ICECE 2008 - 5th International Conference on Electrical and Computer Engineering

Copyright and Reprint Permission

IEEE Catalog Number	CFP0868A-PRT
ISBN	978-1-4244-2014-8
Library of Congress	2007943881

Additional copies of this publication are available from

ICECE 2008 Conference Secretariat
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
Phone: +880-2-9665650-80/7113

ICECE 2008 Committees

Organizing Chair

Aminul Hoque
Department of EEE, BUET

Organizing Secretary

Newaz M. Syfur Rahim
Department of EEE, BUET

Technical Chairs

Mohammad Ali Choudhury
Department of EEE, BUET

Satya Prasad Majumder
Department of EEE, BUET

Technical Secretary

Kazi Mujibur Rahman
Department of EEE, BUET

Web Administrator

Md. Nasim Ahmed Dewan
Department of EEE, BUET

Organizing Committee

ABM Siddique Hossain
Department of EEE, BUET

Md. Abdul Matin
Department of EEE, BUET

Saiful Islam
Department of EEE, BUET

Md. Quamrul Ahsan
Department of EEE, BUET

Shahidul Islam Khan
Department of EEE, BUET

M. M. Shahidul Hassan
Department of EEE, BUET

S. Shahnawaz Ahmed
Department of EEE, BUET

Md. Saifur Rahman
Department of EEE, BUET

Quazi D. M. Khosru
Department of EEE, BUET

Md. Shafiqul Islam
Department of EEE, BUET

Md. Kamrul Hasan
East West University

Md. Aynal Haque
Department of EEE, BUET

S. M. Mominuzzaman
Department of EEE, BUET

Md. Jahangir Alam
Department of EEE, BUET

Md. Shah Alam
Department of EEE, BUET

Hamidur Rahman
Department of EEE, BUET

M. Imamul H. Bhuiyan
Department of EEE, BUET

Md. Farhad Hossain
Department of EEE, BUET

Mohammad Ariful Haque
Department of EEE, BUET

Mahbub Alam
Department of EEE, BUET

Imtiaz Ahmed
Department of EEE, BUET

Chief Technical Officer
Grameen Phone Ltd.

Md. Rabiul Alam
Energypac Engineering Ltd.

Chair
IEEE Bangladesh Section

Chair
IEEE EDS, Bangladesh Chapter

Head
Department of CSE, BUET

Technical Committee

Pran Kanai Saha
Department of EEE, BUET

A. B. M. Harun-Ur-Rashid
Department of EEE, BUET

Md. Quamrul Huda
Department of EEE, BUET

Md. Nasim Ahmed Dewan
Department of EEE, BUET

Abdul Hasib Chowdhury
Department of EEE, BUET

Md. Shofiqul Islam
Department of EEE, BUET

Md. Itrat Bin Shams
Department of EEE, BUET

Rajib Mikail
Department of EEE, BUET

Shafat Jahangir
Department of EEE, BUET

M. Rezwan Khan
United International University

Anisul Haque
East West University

International Advisory Committee

A. H. M. Zahirul Alam
IIUM, Malaysia

Abul N. Khondker
Clarkson University, USA

Anowar Hossain
University of Waterloo, Canada

Ashraful Alam
Purdue University, USA

B. M. Azizur Rahman
City University London, UK

Hamidul Haque
NTU, Singapore

Hiroshi Iwai
Tokyo Inst. of Technology, Japan

Iqbal Hossain
University of Akron, USA

M. Ataul Karim
Old Dominion University, USA

M. Atiquzzaman
Oklahoma University, USA

M. Azizur Rahman
MUN, Canada

M. H. Rashid
University of South Florida, USA

Monzur Murshed
Monash University, Australia

Mohiuddin Majumder
Intel Corporation, USA

Rezaul Karim Beg
Victoria University, Australia

S. Alam
University of South Alabama, USA

Saifur Rahman
Virginia Tech, USA

Shahriar S. Ahmad
Intel Corporation, USA

Takamaro Kikkawa
Hiroshima University, Japan

Takashi Yahagi
Chiba University, Japan

Tapan K. Saha
University of Queensland, Australia

Contents

SESSION 1A1 (9:00AM – 11:00 AM) SUNDAY 21 DECEMBER 2008

POWER SYSTEM - I

SESSION CHAIR: M. A. RAHMAN, MEMORIAL UNIV. OF NEWFOUNDLAND, CANADA

SESSION 1B1 (9:00AM-11:00AM) SUNDAY 21 DECEMBER 2008

BIOMEDICAL SIGNAL PROCESSING - I

SESSION CHAIR: MD. AYNAL HAQUE, EEE, BUET, BANGLADESH

SESSION 1C1 (9:00AM-10:30AM) SUNDAY 21 DECEMBER 2008

COMMUNICATION SYSTEMS – I

SESSION CHAIR: SAIFUL ISLAM, EEE, BUET, BANGLADESH

SESSION 1D1 (9:00AM-10:30AM) SUNDAY 21 DECEMBER 2008

SMICONDUCTOR DEVICES - I

SESSION CHAIR: M. REZWAN KHAN, UNITED INTL. UNIVERSITY, BANGLADESH

INVITED PAPER SESSION 1C1A (10:30AM-11:00AM) SUNDAY 21 DECEMBER 2008

SPEAKER: B. M. AZIZUR RAHMAN, CITY UNIVERSITY OF LONDON, UK

SESSION CHAIR: SAIFUL ISLAM, EEE, BUET, BANGLADESH

INVITED PAPER SESSION 1D1A (10:30AM-11:00AM) SUNDAY 21 DECEMBER 2008

SPEAKER: CHANDAN KUMAR SARKAR, JADAVPUR UNIVERSITY, INDIA

SESSION CHAIR: M. REZWAN KHAN, UNITED INTL. UNIVERSITY, BANGLADESH

SESSION 1A2 (11:15AM-1:15PM) SUNDAY 21 DECEMBER 2008

POWER ELECTRONICS – I

SESSION CHAIR: M. H. RASHID, UNIVERSITY OF SOUTH FLORIDA, USA

SESSION 1B2 (11:15AM-1:15PM) SUNDAY 21 DECEMBER 2008

COMMUNICATION SYSTEMS – II

SESSION CHAIR: B. M. AZIZUR RAHMAN, CITY UNIVERSITY OF LONDON, UK

SESSION 1C2 (11:15AM-1:15PM) SUNDAY 21 DECEMBER 2008

MEMS AND NANO TECHNOLOGY

SESSION CHAIR: SAIF ISLAM, UCDAVIS, USA

SESSION 1D2 (11:15AM-1:15PM) SUNDAY 21 DECEMBER 2008

VLSI – I

SESSION CHAIR: A. B. M. H. RASHID, EEE, BUET, BANGLADESH

SESSION 1A3 (2:15PM-4:15PM) SUNDAY 21 DECEMBER 2008

POWER SYSTEM – II

SESSION CHAIR: S. SHAHNAWAZ AHMED, EEE, BUET, BANGLADESH

SESSION 1B3 (2:15PM-4:15PM) SUNDAY 21 DECEMBER 2008

DIGITAL SIGNAL PROCESSING

SESSION CHAIR: MD. KAMRUL HASAN, EAST WEST UNIVERSITY, BANGLADESH

SESSION 1C3 (2:15PM-4:15PM) SUNDAY 21 DECEMBER 2008

SEMICONDUCTOR DEVICES – II

SESSION CHAIR: M. M. SHAHIDUL HASSAN, EEE, BUET, BANGLADESH

SESSION 1D3 (2:15PM-4:15PM) SUNDAY 21 DECEMBER 2008

COMMUNICATION NETWORK – I

SESSION CHAIR: PRAN KANAI SAHA, EEE, BUET, BANGLADESH

SESSION 2A1 (9:00AM-11:00AM) MONDAY 22 DECEMBER 2008

POWER ELECTRONICS – II

SESSION CHAIR: SHAHIDUL I. KHAN, EEE, BUET, BANGLADESH

SESSION 2B1 (9:00AM-11:00AM) MONDAY 22 DECEMBER 2008

OPTOELECTRONIC DEVICES AND PHOTONICS

SESSION CHAIR: MD. QUAMRUL HUDA, EEE, BUET, BANGLADESH

SESSION 2C1 (9:00AM-11:00AM) MONDAY 22 DECEMBER 2008

COMMUNICATION NETWORK – II

SESSION CHAIR: ABM SIDDIQUE HOSSAIN, EEE, BUET, BANGLADESH

SESSION 2D1 (9:00AM-11:00AM) MONDAY 22 DECEMBER 2008

COGNITIVE INTELLIGENCE AND DATABASE SYSTEMS – I

SESSION CHAIR: MD. SAIFUR RAHMAN, EEE, BUET, BANGLADESH

SESSION 2A2 (2:15PM-4:15PM) MONDAY 22 DECEMBER 2008

POWER SYSTEMS AND DRIVES

SESSION CHAIR: MD. QUAMRUL AHSAN, EEE, BUET, BANGLADESH

SESSION 2B2 (2:15PM-4:15PM) MONDAY 22 DECEMBER 2008

VLSI – II

SESSION CHAIR: DIDAR ISLAM, MD & CEO, POWER IC, BANGLADESH

SESSION 2C2 (2:15PM-4:15PM) MONDAY 22 DECEMBER 2008

MICROWAVE AND RF SYSTEMS

SESSION CHAIR: A. H. M. ZAHIRUL ALAM, ECE, IIUM, MALAYSIA

SESSION 2D2 (2:15PM-4:15PM) MONDAY 22 DECEMBER 2008

THIN FILM TECHNOLOGY

SESSION CHAIR: MD. SHAFIQUL ISLAM, EEE, BUET, BANGLADESH

SESSION 2A3 (5:30PM-6:30PM) MONDAY 22 DECEMBER 2008

CONTROL SYSTEMS AND POWER CONVERTERS

SESSION CHAIR: MD. ENAMUL BASHER, EEE, BUET, BANGLADESH

SESSION 2B3 (5:30PM-6:30PM) MONDAY 22 DECEMBER 2008

COMMUNICATION SYSTEMS AND NETWORKS

SESSION CHAIR: S. P. MAJUMDER, EEE, BUET, BANGLADESH

SESSION 2C3 (5:30PM-6:30PM) MONDAY 22 DECEMBER 2008

MATERIALS AND SEMICONDUCTOR DEVICES

SESSION CHAIR: QUAZI D. M. KHOSRU, EEE, BUET, BANGLADESH

SESSION 2D3 (5:30PM-6:30PM) MONDAY 22 DECEMBER 2008

COGNITIVE INTELLIGENCE AND DATABASE SYSTEMS – II

SESSION CHAIR: MD. SAIDUR RAHMAN, CSE, BUET, BANGLADESH

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Improvement of Load Frequency Control With Fuzzy Gain Scheduled SMES Unit Considering Governor Dead-Band and GRC

M.R.I. Sheikh, S.M. Muyeen, R. Takahashi, Toshiaki Murata and Junji Tamura
Kitami Institute of Technology, 165 Koen-cho, Kitami, Hokkaido, 090-8507, Japan
Email: sheikh@pullout.elec.kitami-it.ac.jp

Abstract - Since a Superconducting Magnetic Energy Storage (SMES) unit with a self-commutated converter is capable of controlling both the active and reactive power simultaneously and quickly, increasing attention has been focused recently on power system stabilization by SMES control. In this study, a fuzzy gain scheduled supplementary control scheme with SMES unit is proposed and applied to Automatic Generation Control (AGC) in power system for the improvement of Load Frequency Control (LFC). The performances of the system for load changes in the areas in the interconnected power system are studied. The computer simulation of the interconnected power system shows that SMES unit with the proposed gain scheduled supplementary controller can perform a more effective primary frequency control for multi area power system.

I. Introduction

Automatic generation control is a very important subject in power system operation for supplying sufficient and reliable electric power with quality. Frequency variations in interconnected power systems can cause large-scale serious instability problems. LFC is one of control schemes to provide the stable and reliable operation in multi-area power systems. For stable operation, constant frequency and active power balance must be provided. To improve the stability of the power networks, it is necessary to design LFC systems that control the power generation and active power on tie-lines. In an interconnected power system, as the load demand varies randomly, the area frequency and tie-line power interchange also vary. The objective of LFC is to minimize the transient deviations in these variables and to ensure their steady state values to be zero. The LFC by only a governor control of synchronous generators imposes a limit on the degree to which the deviations in frequency and tie-line power exchange can be minimized. However, as the fundamental purpose of LFC is solving the problem of an instantaneous mismatch between the generation and demand of active power, the incorporation of a fast-acting energy storage device in the power system can improve the performance under such conditions. But fixed gain controllers based on classical control theories are presently used. They are not sufficient for the case with changing operating point during a daily cycle [1–4] and also not suitable for all operating conditions. Therefore, variable structure controller [5–7] has been proposed for AGC. For designing controllers based on these techniques, the perfect model is required which has to track the state variables and satisfy system constraints. Therefore it is difficult to apply these adaptive control techniques to AGC in practical implementations. In multi area power system, if a load variation occurs at any one of the areas in the system, the frequency related with this area is affected first and then that of other areas are also affected from this perturbation through tie-lines.

In this study, the same gain scheduled controller is used to implement AGC in the interconnected system having two areas including SMES units when a step load perturbation occurs in one or both areas. In the model system, each area in the interconnected system includes steam reheat turbines and generation rate constraints. We reported a work [8] for LFC by fuzzy gain scheduled SMES. However, in our previous study [8] the governor dead-band (DB) and generation rate constraints (GRC) were not considered. In the present work effect of boiler system and governor DB and GRC are also considered, by which the worst situation of power system can be considered. When a small load disturbance in any area of the interconnected system occurs, tie-line power deviations and power system frequency oscillations continue for a long duration. To damp out the oscillations in a short time, automatic generation control including a SMES unit with the proposed gain scheduled supplementary controller is used. The basic objective of the supplementary control is to restore balance between each area load and generation for a load disturbance. This is met when the control action maintains the frequency and the tie-line power interchange at the scheduled values. The supplementary controller with integral gain K_{Ii} is therefore made to act on area control error, which is a signal obtained from tie-line power flow deviation added to frequency deviation weighted by a bias factor β.

$$ACE_i = \sum_{j=1}^{n} \Delta P_{tie,\,i\,j} + \beta_i \Delta f_i \qquad (1)$$

where the suffix i refer to the control area and j refer to the number of generator.

Using fuzzy logic, the integrator gain (K_{Ii}) of supplementary controller is so scheduled that it compromise between fast transient recovery and low overshoot in dynamic response of the system. It is seen that with the addition of gain scheduled supplementary controller, a simple controller scheme for SMES is sufficient to improves effectively the damping of the oscillations after the load deviation in one or both of the areas in the interconnected system.

II. Integration of SMES with Two-Area Power System

Figure 1 shows the two-area power system with SMES unit used in the analyses. Two areas are connected by a weak tie-line. When there is a sudden rise in power demand in a control area, the stored energy is almost immediately released by the SMES through its power conversion system (PCS). As the governor control mechanism starts working to set the power system to the new equilibrium condition, the SMES coil stores energy back to its nominal level. Similar action happens when there is a sudden decrease in load demand. Basically, the operation speed of governor-turbine system is slow compared with that of the excitation system. As a result, fluctuations in terminal voltage can be corrected by the excitation system very quickly, but fluctuations in generated power or frequency are corrected slowly. Since load frequency control is primarily concerned with the real power/frequency behavior, the excitation system model will not be required in the approximated analysis. This important simplification paves the way for constructing the simulation model shown in Fig. 1.

All the governors have dead-band which affects the stability of the system and produces a continuous sinusoidal oscillation of natural period. So effects of governor dead-band are studied in relation to AGC. The limiting value of dead-band is specified as 0.06%. Also in practical steam turbine, due to thermodynamic and mathematical constraints, there is a limit to the rate at which its output power (dP_t/dt) can be changed. This limit is referred to as Generation rate constraint (GRC). In practice, there exists a maximum limit on the rate of change in the generating power of a steam plant. In the presence of GRC, the dynamic responses of the system experience larger overshoots and longer settling time compared to the case without considering the GRC. Hence, if the load change are too fast under transient conditions, then system nonlinearities will prevent its achievement. Moreover, if the parameters of the controller are not chosen properly, the system may become unstable. So considering these, the GRC is taken into account by adding a limiter to the turbine as shown in Fig. 2, with a value of 0.1 p.u. MW/min [9] as shown in eq.(2). This is a typical value up to 3.4 MW/second. All parameters are same as that used in [8].

$$\Delta\dot{P}_{generation} = 0.1 \text{ p.u.MW/min} = 0.0017 \text{p.u.MW/sec} = \delta \qquad (2)$$

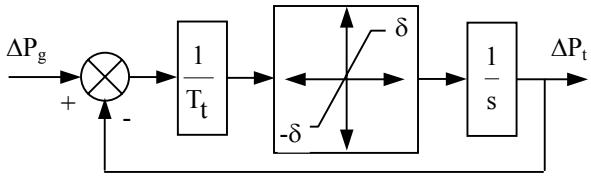


Fig. 2: A non-linear turbine model with GRC

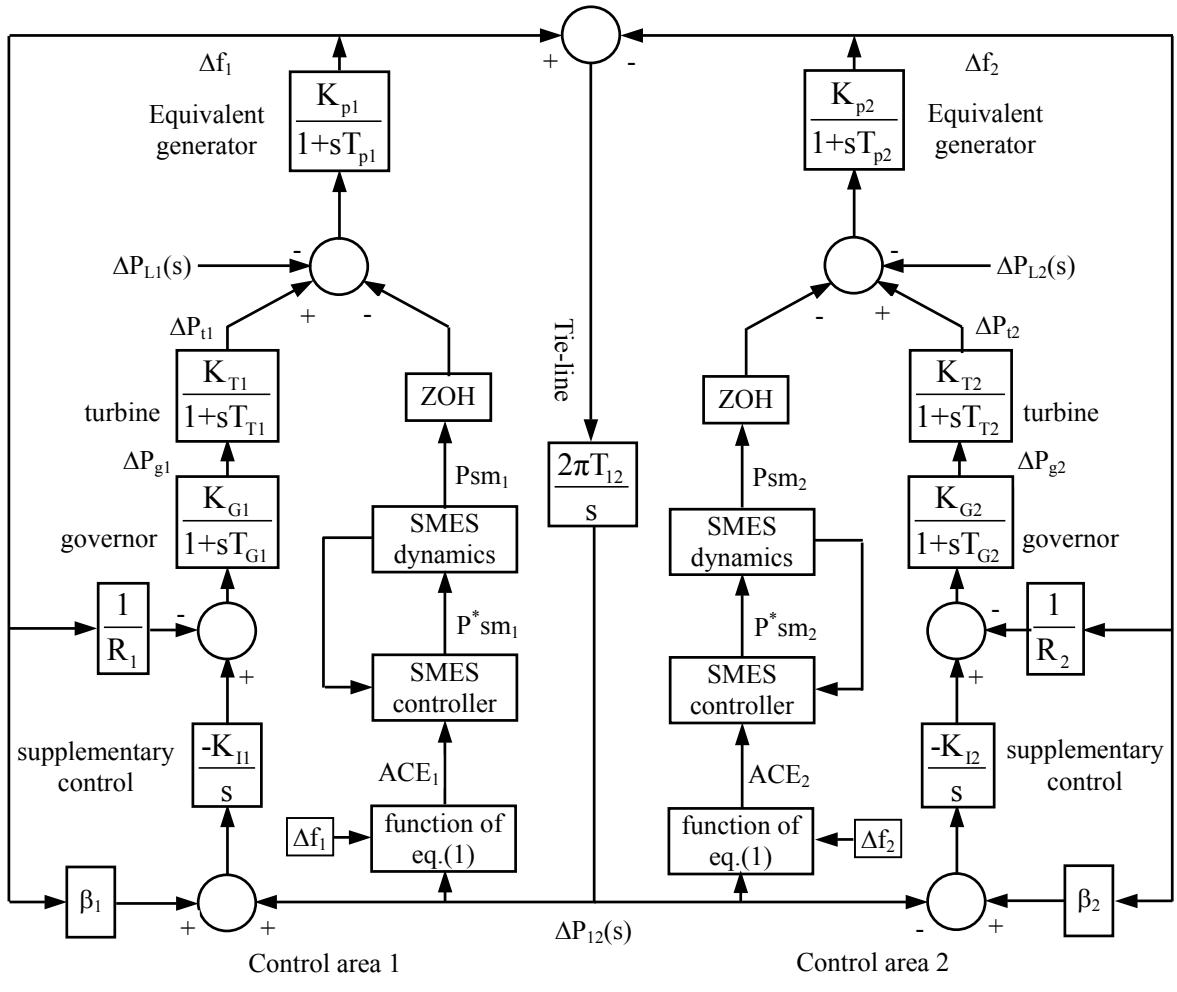


Fig. 1 Simulation model for the two-area power system

III. Optimization of the Integral Gain, K_I and Frequency Bias Factors, β in Multi-Area Power System

Figure 3 shows the frequency deviations for different values of K_I for a specific load change. It is observed that a higher value of K_I results in reduction of maximum deviation of the system frequency but the system oscillates for longer times. Decreasing the value of K_I yields comparatively higher maximum frequency deviation at the beginning but provides very good damping in the later cycles. These initiate a variable K_I, which can be determined from the frequency error and its derivative. Obviously higher values of K_I is needed at the initial stage and then it should be changed gradually depending on the system frequency change.

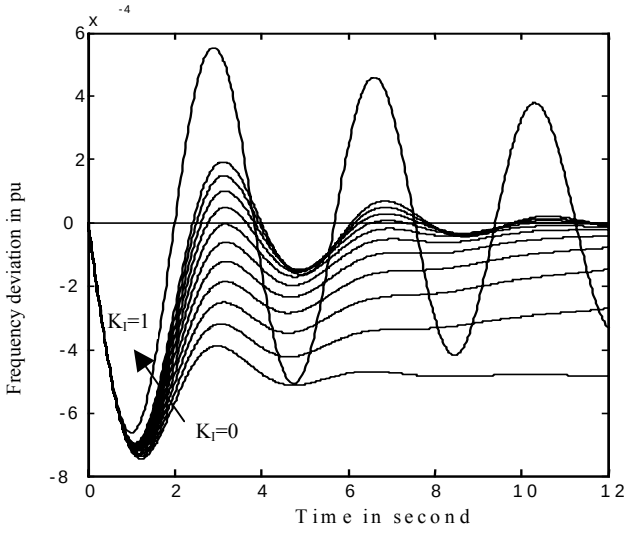


Fig. 3 Frequency deviation step response for different values of K_I

Dynamic performance of the AGC system would obviously depend on the value of frequency bias factors, $\beta_1 = \beta_2 = B$ and integral controller gain value, $K_{I1}=K_{I2}=K_I$. In order to optimize B and K_I the concept of maximum stability margin is used, evaluated by the eigen values of the closed loop control system [7].

For a fixed gain supplementary controller, the optimal values of K_I and B are chosen, here, on the basis of a performance index (PI) given in eq.(3) for a specific load change. The Performance Index (PI) curves are shown in Fig. 4 with considering governor dead-band (DB) and generation rate constraints (GRC).

$$P.I. = \int_0^{40} \left(\Delta P_{tie}^2 + w_1 \Delta f_1^{\,2} + w_2 \Delta f_2^{\,2} \right) dt \qquad (3)$$

Where, w_1 and w_2 are the weight factors. The weight factors w_1 and w_2 both are chosen as 0.25 for the system under consideration.

From Fig. 4, in the presence of DB & GRC it is observed that the value of integral controller gain, K_I = 0.28 and frequency bias factors, B=0.15 which occurs at PI= 0.0363.

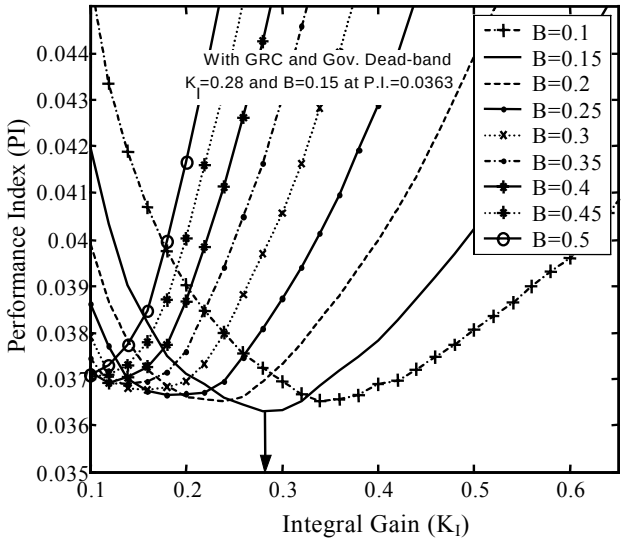


Fig. 4 The optimal integral controller gain, K_I and frequency bias factor, B with DB and GRC

IV. Fuzzy Gain Scheduler PI Control (FGSPI)

Figure 5 shows the membership functions for PI control system with a fuzzy gain scheduler. The approach taken here is to exploit fuzzy rules and reasoning to generate controller parameters. The triangular membership functions for the proposed FGSPI controller of the three variables (e_t, $c\dot{e}_t$, K_{Ii}) are shown in Fig. 5, where frequency error (e_t) and change of frequency error ($c\dot{e}_t$) are used as the inputs of the fuzzy logic controller. K_{Ii} (i=1,2) is the output of fuzzy logic controller. Considering these two inputs, the output of gain K_{Ii} is determined. The use of two input and single output variables makes the design of the controller very straightforward. A membership value for the various linguistic variables is calculated by the rule given by

$$\mu\left(e_t, c\dot{e}_t\right) = \min\left[\mu\left(e_t\right), \mu\left(c\dot{e}_t\right)\right] \qquad (4)$$

The equation of the triangular membership function used to determine the grade of membership values in this work is as follows:

$$A\left(x\right) = \frac{\left(b - 2\left|x-a\right|\right)}{b} \qquad (5)$$

Where A(x) is the value of grade of membership, 'b' is the width and 'a' is the coordinate of the point at which the grade of membership is 1 and x is the value of the input variables. The control rules for the proposed strategy are very straightforward and have been developed from the viewpoint of practical system operation and by trial and error methods. The fuzzy rule base for the FGSPI controller is shown in Table I.

The membership functions, knowledge base and method of defuzzification determine the performance of the FGSPI controller in a multi area power system as shown in eq. (6).

$$K_{Ii} = \frac{\sum_{j=1}^{n} \mu_j u_j}{\sum_{j=1}^{n} \mu_j} \quad (6)$$

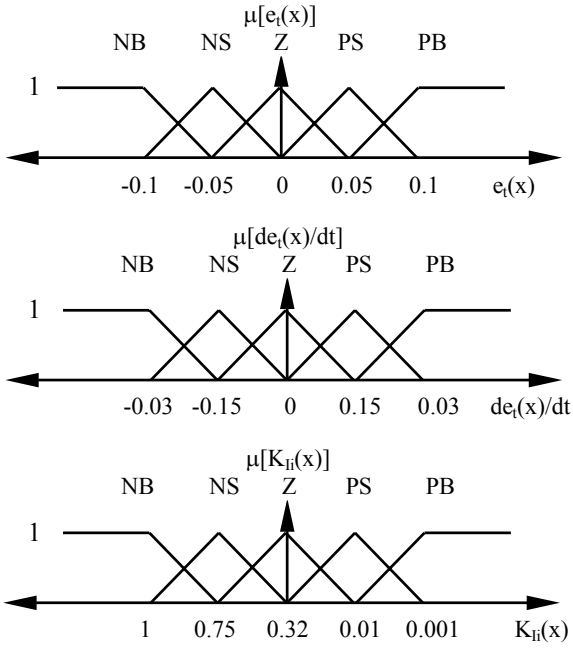


Fig. 5 Membership functions for the fuzzy variables

Table I
Fuzzy Rule Base for FGSPI Controller

ce \ e	NB	NS	Z	PS	PB
NB	PB	PB	PB	PS	Z
NS	PB	PB	PS	Z	Z
Z	NS	NS	Z	NS	NB
PS	Z	Z	NS	PB	NB
PB	Z	NS	NB	NB	NB

V. Control System of SMES

The schematic diagram in Fig. 6 shows the configuration of a thyristor controlled SMES unit, which is incorporated in each control area of power system for LFC as shown in Fig. 7. The converter firing angle controls the DC voltage V_{sm} appearing across the inductor to be continuously varied between a wide range of positive and negative values. The inductor is initially charged to its rated current by applying a low positive voltage. Once the current reaches the rated value, it is maintained constant by reducing the voltage across the inductor to zero.

Figure 8 outlines the proposed simple control scheme for SMES, which is incorporated in each control area to reduce the instantaneous mismatch between demand and generation. ACE_i (i=1,2) in each control area is taken as the control input signal for SMES. It is desirable to restore the inductor current to its rated value as quickly as possible after a system disturbance, so that the SMES unit can respond properly to any subsequent disturbance. So inductor current deviation is sensed and used as negative feedback signal in the SMES control loop to achieve quick restoration of current and SMES energy level.

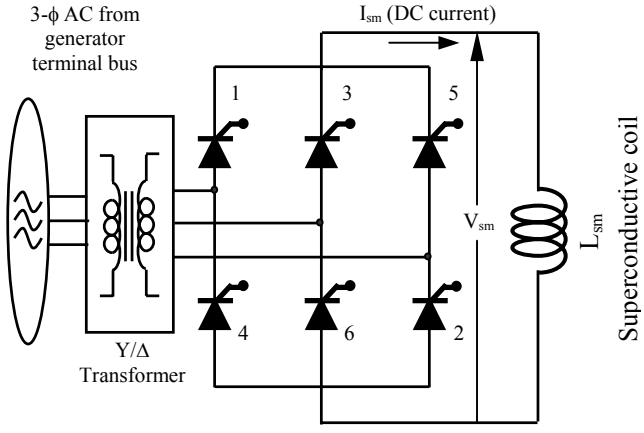


Fig. 6. SMES unit with 6-pulse bridge AC/DC thyristor controlled converter

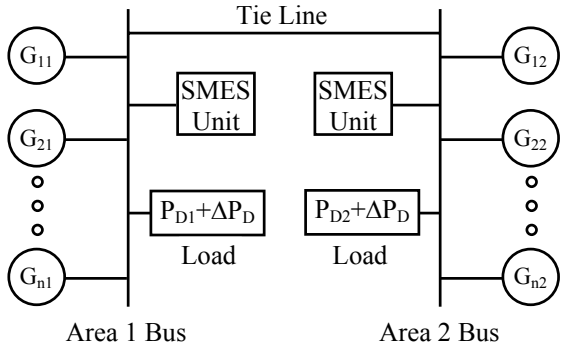


Fig. 7 Configuration of SMES in a two-area power system

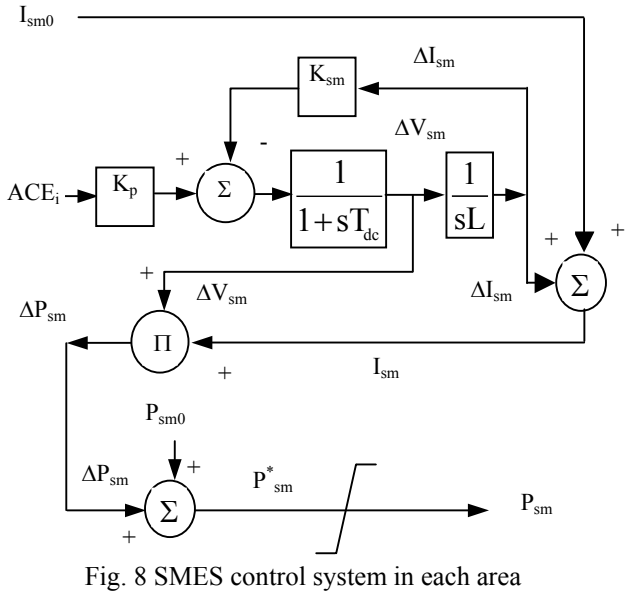


Fig. 8 SMES control system in each area

VI. Simulation Results

To demonstrate the usefulness of the proposed controller, computer simulations were performed using the MATLAB environment under different operating conditions. The system performances with gain scheduled SMES and fixed gain SMES are shown in Fig. 9 through Fig. 14. Three cases studies are conducted.

Case I: a step load increase (ΔP_{L1}=0.015 pu MW) in area1.

In this case, it is seen from Fig. 9 that the tie-line power deviation are reduced with the proposed gain scheduled SMES controller and the deviations are negetive. Thus sensing the input signal ACE_i in the control areas SMES provide sufficient compensation, and it is seen from Fig. 10 that SMES in area1 is discharging energy and SMES in area2 is charging energy to keep the frequency deviations in both areas minimum. From Fig. 10, it is also seen that FGSPI controller of the loaded area determines the integral gain K_I to a scheduled value to resotore the frequency to its nominal value, and FGSPI controller of the unloaded area reamains unscheduled and selects the critical value as its integral gain. Finally it is seen that the damping of the system frequency is not satisfactory for the fixed gain controller. But the proposed gain scheduled supplementary controller significantly improves the system performances.

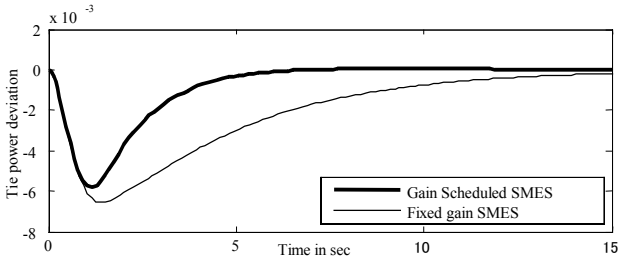


Fig. 9 Performance of tie power deviation [**Case I**]

Case II: the same step load increase in both areas.

In this case, the same load increase, $\Delta P_{L1}=\Delta P_{L2}$= 0.01 p.u MW, is applied to both areas. It is seen from Fig. 11 that the tie-line power deviation is zero. Thus SMES compensation depends on Δf_i in both areas. As the load change is same in both areas, the SMES in both areas provide same compensation. Finally it is seen from Fig. 13 that FGSPI controller of both the loaded areas determine the integral gain K_{Ii} (i=1,2) to a scheduled value to resotore the frequency to its nominal value. Due to this, the damping of the system frequency is also significantly improved with the proposed controller.

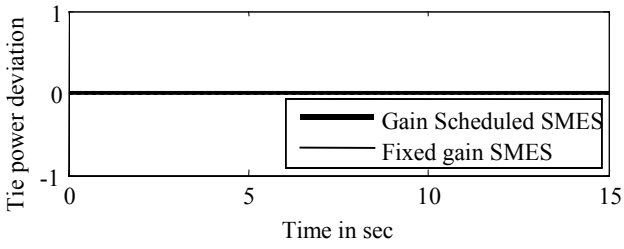


Fig. 11 Performance of tie power deviation [**Case II**]

Case III: the different step load increases are applied to each Area.

In this case, as each area is loaded by the different increase, each area adjusts their own load. Figure 12 shows the tie power deviation but the magnitude is small. So the SMES controller in both areas dominated on Δf_i. As ΔP_{L1}=0.01 p.u MW & ΔP_{L2}=0.015 p.u MW, it is seen from Fig. 14 that SMES in area2 provided more compensation than area1. Finally frequency deviations restore to its nominal value with the gain scheduled SMES controller.

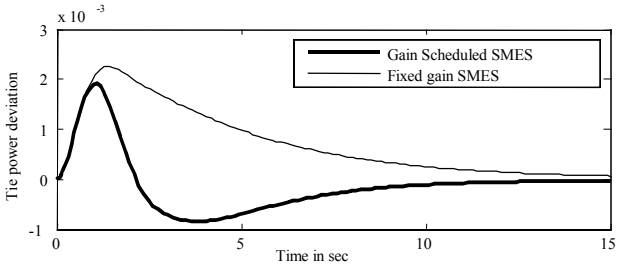


Fig. 12 Performance of tie power deviation [**Case III**]

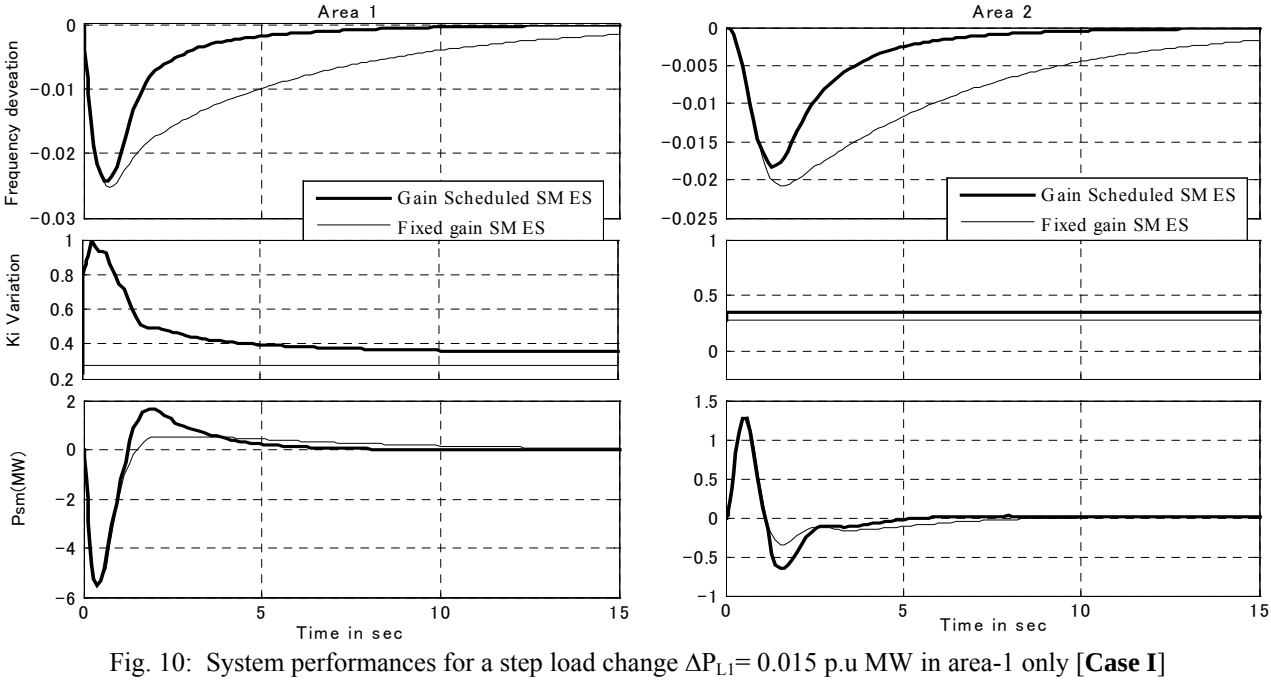


Fig. 10: System performances for a step load change ΔP_{L1}= 0.015 p.u MW in area-1 only [**Case I**]

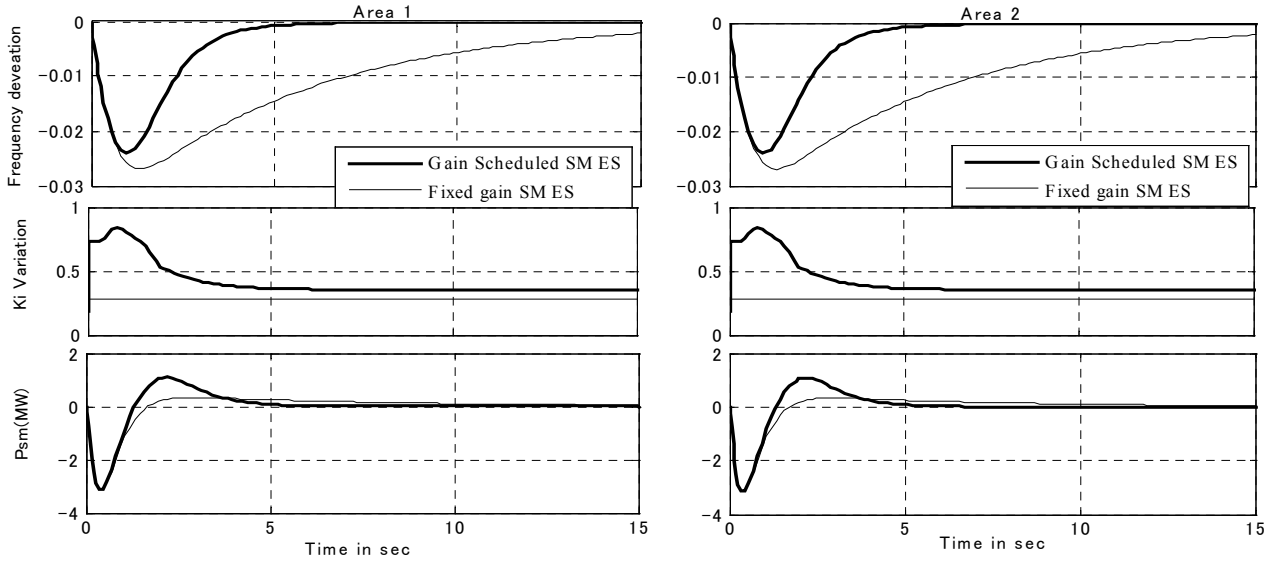


Fig. 13: System performances for a step load change $\Delta P_{L1}=\Delta P_{L2}= 0.01$ p.u MW in both areas [**Case II**]

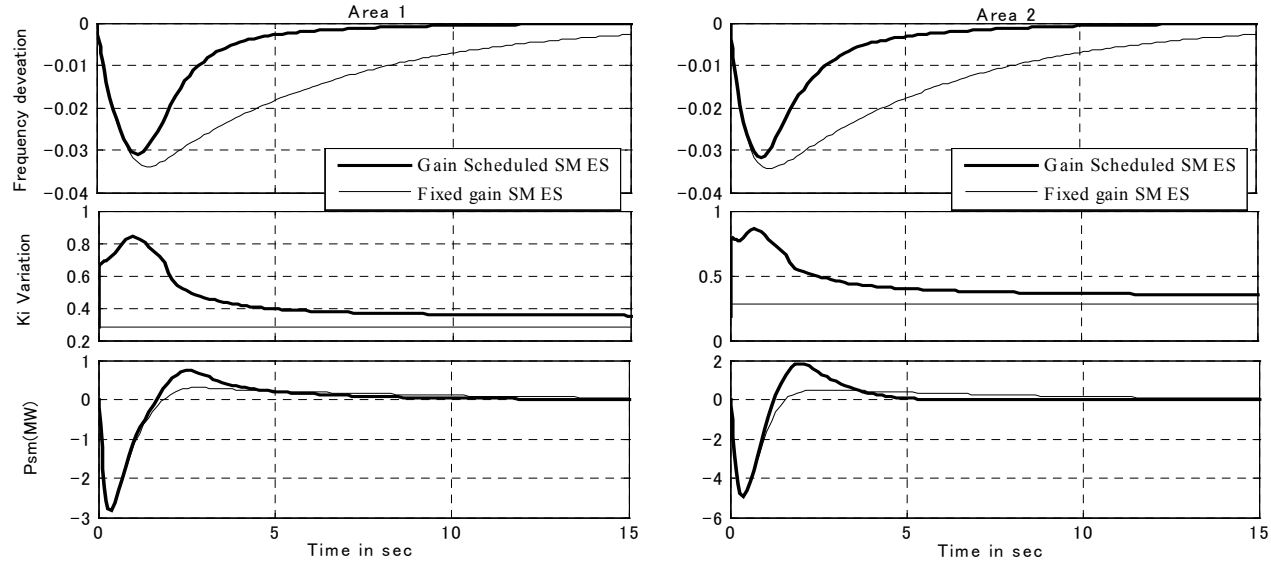


Fig. 14: System performances for a step load change $\Delta P_{L1}=0.01$ p.u MW in area1& $\Delta P_{L2}= 0.015$ p.u MW in area2 [**Case III**]

VII. Conclusions

The simulation studies are carried out on a two-area power system considering DB and GRC to investigate the impact of the proposed intelligently controlled SMES on the power system dynamic performance. The results show that the scheme is very powerful in reducing the frequency and tie-power deviations under a variety of load perturbations. On line adaptation of supplementary controller gain associated with SMES makes the proposed intelligent controllers more effective and are expected to perform optimally under different operating conditions.

References:

[1] Benjamin NN, Chan WC.:"Multilevel Load-frequency Control of Inter-Connected Power Systems", *IEE Proceedings, Generation, Transmission and Distribution*,1978; No.125: pp.521–526.

[2] Nanda J, Kavi BL.:"Automatic Generation Control of Interconnected Power System", *IEE Proceedings, Generation, Transmission and Distribution*, 1988; No.125(5): pp.385–390.

[3] Das D, Nanda J, Kothari ML, Kothari DP.:"Automatic Generation Control of Hydrothermal System with New Area Control Error Considering Generation Rate Constraint", *Electrical Machines and Power System* 1990; 18:461–471.

[4] Mairaj uddin Mufti, Shameem Ahmad Lone, Sheikh Javed Iqbal, Imran Mushtaq:"Improved Load Frequency Control with Superconducting Magnetic Energy Storage in Interconnected Power System", *IEEJ Transaction*, 2007, vol. 2,pp. 387-397.

[5] Benjamin NN, Chan WC.:"Variable Structure Control of Electric Power Generation", *IEEE Transactions on Power Apparatus and System* 1982; 101(2):376–380.

[6] Sivaramaksishana AY, Hariharan MV, Srisailam MC.:"Design of Variable Structure Load-Frequency Controller Using Pole Assignment Techniques", *International Journal of Control* 1984; 40(3):437–498.

[7] Tripathy SC, Juengst KP.:"Sampled Data Automatic Generation Control with Superconducting Magnetic Energy Storage", *IEEE Transactions on Energy Conversion* 1997; 12(2):187–192.

[8] M.R.I. Sheikh, S.M. Muyeen, Rion Takahashi, Toshiaki Murata and Junji Tamura "Improvement of Load Frequency Control with Fuzzy Gain Scheduled Superconducting Magnetic Energy Storage Unit", *International Conference of Electrical Machine (ICEM, 08),* Conference CD, Paper ID-1026, 06-09 September, 2008, Portugal.

[9] C.T. Pan,C. M. Lian, "An Adaptive Controller For Power System Load-Frequency Control", *IEEE Transactions on Power System*, Vol. 4, No. 1, February, 1988.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Computerized Modelling of Hybrid Energy System—Part I: Problem Formulation and Model Development

Ajai Gupta, R. P. Saini, and M. P. Sharma

Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee
Roorkee, Uttarakhand-247667, India
E-mail: ajai_abjc@yahoo.com (Ajai Gupta)

Abstract – A well designed hybrid energy system can be cost effective, has a high reliability and can improve the quality of life in remote rural areas. The economic constraints can be met, if these systems are fundamentally well designed, use appropriate technology and make use effective dispatch control techniques. The first part of this tri-series paper, presents the analysis and design of a mixed integer linear mathematical programming model (time series) to determine the optimal operation and cost optimization for a hybrid energy generation system consisting of a photovoltaic array, biomass, biogas, small/micro hydro, a battery bank and a fossil fuel generator. The optimization is aimed at minimizing the cost function based on demand and potential constraints. Further, mathematical models of all other components of hybrid energy system are also developed. This is the generation mix of the remote rural area of India; it may be applied to other rural areas also.

***Index terms*-Hybrid Energy System, Micro hydro, Solar Photovoltaic System, Biomass energy system, Diesel generator, Battery storage, Integer programming.**

I. Introduction

Hybrid energy systems (HES) generally integrate renewable energy sources with fossil fuel powered diesel/petrol generator to provide electric power where the electricity is either fed directly into the grid or to batteries for energy storage. The role of integrating renewable energy in a hybrid energy system is primarily to save diesel fuel. Examples of renewable energy sources commonly used in hybrid configurations are small wind turbines, photovoltaic systems, micro-hydro, biomass, fuel cells and stirling engines.

A hybrid energy system consists of two or more energy systems, an energy storage system, power conditioning equipment and a controller [1]. A hybrid energy system may or may not be connected to the grid. They are generally independent of large centralized electric grids and are used in rural remote areas. In these systems it is possible for the individual power sources to provide different percentages of the total load.

The Hybrid Energy System has received much attention over the past decade. It is a viable alternative solution as compared to systems, which rely entirely on hydrocarbon fuel. Apart from the mobility of the system, it also has longer life cycle. In particular, the integrated approach [2] makes a hybrid system to be the most appropriate for isolated communities of a rural remote area.

For systems employing totally clean renewable energy, high capital cost is an important barrier. However, we can produce green power by adding different renewable energy sources to diesel generator and battery, which is also called a hybrid system. This kind of system can compromise investment cost, diesel fuel usage cost and also operation and maintenance costs.

There are generally two accepted hybrid energy system configurations:

- Systems based mainly on diesel generators with renewable energy used for reducing fuel consumption.
- Systems relying on the renewable energy source with a diesel generator used as a back-up supply for extended periods of low renewable energy input or high load demand.

Optimization of a hybrid energy system is site specific and it depends upon the resources available and the load demand. The aim of this study is to identify the most economic and appropriate power supply for a selected remote rural area.

II. Literature Review Conclusions

Literature review reveals that the modelling of hybrid energy system and their application in decentralized mode are quite limited. The models applied, are based on one of the available resources, while the literature has focused on one or two available resources. Further, attempts for developing optimum energy mix of different resources for meeting the energy needs of the rural people are also limited.

Application of models for matching the projected energy demand with mix of sources at decentralized level is limited. The models developed so far mainly focus on rural areas and not individual villages, clusters of villages, blocks, or district. Very limited efforts are reported for block level planning and are not based on any optimization approach [3-4].

III. Problem Formulation

From the literature, it is observed that lot of work has been carried out for modelling of hybrid energy systems.

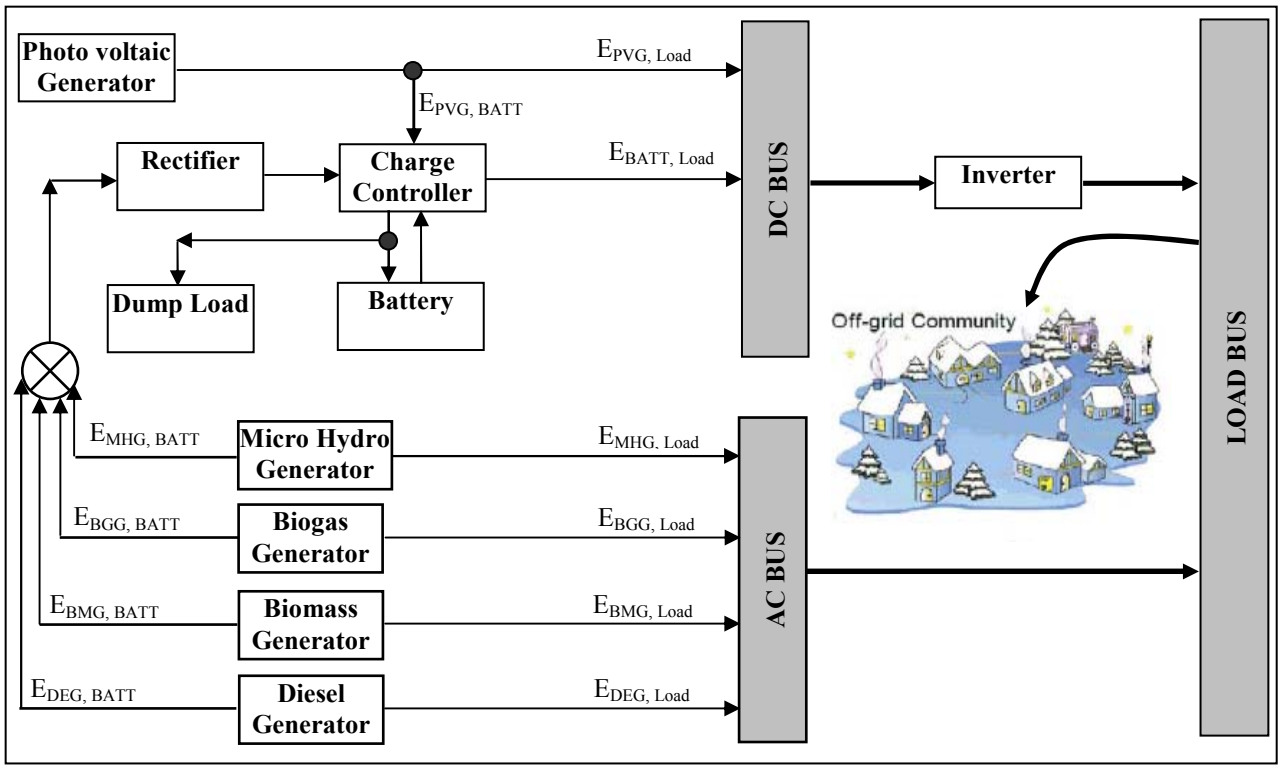


Fig. 1 The configuration of hybrid energy system

The renewable energy sources considered under these studies are mainly solar, biomass, and wind but a very little work has been reported for the modelling of hybrid energy systems involving small hydro/micro hydro power in combination with conventional system.

The main objective of the study is to develop a model of Hybrid Energy System with more emphasis on small hydro power (SHP)/micro hydro for a remote rural area in cost effective manner.

In this paper, a mixed integer linear programming model, which is a special type of mathematical programming model, is developed to solve the problem of designing hybrid energy systems. Details of the model development are discussed in the next section.

IV. Hybrid Energy System Configuration

The block diagram for a typical stand-alone hybrid energy system is shown in Fig. 1. The system consists of micro-hydro, biogas, biomass (fuelwood), photovoltaic generator, battery bank storage, back up diesel generator, and dump load. Provisions for the availability of both AC and DC buses are made using electronic converters.

In this study, a hybrid energy system is based on a generalized three-bus configuration the three buses are a DC bus, an AC bus and a load bus. Technologies that generate DC current — photovoltaic array and battery — are connected to DC bus. Technologies that generate AC current — micro hydro generator, biogas generator, biomass (fuel wood) generator, and diesel generator — are connected to AC bus. An inverter, or a DC-to-AC converter, is used to convert DC current to AC current (from the DC bus to serve the AC load). An AC-to-DC converter is used to convert AC current from micro hydro generator, biogas generator, biomass (fuel wood) generator and diesel generator to DC current to charge the battery. The energy from all generators (either renewable or non renewable) is allowed to charge battery bank.

As shown in Fig. 1 the energy generated from all generators can be directed to serve the load and charge the battery. These relationships are expressed in eq. 1-1 through 1-5, for photovoltaic generator (PVG), micro hydro generator (MHG), biogas generator (BGG), biomass generator (BMG) and diesel generator (DEG) respectively.

$$E_{PVG}(t) = E_{PVG,\,load}(t) + E_{PVG,\,BATT}(t) \quad 1\text{-}1$$
$$E_{MHG}(t) = E_{MHG,\,load}(t) + E_{MHG,\,BATT}(t) \quad 1\text{-}2$$
$$E_{BGG}(t) = E_{BGG,\,load}(t) + E_{BGG,\,BATT}(t) \quad 1\text{-}3$$
$$E_{BMG}(t) = E_{BMG,\,load}(t) + E_{BMG,\,BATT}(t) \quad 1\text{-}4$$
$$E_{DEG}(t) = E_{DEG,\,load}(t) + E_{DEG,\,BATT}(t) \quad 1\text{-}5$$

In any hour (t), the energy available to charge the battery bank (BB) is shown in eq. 1-6, and the energy available from the battery to serve the load are shown in eq. 1-7.

$$E_{BATT,\,in}(t) = \eta_{CC} \times \eta_{CHG}\,[E_{PVG,\,BATT}(t) + \eta_{REC} \times (E_{MHG,\,BATT}(t) + E_{BGG,\,BATT}(t) + E_{BMG,\,BATT}(t) + E_{DEG,\,BATT}(t))] \quad 1\text{-}6$$

$$E_{BATT,\,load}(t) = \eta_{DCHG}\,[E_{BATT,\,in}(t)] \quad 1\text{-}7$$

Finally, the total energy available to serve the load is written in eq. 1-8.

$$E_{load}(t) = [E_{MHG,\,load}(t) + E_{BGG,\,load}(t) + E_{BMG,\,load}(t) + E_{DEG,\,load}(t) + \eta_{inv}\,(E_{PVG,\,load}(t) + E_{BATT,\,load}(t))] \quad 1\text{-}8$$

Where, Ex = Energy output from technology x in kWh.
X = x stands for MHG, BGG, BMG, PVG, DEG, and BB.
$E_{x,\,load}$ = Energy output from x directed to load.
$E_{x,\,BATT}$ = Energy output from x directed to battery.
$E_{BATT,\,in}$ = Energy input to battery.

V. System Component Modelling

A. Mathematical Model of Hydro Generator

- Micro hydro Electricity: The hydro power usually refers to the generation of shaft power from falling water. The power is then used for direct mechanical purposes or, more frequently, for generating electricity. A micro-hydro installation can produce power continuously thereby generating larger amounts of energy than PV or wind turbines of similar size, except in adverse weather conditions such as droughts or freezing temperatures.
- Mathematical model: The theoretical electrical power generated by the micro-hydro unit is given by

$$P_{MHG}(t) = 9.81\ Q \times \rho \times h \quad (2.1)$$

and the total energy in kWh in a hour t is given by

$$E_{MHG} = P_{MHG}(t) \times \eta \quad (2.2)$$

Where, Q = discharge in m^3/sec
ρ = density of water = 1000 kg/m^3
h = head in m
η = overall efficiency of the micro hydro generator

B. Mathematical Model of Biomass Generator

- Biomass (fuel wood) based electricity: As an energy resource, biomass is very versatile in terms of the variety of forms and the number of options available for its utilization. Biomass is a generic term used to denote all substances that originate from biological sources (e.g. plants, trees, crops, animals and humans). To simplify our study, we will concentrate only on fuel wood in this section.

A biomass gasifier based electricity generation system consists of biomass preparation unit, biomass gasifier, gas cooling and cleaning system, internal combustion engine suitable for operation in dual fuel mode, and electric generator. Biomass preparation unit is used to cut the collected biomass to proper size for feeding into biomass gasifier where gasification takes place under controlled thermal conditions. Gasification is a thermo-chemical process, converting the biomass to producer gas to be used in an internal combustion engine. The producer gas can be used to generate power using a diesel engine with diesel as pilot fuel and producer gas as main fuel. This mode of operation is known as dual fuel mode.

- Mathematical model: The annual delivered electricity output (E_{Annual}) of a biomass gasifier energy system with rated power output (P_{BMG}) of electricity generator is dependent on its capacity utilization factor (CUF). Assuming a 21 % conversion efficiency from fuelwood to electricity, it can be modelled using the following expression:

$$E_{Annual} = P_{BMG}\ (8760 \times CUF) \quad (3.1)$$

And hourly energy output (E_{BMG}) is given by

$$E_{BMG}(t) = P_{BMG}(t) \quad (3.2)$$

C. Mathematical Model of Biogas Generator

- Biogas based electricity: The concept of a biogas based energy systems is to some degree similar to biomass based energy systems. Biogas can be produced from livestock manure and human sewage. Farms where animal graze and sewage plants are ideal places to produce energy from biogas.

A biogas based electricity generation system consists of a digester, a biogas collection tank, a generator as well as the piping and controls required for successful operation. The biogas is produced in the anaerobic digester, where anaerobic fermentation takes place, which is provided everyday with livestock manure in the form of cattle dung. Anaerobic fermentation is the fermentation in the absence of air of cellulose containing organic materials, like cattle dung, poultry droppings etc. This oxygen deficient fermentation results in the production of combustible gas called biogas which contains 50-60 % methane, 30-40 % carbon dioxide, 1-5 % hydrogen and traces of nitrogen etc. This biogas can be used to generate power using a diesel engine with diesel as pilot fuel and biogas as main fuel. This mode of operation is known as dual fuel mode.

- Mathematical model: The annual delivered electricity output (E_{Annual}) of a biogas based energy system with rated power output (P_{BGG}) of electricity generator is dependent on its capacity utilization factor (CUF). Assuming a 27 % conversion efficiency from biogas to electricity, it can be modelled using the following expression:

$$E_{Annual} = P_{BGG}\ (8760 \times CUF) \quad (4.1)$$

And hourly energy output (E_{BGG}) is given by

$$E_{BGG}(t) = P_{BGG}(t) \quad (4.2)$$

D. Mathematical model of SPV Generator

- PV electricity: Solar photovoltaic (SPV) technology involves the direct conversion of sunlight into electricity through the use of photovoltaic modules. PV cells are composed mainly of silicon (Si). When PV cells are joined physically and electrically and placed into a frame they form a solar panel or PV module. Panels joined together form a solar array.
- Mathematical model: The sunlight impinging on panels, i.e. irradiance or insolation (incoming solar radiation), is measured in units of watts per square meter (W/m^2). The PV system power output (DC) has approximately a linear relationship to the insolation. Using the solar radiation available on the tilted surface the hourly energy output of the PV generator, can be calculated according to following equation:

$$E_{PVG}(t) = G(t) \times A \times \eta_{PVG} \quad (5.1)$$

Where, G (t) = irradiance in kWh/m^2
A = surface area of the PV modules in m^2
E_{PVG} (t) = hourly energy output from PV in kWh.

All the energy losses in a PV generator, including connection losses, wiring losses and other losses, are assumed to be zero. Eq. 5.1 assumes that PV generator has a tracking system and a maximum power point tracker (i.e. $\eta = 1$). It also assumes that the temperature effects (on PV cells) are ignored.

E. Mathematical Model of Diesel Generator

Conventional generators are normally diesel engines coupled to generator. Diesel generators supply energy in one of two ways. Either they generate only the power

needed by the load (load following), or they generate at nominal power and the surplus energy (if any), is used to charge the battery bank. In this study, a diesel generator of both kinds was considered. The generator model is designed in such a way that the diesel generator is always operating between 80-100 % of their kW rating, While operating in conjunction with the battery bank and other renewable generators. Energy generated by diesel generator in an hour t is defined by the following expression

$$E_{DEG}(t) = P_{DEG}(t) \times \eta_{DEG} \quad (6.1)$$

Where, $E_{DEG}(t)$ = Energy generated by diesel generator, $P_{DEG}(t)$ = Diesel generator capacity, kW.
η_{DEG} = Diesel generator efficiency.

F. Mathematical Model of Battery Bank

Several mathematical models for lead-acid batteries were studied. Some of these models are dynamic and others are static. The battery state of charge (SOC) is the cumulative sum of the daily charge/discharge transfers. At any hour t the state of battery is related to the previous state of charge and to the energy production and consumption situation of the system during the time from t–1 to t. During the charging process, when the total output of all generators is greater than the load demand, the available battery bank capacity at hour t can be described by

$$E_{BATT}(t) = E_{BATT}(t\text{-}1) + [E_{SUR\text{-}AC}(t) + E_{SUR\text{-}DC}(t)] \times \eta_{CHG} \quad (7.1)$$

$$E_{SUR\text{-}AC}(t) = [E_{MHG}(t) + E_{BGG}(t) + E_{BMG}(t) + E_{DEG}(t) - E_{LOAD}(t)] \times \eta_{REC} \times \eta_{CC} \quad (7.2)$$

$$E_{SUR\text{-}DC}(t) = [E_{PVG}(t) - (E_{LOAD}(t) - E_{MHG}(t) - E_{BGG}(t) - E_{BMG}(t) - E_{DEG}(t))/\eta_{INV}] \times \eta_{CC} \quad (7.3)$$

On the other hand, when the load demand is greater then the available energy generated, the battery bank is in discharging state. Therefore, the available battery capacity at hour t can be expressed as:

$$E_{BATT}(t) = E_{BATT}(t\text{-}1) - E_{NETLOAD}(t)/\eta_{INV} \quad (7.4)$$

Where, $E_{BATT}(t)$ = Energy stored in battery at hour t.
$E_{BATT}(t\text{-}1)$ = previously stored energy in battery.
$E_{SUR\text{-}AC}(t)$ = Amount of surplus energy from AC sources.
$E_{SUR\text{-}DC}(t)$ = Amount of surplus energy from DC sources.
η_{CHG} = Battery charging efficiency.
η_{CC} = Charge Controller efficiency.

Meanwhile, the charged quantity of the battery is subject to the following constraints:

$$SOC_{min} \leq SOC(t) \leq SOC_{max}$$

The maximum value of SOC is 1, and the minimum SOC is determined by maximum depth of discharge (DOD),

$$SOC_{min} = 1\text{-}DOD$$

G. Mathematical Model of Charge Controller

To prevent overcharging of a battery, a charge controller is used to sense when the batteries are fully charged and to stop or decrease the amount of energy flowing from the energy source to the batteries. The modelling of the charge controller is presented below:

$$E_{OUT}(t) = E_{IN}(t) \times \eta_{CC} \quad (8.1)$$

Where,
$E_{OUT}(t)$ = Energy output from charge controller, kWh.
$E_{IN}(t)$ = Energy input to charge controller, kWh.

H. Mathematical Model of Inverter

The photovoltaic generator and battery produce DC power and therefore when the hybrid energy system contains an AC load, a DC/AC conversion is required. This is the reason why this section presents the inverter model. The inverter is characterized by a power dependent efficiency. The role of the inverter is to keep on the AC side the voltage constant at the rated voltage 230 volt and to convert the input power in the output power with the best possible efficiency. The inverter model for photovoltaic generator and battery bank are given below:

$$E_{INV}(t) = E_{PVG}(t) \times \eta_{INV} \quad (9.1)$$

$$E_{INV}(t) = E_{BATT}(t) \times \eta_{INV} \quad (9.2)$$

Where, $E_{INV}(t)$ = Energy output from Inverter, kWh
$E_{BATT}(t)$ = Energy discharged by battery, kWh.

I. Mathematical Model of Dump Load

The dump energy is defined as the energy produced by the renewable generators or diesel generator but unused when the load does not need all the energy and the battery has reached its maximum capacity and can not store more energy. In this study, a conventional electric water heater is assumed as dump load. The dump energy, defined as the energy produced and not used by the system, for hour t is calculated as follows:

$$D(t) = [E_{SUR\text{-}AC}(t) + E_{SUR\text{-}DC}(t)] - (E_{BATMAX} - E_{BATT}(t\text{-}1)/\eta_{CHG})]$$

$$E_{SUR\text{-}AC}(t) = [E_{MHG}(t) + E_{BGG}(t) + E_{BMG}(t) + E_{DEG}(t) - E_{LOAD}(t)] \times \eta_{INV} \times \eta_{CC} \quad (10.1)$$

$$E_{SUR\text{-}DC}(t) = [E_{PVG}(t) - (E_{LOAD}(t) - E_{MHG}(t) - E_{BGG}(t) - E_{BMG}(t) - E_{DEG}(t))/\eta_{INV}] \times \eta_{CC} \quad (10.2)$$

Where, D (t) = Total dump energy at time t, kWh.
E_{BATMAX} = Maximum capacity of battery, kWh.

VI. Development of Model

This section describes an optimization technique developed for the design of stand-alone hybrid energy systems. The stand-alone hybrid energy system in this research serves the isolated demand in clusters of remote villages. The design takes into account the unit cost of different resources based on life cycle costs. The results are optimal suggestion for optimum configurations based on locally available energy resources. The problem is formulated as a mixed-integer linear programming developed in C++ [5-8].

A. Assumptions

In order to state a model which is both sufficiently general and accurate for describing all types of energy flow, we make the following assumptions:

- We consider the system in steady state.
- We consider the steady state power, efficiency, and energy only, no other values are used for the system description.

- The model incorporates conservation laws (e.g. conservation of energy flow) but no constitutional laws (e.g. relation between voltage and current).
- It is assumed that only AC appliances are used and are connected to the load bus.
- A time horizon of one hour is used throughout this study.

The unit of measurement for power is kW and that for electrical energy in kWh. It is assumed that all the system components shown in Fig. 1 are installed at a specific site.

B. Objective Function

The objective function to determine the optimum cost of a hybrid energy system is illustrated in expression given below:

Minimize:

$$TC = \sum_{j=1}^{6} \sum_{d=1}^{dn} \sum_{t=1}^{24} [C_j \times E_{jdt}] \quad (11.1)$$

Where, TC is the total optimized cost of providing energy, for all end uses for operation of the system; C_j cost/unit of the jth generating unit (Rs/kWh); E_{jdt} optimal amount of the energy of the generating unit j for end use in a day d, hour t for a particular month; dn is number of day depending upon a particular month; and j is a type of energy source. E_{jdt} is a continuous decision variable in the objective function.

C. Integer variable restriction

Integer variables are used for representing the operation or otherwise of the renewable (RG), conventional generator (DEG) and battery bank (BB). This is given by:

$$X_j = \begin{cases} 1 & \text{If unit j serves the load directly} \\ 0 & \text{Otherwise} \end{cases}$$

Where, j = MHG, or BGG, or BMG, or PVG, or DEG, or BB.

D. General constraints

1. Energy balance constraints: The hourly energy demand must be satisfied by the amount of energy generated from all distributed generation units. Mathematically, it is given by the following expression for all day d, hour t for a particular month.

$$E_{PVG}(t) + E_{MHG}(t) + E_{BGG}(t) + E_{BMG}(t) + E_{DEG}(t) + E_{BATT}(t) \geq E_{LOAD}(t) \quad (11.2)$$

3. Monthly generation constraints: For all generating units, their monthly generation may not exceed their availability.

4. Individual capacity constraints: For all type of generators, their hourly outputs are limited by their total generating capacity. For all hours t

$$E_{jdt} \leq P_j$$

5. Unit generation limit: Each generation unit has its maximum and minimum generation limit. Therefore, in order to avoid damages, the generation of each unit must meet constraint as follows:

$$P_{j\,min} \leq P_j(t) \leq P_{j\,max}$$

6. Non-negativity constraints: Since electrical energy flow and power flow can not be negative in the solution. Hence, all the decision variables are non-negative.

E. Decision variables and constraints associated with Photovoltaic Generator (PVG)

Decision variables of photovoltaic generator are $X_{j=PVG}$, and $E_{jdt,\ j=PVG}$. The former is an integer decision variable representing a decision to select or not select a photovoltaic generator in an hour t; where as the latter is a continuous decision variable representing power generation from photovoltaic generator in day d, and hour t of a particular month. The relationship between energy output from photovoltaic generator and insolation (G) is shown in eq. 11.3 & 11.4 respectively.

$$E_{\underset{j=PVG}{jdt}}(t) = E_{\underset{j=PVG}{jdt,\ Load}}(t) + E_{\underset{j=PVG}{jdt,\ BATT}}(t) = P_{j=PVG}(t) \times \eta_{INV} \times X_{j=PVG} \quad (11.3)$$

$$P_{j=PVG}(t) = G(t) \times A \times \eta_{PVG} \quad (11.4)$$

Where, $P_{j\ =\ PVG}(t)$ is the power available from photovoltaic generator in hour t; G (t) is the solar radiation in day d, and hour t for a particular month; and the last factor is the integer decision variable.

F. Decision variables and constraints associated with Micro hydro Generator (MHG)

Decision variables of micro hydro generator are $X_{j=MHG}$, and $E_{jdt,\ j=MHG}$. The former is an integer decision variable representing a decision to select or not select micro hydro generator. The latter is a continuous decision variable, representing power generation from the MHG in a day d, hour t for a particular month. The following equations represent the generation characteristic of MHG. They imply that the power generation from the MHG at any hour t can take the value at its maximum generation capacity (which is its rated power), if the MHG is selected. Where $P_{j=MHG}$ is the rated power of MHG.

$$E_{\underset{j=MHG}{jdt}}(t) = E_{\underset{j=MHG}{jdt,\ Load}}(t) + E_{\underset{j=MHG}{jdt,\ BATT}}(t) \leq P_{j=MHG}(t) \times X_{j=MHG} \quad (11.5)$$

G. Decision variables and constraints associated with Biogas Generator (BGG)

Decision variables of biogas generator are $X_{j\ =\ BGG}$, and $E_{jdt,\ j\ =\ BGG}$. The former is an integer decision variable representing a decision to select or not select biogas generator. The latter is a continuous decision variable, representing power generation from the BGG in a day d, hour t for a particular month. The following equations represent the generation characteristic of BGG. They imply that the power generation from the BGG at any hour t can take the value at its maximum generation capacity (which is its rated power), if the BGG is selected. Where $P_{j\ =\ BGG}$ is the rated power of BGG.

$$E_{\underset{j=BGG}{jdt}}(t) = E_{\underset{j=BGG}{jdt,\ Load}}(t) + E_{\underset{j=BGG}{jdt,\ BATT}}(t) \leq P_{j=BGG}(t) \times X_{j=BGG} \quad (11.6)$$

H. Decision variables and constraints associated with Biomass Generator (BMG)

Decision variables of biomass generator are $X_{j\ =\ BMG}$, and $E_{jdt,\ j\ =\ BMG}$. The former is an integer decision variable representing a decision to select or not select biomass. The latter is a continuous decision variable, representing power generation from the BMG in a day d, hour t for a particular month. The following equations represent the generation

characteristic of BMG. They imply that the power generation from the BMG at any hour t can take the value at its maximum generation capacity (which is its rated power), if the BMG is selected. Where $P_{j = BMG}$ is the rated power of BMG.

$$E_{\substack{jdt \\ j=BMG}}(t) = E_{\substack{jdt,\ Load \\ j=BMG}}(t) + E_{\substack{jdt,\ BATT \\ j=BMG}}(t) \le P_{j=BMG}(t) \times X_{j=BMG} \quad (11.7)$$

I. Decision variables and constraints associated with Diesel Engine Generator (DEG)

Decision variables of DEG are $X_{j = DEG}$, and $E_{jdt,\ j = DEG}$. The former is an integer decision variable representing a decision to select or not select DEG. The latter is a continuous decision variable, representing power generation from the DEG in a day d, hour t for a particular month. The following equations represent the generation characteristic of DEG. They imply that the power generation from the DEG at any hour t can take the value zero, or any value between its minimum generation (which is assumed to be 80% of its rated power), and its maximum generation (which is its rated power), if the DEG is selected. Where $P_{j = DEG}$ is the rated power of DEG.

$$E_{\substack{jdt \\ j=DEG}}(t) = E_{\substack{jdt,\ Load \\ j=DEG}}(t) + E_{\substack{jdt,\ BATT \\ j=DEG}}(t) \le P_{j=DEG}(t) \times X_{j=DEG} \quad (11.8)$$

$$E_{\substack{jdt \\ j=DEG}}(t) = E_{\substack{jdt,\ Load \\ j=DEG}}(t) + E_{\substack{jdt,\ BATT \\ j=DEG}}(t) = 0 \quad \text{or}$$

$$E_{\substack{jdt \\ j=DEG}}(t) = E_{\substack{jdt,\ Load \\ j=DEG}}(t) + E_{\substack{jdt,\ BATT \\ j=DEG}}(t) = [0.8\ P_{j=DEG}(t),\ 1.0\ P_{j=DEG}(t)]$$

J. Decision variables and constraints associated with Battery bank (BB)

Decision variables of battery bank (BB) are $X_{j = BATT}$, and $E_{jdt,\ j\ =\ BATT}$. The former is an integer decision variable representing a decision to select or not select battery. The latter is a continuous decision variable, representing discharging from the battery in a day d, hour t for a particular month. The following equations represent the generation characteristic of battery bank. They imply that the discharging from the battery at any hour t can take the value zero, or any value between its minimum discharge capacity (which is assumed to be 20 % of its rated capacity), and its maximum discharge capacity (which is its battery rated capacity), if the battery is selected. Where $P_{j = Batt}$ is the rated capacity of battery.

$$E_{\substack{jdt \\ j=BATT}}(t) = E_{\substack{jdt,\ Load \\ j=BATT}}(t)/\eta_{INV} \le \eta_{INV} \times P_{j=BATT}(t) \times X_{j=BATT} \quad (11.9)$$

$$E_{\substack{jdt \\ j=BATT}}(t) = E_{\substack{jdt,\ Load \\ j=BATT}}(t)/\eta_{INV} = 0 \quad \text{or}$$

$$E_{\substack{jdt \\ j=BATT}}(t) = E_{\substack{jdt,\ Load \\ j=BATT}}(t)/\eta_{INV} = \eta_{INV} \times [0.2\ P_{j=BATT}(t),\ 1.0\ P_{j=BATT}(t)]$$

This research assumes that SOC_{min} and SOC_{max} equal 20 % and 100 % of the battery capacity, respectively. It is also assumed that the initial SOC of the battery is 100 % at the beginning of the simulation. At any hour t

$$SOC(t) \le SOC_{max}$$
$$SOC(t) \ge SOC_{min}$$

Lastly, in order for the system with battery to be sustained over a long period of time, the battery SOC at the end of the day must be greater than a percentage of its SOC_{max}. This study assumes 80 % as shown below:

$$SOC(t = 24) \ge 0.8\ SOC_{max}$$

VII. Conclusions

Taking into consideration the scientific interest concerning the capabilities of hybrid energy systems with energy storage to fulfil the electrification needs of a rural remote area, a detailed mathematical model of describing the operational behaviour of the basic hybrid energy system components is presented.

The proposed model employs generalized integer linear programming to determine the optimum unit cost and operation of the hybrid energy system with a storage facility, using hourly, daily and monthly load demand. The model is shown to be sufficiently accurate. It uses the original load curves instead of the load duration curves, accurately reflecting the time dependency of the storage operation policies over a year.

The proposed model can be used in planning studies to determine the optimum design of a autonomous hybrid energy system.

References

[1] R. Ramakumar, N. G. Butler, and A. P. Podriguez, "Economic aspects of advanced energy technologies," in *Proc. 1993 of IEEE* vol. 81, No. 3, pp. 318-332, March 1993.

[2] R. Ramakumar, I. Abouzahr, and K. Asenyayi, "A knowledge-based approach to the design of integrated renewable energy systems," *IEEE Trans. on Energy Conversion,* vol. 7, No. 4, pp. 648-657, 1992.

[3] S. Jebaraj and S. Iniyan, "A review of energy models," *Renewable & Sustainable Energy Reviews*, vol. 10, pp. 281-311, 2006.

[4] R. B. Hiremath, S. Shikha and N. H. Ravindranath, "Decentralized energy planning; modeling and application," *Renewable & Sustainable Energy Reviews*, vol. 11, pp. 729-752, 2007.

[5] M. Chen, R. Atta-Konadu, "Mathematical programming model for energy system design," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Vol. 19, No. 8, pp. 789-801, March 1997.

[6] A. Arivalagan, B. G. Raghavendra, and A. R. K. Rao, "Integrated energy optimization model for a cogeneration based energy supply system in the process industry," *Electrical Power and Energy Systems*, Vol. 17, No. 4, pp. 227-233, 1995.

[7] M. Pipattanasomporn, "A study of remote area internet access with embedded power generation," Ph.D. dissertation, Dept. Elec. Engg., Virginia Polytechnic Institute and State Univ., Alexandria, Virginia, 2004.

Biographies

Ajai Gupta was born in Bareilly, India. He received the B. Tech in Electrical Engineering from I.E.T. Rohilkhand University, Bareilly in 2000 and M. Tech in Instrumentation and Control from Aligarh Muslim University, Aligarh in 2004 respectively. Currently he is a Research Scholar at Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee, India.

Dr. R. P. Saini is serving as an Associate Professor at A.H.E.C, Indian Institute of Technology, Roorkee, India.

Dr. M. P. Sharma has been working as an Associate Professor at Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee, India since the last 20 year.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Computerized Modelling of Hybrid Energy System—Part II: Combined Dispatch Strategies and Solution Algorithm

Ajai Gupta, R. P. Saini, and M. P. Sharma

Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee
Roorkee, Uttarakhand-247667, India
E-mail: ajai_abjc@yahoo.com (Ajai Gupta)

Abstract – Computer simulation is an increasingly popular tool for determining the most suitable hybrid energy system type, design and control for an isolated community or a cluster of villages. This paper presents the development of the optimum control algorithm based on combined dispatch strategies, to achieve the optimal cost of battery incorporated hybrid energy system for electricity generation, during a period of time by solving the mathematical model, which was developed in Part I. The main purpose of the control system proposed here is to reduce, as much as possible, the participation of the diesel generator in the electricity generation process, taking the maximum advantage of the renewable sources available. The overall load dispatch scenario is controlled by the availability of renewable power, total system load demand, diesel generator operational constraints and the proper management of the battery bank. The incorporation of a battery bank makes the control operation more practical and relatively easier.

***Index terms* – Combined dispatch strategy, Load Following Strategy, Cycle Charging Strategy, Algorithm.**

I. Introduction

One of the most important aspects of hybrid energy system is to supply power to the customers economically. The problem of deciding how power supplies are shared among generators in a system in the most economic manner has been studied extensively and various mathematical programming methods and optimisation techniques have been developed and applied [1].

When designing a hybrid energy system, both the sizing of the elements and the most adequate control strategy must be obtained. Obtaining a good control strategy is essential, since the performance of a hybrid energy system can be significantly affected by relatively small changes made in the control strategy.

As mentioned before, a very important aspect when designing a hybrid energy system is the determination of the most adequate control strategy. Numerous studies have developed control strategy applicable to hybrid energy systems. In [2], several control strategies for PV-Diesel systems including batteries are described. Likewise, the HOMER optimization model [3] uses relatively simple strategies based on the ones studied by Barley et. al. and it is able to obtain an optimal design of a hybrid energy system by selecting the most appropriate strategy. Hybrid 2 [4], outperforms HOMER and generates many control strategies; nevertheless, it is a simulation program, not an optimization tool.

M. Ashari and C. V. Nayar presents dispatch strategies for the operation of a solar photovoltaic (PV)-diesel-battery hybrid power system using 'set-point' [5]. A novel strategy, optimized by genetic algorithm, to control stand-alone hybrid renewable electrical systems with hydrogen storage has been proposed by R. Dufo-Lopez et al. [6].

The cost of hybrid energy system depends on two factors: the size of the individual component and the dispatch strategy. In order to predict the cost and the operation of systems, researchers have developed computer programs for simulation and analysis based on different preferences and approach.

Hence, this paper presents a combined dispatch strategies based control algorithm (computer program) for the operation of a battery incorporated hybrid energy system for determination of the optimum unit cost (with high efficiency of diesel generator) of system, while taking into account the component and system constraints.

II. Concept and Dispatch Strategies

A. Dispatch strategy concept

The control system is one of the most important elements of any system. As their name indicates it is entrusted with controlling to the rest of the components through the inputs and outputs control. The control system is necessary for energy sources selection, battery working level, protections etc.

There are two distinct levels in the control of a hybrid energy system:

- Dynamic control, which deals with control of the frequency and magnitude of the output voltage of the system, and
- Dispatch control, which deals with the flow of energy in the system from the various sources to load.

The dispatch strategy for a hybrid energy system is a control algorithm for the interaction among various system components. The control strategy determines the energy flows from the various sources like diesel generator and different types of renewable generators, towards the user

loads, and dump load, including the charging and discharging of the energy storage systems, on a time scale of minutes to hours, in such a way as to optimize system performance in terms of operating cost.

B. Simple dispatch strategies

The various dispatch strategies that are modelled in this study are described [7]:

1) Battery charging strategy: When the power generated from renewable generators exceeds the load demand, the excess energy will charge the battery bank, provided they are not already fully charged. Opportunities for this type of charging occur frequently in systems located in high solar penetration regions.

The other possibility is to use the diesel generator to charge the battery bank. This technique is known as "cycle charging". The battery charging continues until:

- Maximum battery SOC (or 80 % of rated capacity) is reached.
- The renewable power is not sufficient to meet the load.

2) Battery discharge strategy: Battery power may be used to meet the netload in a time step. Diesel power is not used to charge the batteries; the diesel operating point is set to match the instantaneous netload. The discharging of the battery is continued until the minimum battery SOC for discharge is reached. The battery discharging continues until:

- Minimum battery SOC for discharge is reached.
- The renewable power is sufficient to meet the load.
- The renewable power is sufficient to meet the load as well as continue to charge the batteries.
- Netload (or netload + battery load) is equal to or greater than diesel minimum operating power.

3) Load-Following strategy: Battery power is never used to meet the netload in a time step. Diesel power is not used to charge the batteries; the diesel operating point is set to match the instantaneous load. However, if excess power is available because of prescribed diesel minimum operating power, this power will charge the batteries. A minimum diesel run time may also be applied to avoid excessive start/stop frequency.

4) Cycle charge strategy: Diesel power may be used to meet the netload in a time step, whenever the diesel runs, it runs at full power (or as high as possible without dumping excess power); the difference between the diesel power and the netload is stored in the batteries. The diesel continues running for its prescribed minimum run time; after that, the diesel continues running until one of the condition is met:

- The prescribed SOC set point has been met, or
- The renewable power is sufficient to meet the load.
- The renewable power is sufficient to meet the load as well as continue to charge the batteries.

5) Peak shaving strategy: Battery power is never used to meet the netload within a time step; it is only used to buffer instantaneous fluctuations around the netload. Diesel power is not used to charge the batteries, unless excess power is available because of prescribed diesel minimum operating power. Renewable power in excess of the load will charge the batteries.

III. Proposed Combined Dispatch Strategy

In a system that has different renewable energy sources, the output from these renewable energy generators is generally subtracted from the hourly demand to determine the hourly net load. By observing hourly operation of proposed hybrid energy system (see part 1), there are five possible dispatch strategies to meet the net load [6, 8]. If net load is zero or negative in a particular hour then battery charging strategy is used to absorb the surplus power (all or a fraction), generated by renewable generators. Alternatively, if netload is positive, then all five dispatch strategies are used to operate and control the system. The summarized strategies are:

a) Battery charging strategy: the use of only the battery to absorb the surplus power.
b) Battery discharging strategy: the use of only the battery to cover the load demand.
c) Load following strategy: the diesel generator is ON to follow the load, no charge to the battery and no discharge from the battery.
d) Cycle charging strategy: the diesel generator is ON to cover the load demand and charge the battery.
e) Peak shaving strategy: the use of battery to shave the peak demand.

The combination of these strategies results some conflicting objectives of the system operation as follows:

1. Minimize the system unit cost.
2. Keeping the output of the diesel generator constant with high efficiency.
3. Minimize the fuel consumption of the diesel generator.
4. Minimize the frequency of diesel generator starts/stops.
5. Maximize the utilization rate of the renewable energy resources.

Let the power generated in the system at the time t, P (t), can be expressed as follows:

$$P(t) = P_{MHG}(t) + P_{BGG}(t) + P_{BMG}(t) + P_{PVG}(t) \quad (1)$$

Where, $P_{BGG}(t)$ = power generated by biogas generator.
$P_{MHG}(t)$ = power generated by micro-hydro generator.
$P_{BMG}(t)$ = power generated by biomass generator.
$P_{PVG}(t)$ = power generated by photovoltaic generator.
P (t) = total power generated by renewable generators at time t.

And the hourly load demand remaining for diesel generators & battery bank is given by

$$NL(t) = L(t) - P(t) \quad (2)$$

Where, L (t) = Total hourly load demand at time t.
NL (t) = Netload demand at time t.

This system has three conditions of operation to supply the netload, which are given below:

A. Condition-1 [P (t) ≥ L (t)]

When the power generated from renewable generators exceeds the load demand for the system, the excess energy will charge the battery bank (provided they are not already fully charged). Opportunities for this type of charging

occur frequently in system with high solar insolation (battery charging strategy).

There are two possibilities in this condition. In the first case, the battery has not reached the maximum capacity and the excess power will charge the battery. In other case, the excess power (all or some fraction) will be lost because the battery bank is fully charges.

B. Condition-2 [P (t) < L (t)] & [NL (t) < P_{DEG} (t)]

When the power generated from renewable generators are lower than their corresponding power demands, then first battery bank will be discharged (because unit cost of battery bank is lower than the unit cost of diesel generator) by the amount necessary to cover the netload and the diesel generator turns off (battery discharge strategy).

If the batteries are not able to supply such a netload, or the strategy does not allow it, then the netload is met by diesel generator. There are two possible dispatch strategies in this condition. These are the cycle charging strategy and the load following strategy. The cycle charging strategy, which is discussed above, is generally used in combination with the load following strategy. The decision to use one or the other strategy, i.e., the decision to generate power equal to the netload (load following) or greater than the netload to charge the battery (cycle charging), can be arbitrary selected. When the netload on the diesel generator is low, the diesel output can be increased, with the excess power produced being used to charge the storage batteries. The advantages of cycle charging are that it allows the diesel generator to operate at higher power output, where it operates more efficiently; and it allows for longer diesel shutdown periods, thus reducing the rotating cost associated with the diesel generators.

C. Condition-3 [P (t) < L (t)] & [NL (t) > P_{DEG} (t)]

If the netload exceeds the diesel generator rated capacity, the diesel generator will run at full power and batteries will attempt to contribute the difference. This type of simultaneous operation of both battery and diesel generator is known as peak shaving strategy.

IV. Implemented Dispatch Algorithm

A. Assumptions

1. It is assumed that the shape of load curve does not change throughout the planning period.
2. Each renewable generator is assumed to have two possible states: (a) zero outage level (generator is running at full capacity), (b) full outage level (generator is out).
3. (a) Loading of renewable generators is done according to economic merit order, that is, generators are loaded in order of increasing unit generation cost. (b) If the renewable generators are not able to supply such an hourly total demand, then battery bank and diesel generator will be loaded according to economic merit order.
4. It is assumed that the available technologies at the beginning of planning horizon will remain available throughout the planning horizon.

B. Database

The essential data for running the simulation algorithm are collected, organised, and stored in a database. The information stored in the database includes:

- Available renewable energy generator power production capacity.
- Hourly demand load profile of the cluster of villages.
- Diesel generator capacity.
- Battery bank capacity.
- Unit cost of generation of different energy sources.
- Minimum and maximum SOC of battery bank.

C. Constraints

1. Diesel generator constraint:
   - Diesel generator must be operated within the minimum and maximum allowable rating.
   - Diesel generator must be operated above 80 % of the rated capacity as the efficiency of the generator is low at low load.
   - Diesel generator must be turn off at t = 24 to minimise noise in the early hours of the day by meeting the load using battery as much as possible.
2. Battery storage constraints:
   - Battery must be operated between 20 % - 100 % of the capacity.
   - Battery must be charged at least 80 % of maximum capacity at the end of the day. So that it can be operated in the early hours of the next day.
3. Renewable energy source constraints:
   - Renewable energy should be used as much as possible.

D. Optimal economic operation strategy

A computer program was developed to determine the optimum unit cost of hybrid energy system based on energy balance and combined dispatch strategy during a day and summed to give the overall performance over a month. Each day is divided into a 24 one hour segments, with the energy generated and consumed during each segment, taking into account both economic and technical factors.

Fig. 1 through 5 describes the computer algorithm to simulate the operation of the desired hybrid energy system. The following items summarize the key characteristics of the implemented strategy:

According to Fig. 1

1. Renewable power (MHG, BGG, BMG, and PVG) is first sent to meet the residential load demands in order of economic merit. Energy output of each renewable generator is calculated according to equations developed in Part-I (see system components modelling section). If the power supply from renewable generators are smaller than the load demand, then supply the load equal to maximum operating power from renewable generators (RG) and go to step 3. Alternatively, if the power supply from renewable generators is larger than the power demand, then supply the load and go to step 2.

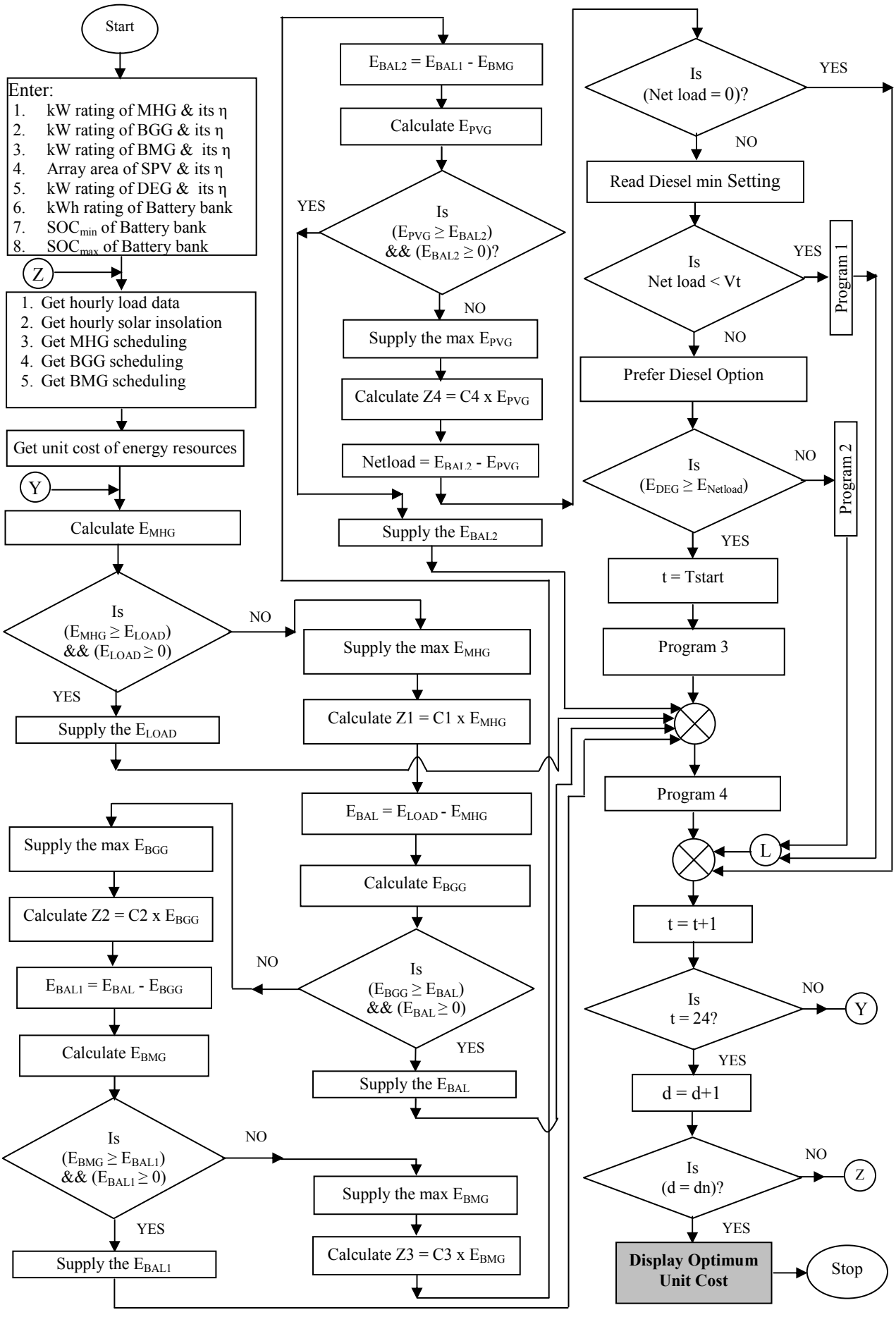


Fig. 1 Flow chart for Main Program

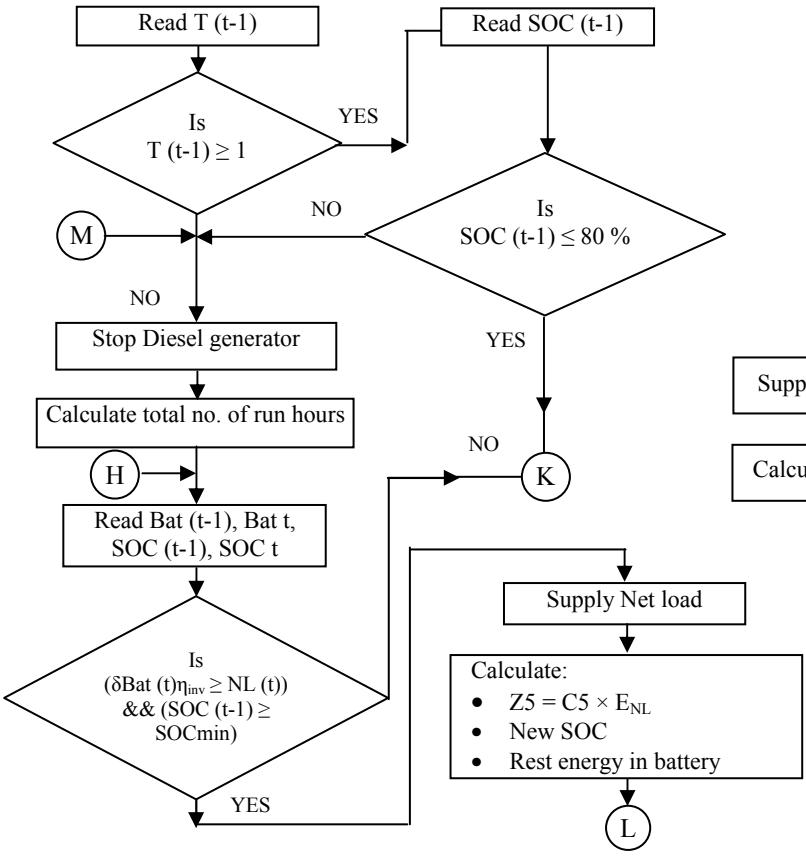


Fig. 2 Flow chart for program 1

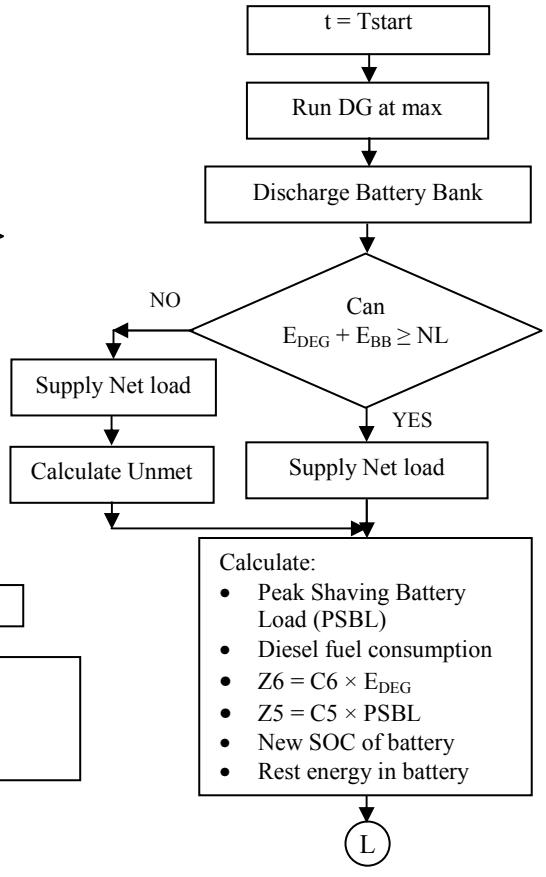


Fig. 3 Flow chart for program 2

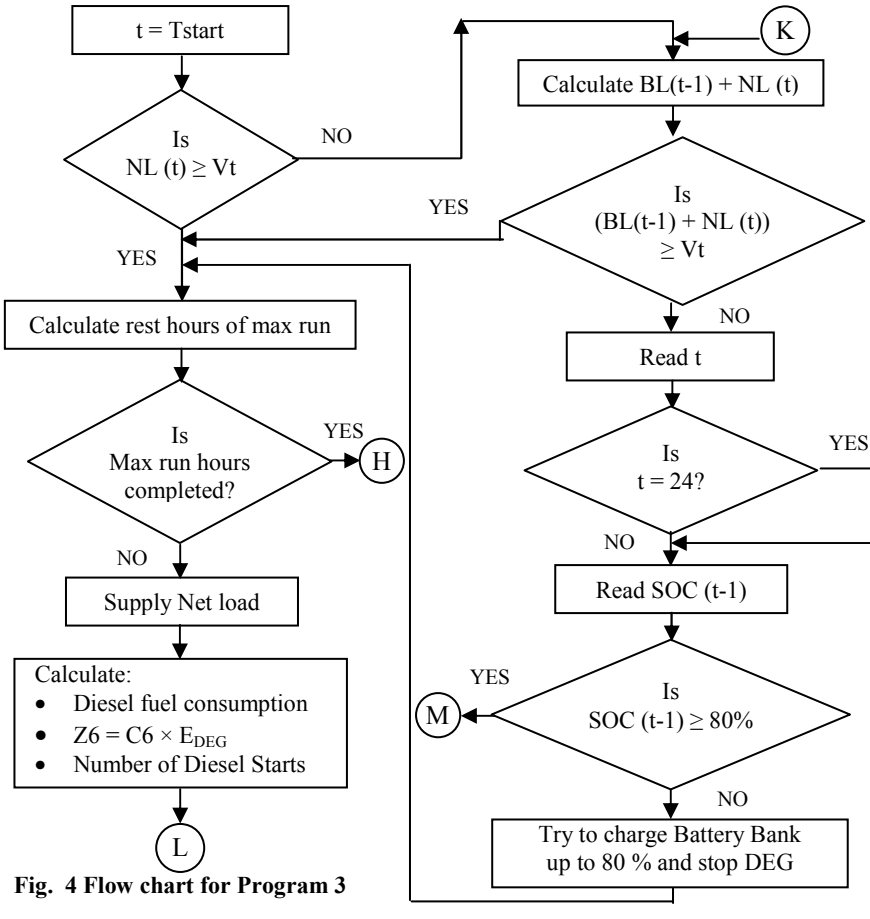


Fig. 4 Flow chart for Program 3

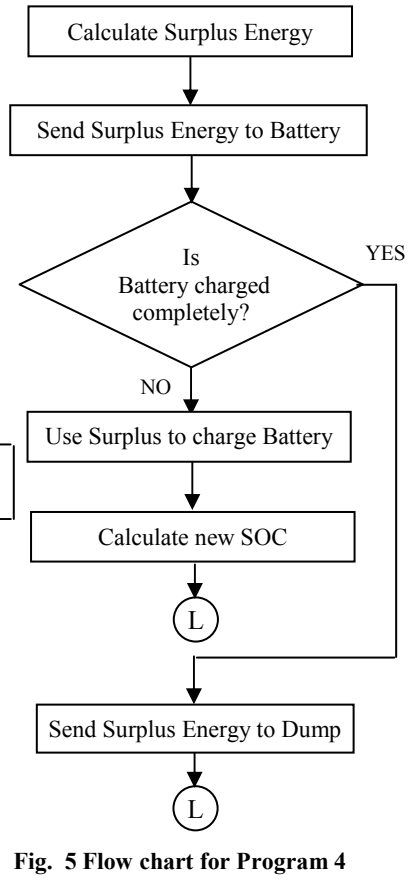


Fig. 5 Flow chart for Program 4

2. Charge all surplus power in battery. Here, if battery becomes full, charge surplus power until battery become full and remaining surplus power is thrown out (goes to dump load). Go to step 8.

3. Energy output from renewable generators is treated as a negative load. So the netload can be determined by the following expression:

Netload = Actual load – renewable generators output

If the netload is equal to zero, then go to step 8. Alternatively, if netload is positive, then go to step 4.

4. If the system has battery storage and diesel generator both, the operation characteristic will be divided according to time that the diesel generator starts or stops. The following description explains the operation of the diesel generator during t = 0 to t = T_{start}; during t = T_{start} to T_{stop}; and during t = T_{stop} to t = 24 of each day.

5. During t = 0 to t = T_{start} of each day: If netload is less than diesel generator minimum operating power, then the system is operated only on the battery.

a) Energy discharged from the battery must be enough to serve hourly demand, i.e. netload/η_{inv}, but not be greater than the maximum allowable discharged energy from the battery (charge controller).

b) Unmet energy is zero if the discharge energy from the battery can cover the netload.

c) If the battery is not sufficient to meet the netload and diesel generator maximum running hours has completed, then there will be a portion of demand unmet.

d) Determine the SOC of the battery at any hour t.

e) No energy is generated from diesel generator during this period. However, whenever SOC reaches SOC_{min}, the diesel generator must be started in order to cover the demand and charge the battery (i.e. go to step 6), otherwise go to step 8.

6. During t = T_{start} to t = T_{stop} of each day: This section is divided into four parts which are given below

a) If diesel generated power is greater than netload and netload is greater than diesel generator minimum operating power, then the system is operated on the diesel generator to cover the netload and to charge the battery. Go to step 8.

b) If diesel generated power is greater than netload and netload is smaller than diesel generator minimum operating power (Vt), then check:

Netload (NL) + battery load (BL) ≥ Vt

If yes, then go to step 6, otherwise check:

at t = 24, SOC < 80 %

If yes, then go to step 6, otherwise go to step 5.

c) If diesel generated power is smaller then netload, then the system is operated on diesel generator and battery both. This is called peak shaving strategy. Diesel generator operates at maximum operating power to cover the maximum part of netload and rest part of netload is covered by battery. Go to step 8.

d) If diesel generator minimum running hours has completed or SOC of battery has reached up to 80% or more, then stop diesel generator and go to step 7.

7. During t = T_{stop} to t = 24 of each day: The system is operated from the battery, so the algorithm is the same as that during t = 0 to t = T_{start} of each day.

8. If time t does not reach the end of simulation, increase t = t + 1 and go to step 1. If time t reaches end of simulation, go to end.

V. Conclusions

This paper proposed an optimized operation control algorithm of hybrid energy system for a rural remote area in consideration of the fluctuating electric power of photovoltaic system due to the solar radiation. It is based upon the combined dispatch strategies. The algorithm presented is capable of efficiently designing a least-cost village electrification system while the diesel generator keeps the output constant with high efficiency in spite of the fluctuating photovoltaic power. Developed algorithm is modular in nature. The time-series optimized simulation is based on detailed component modelling. It can be extended to include other renewable sources of energy such as wind power and fuel cell. This algorithm may be used by policy makers for evaluation purposes when many power generation schemes are taken into account.

References

[1] K. P. Wong, and C. C. Fung, "Simulated annealing based economic dispatch algorithm," *IEE Proceedings-C,* vol. 140, No. 6, pp. 509-515, November 1992.

[2] C. D. Barley, C. B. Winn, L. Flowers and H. J. Green, "Optimal control of remote hybrid power systems Part 1: Simplified model", in Proc. Wind Power '95, Washington DC, pp. xx–xx, 27-30 March 1995.

[3] Hybrid optimization model for electric renewables (HOMER). Website: www.nrel.gov/international/homer.

[4] A hybrid system simulation model - Hybrid 2. Website: www.ecs.umass.edu/mie/labs/rerl/hy2/.

[5] M. Ashari, and C. V. Nayar, "An optimum dispatch strategy using set points for a photovoltaic-diesel-battery hybrid power system," *Solar Energy*, vol. 66, No. 1, pp. 1-9, 1999.

[6] R. Dufo-Lopez, and J. L. Bernal-Agustin, "Design and control strategies of PV-Diesel systems using genetic algorithms," *Solar Energy*, vol. 79, pp. 33-46, 2005.

[7] C. D. Barley, C. B. Winn, "Optimal dispatch strategy in remote hybrid power systems," *Solar Energy*, vol. 58, No. 4-6, pp. 165-179, 1996.

[8] M. Pipattanasomporn, "A study of remote area internet access with embedded power generation," Ph.D. dissertation, Dept. Elec. Engg., Virginia Polytechnic Institute and State Univ., Alexandria, Virginia, 2004.

Biographies

Ajai Gupta was born in Bareilly, India. He received the B. Tech in Electrical Engineering from I.E.T. Rohilkhand University, Bareilly in 2000 and M. Tech in Instrumentation and Control from Aligarh Muslim University, Aligarh in 2004 respectively. Currently he is a Research Scholar at Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee, India.

Dr. R. P. Saini is serving as an Associate Professor at A.H.E.C, Indian Institute of Technology, Roorkee, India.

Dr. M. P. Sharma has been working as an Associate Professor at A.H.E. C., Indian Institute of Technology, Roorkee, India since the last 20 years.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Computerized Modelling of Hybrid Energy System—Part III: Case Study with Simulation Results

Ajai Gupta, R. P. Saini, and M. P. Sharma

Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee
Roorkee, Uttarakhand-247667, India
E-mail: ajai_abjc@yahoo.com (Ajai Gupta)

Abstract - This paper presents the results of the application of model (developed in Part-I) and simulation algorithm (developed in Part-II) for determining the techno-economics of battery storage type hybrid energy system intended to supply the load demand of a rural remote area having a cluster of nine villages (grid isolated). The hour-by-hour simulation model is intended to simulate a typical one month period of system operation. For simulation purpose, hourly solar insolation data and load demand data have been generated and used as an input data. The economic analysis has resulted in the calculation of optimized hourly, daily, and monthly system unit cost of proposed hybrid energy system. The obtained results represent also a helpful reference for energy planners in Uttarakhand state and justify the consideration of hybrid energy systems more seriously.

***Index terms*-Renewable Energy, Rural households, Off-grid Electrification, System sizing, Economic analysis.**

I. Introduction

Energy is vital for sustaining life on earth expresses the economic stability of a nation. It is needed to improve the quality of life by exploiting the natural resources. The careless exploitation of these resources ultimately affects the environment on which such systems thrive. The energy problem is, thus, synonymous to ecological and economic problems. The efforts should, therefore, need to find an optimal solution for sustainable energy supply.

The world energy demand has been increasing exponentially and conversely. The conventional energy resources are exhaustible and limited in supply. Therefore, there is an urgent need to conserve what we have in hand and explore the wider use of alternative energy resources. Currently, more than half of the world population lives in rural areas in developing countries. The majority of these population make use of fuels like fuel wood, cattle dung cakes, agricultural wastes etc. using inefficient technologies. Thus the basic energy needs consists of heating, lighting and electricity etc are hardly satisfied; all this contributes to maintain the cycle of poverty.

Moreover, this problem is composed of three ways:

- Firstly, one billion people are connected to the electricity network and need an inexpensive technique of grid connection and electric distribution.
- Secondly, another one billion people, being away from utility network, may be energized through integrated renewable energy system/Hybrid energy system operating in decentralized mode.
- Thirdly, five hundred million people are not very far from a power network, and can exercise both the above solution.

The proper management of available energy resources is a must to meet the energy demand in a sustainable manner locally and globally. In many cases, utility grid extension is impractical because of dispersed population, rugged terrain, or both; thus small, stand-alone hybrid energy systems are likely to be the most viable option.

The Hybrid Energy System (HES) has received much attention over the past decade. It is a viable alternative solution as compared to systems, which rely entirely on hydrocarbon fuel. Apart from the mobility of the system, it also has longer life cycle. In particular, the integrated approach [1-2] makes a hybrid system to be the most appropriate for isolated communities of a remote area.

II. Hybrid Energy Systems

Hybrid energy systems generally integrate renewable energy sources with fossil fuel powered diesel/petrol generator to provide electric power where the electricity is either fed directly into the grid or to batteries for energy storage. The role of integrating renewable energy in a hybrid energy system is primarily to save diesel fuel. Examples of renewable energy sources commonly used in hybrid configurations are wind turbines, photovoltaic systems, micro-hydro, biomass, and fuel cells.

A hybrid energy system consists of two or more energy systems, an energy storage system, power conditioning equipment and a controller. A hybrid energy system may or may not be connected to the grid. They are generally independent of large centralized electric grids and are used in rural remote areas.

For systems employing totally clean renewable energy, high capital cost is an important barrier. However, we can produce green power by adding different renewable energy sources to diesel generator and battery, which is also called a hybrid system. This kind of system can compromise investment cost, diesel fuel usage cost and also operation and maintenance costs [3-5].

Hybrid systems for rural electrification can be configured in three different ways: grid connected off-grid with distribution system and off-grid for direct supply. The

first configuration is able to rely on the grid if the hybrid system has problems. Similarly, feeding the power to the grid, thereby, boosting the voltage and minimizing power cuts strengthen the grid. For off-grid configurations, the hybrid can either be connected to many load centres, or can act as a source of supply for one or two loads, thus avoiding the need of a distribution system. An isolated off-grid system is usually used to charge batteries or supply power to small rural industry/households.

III. Hybrid Energy Systems For Village Electrification

To provide energy services to remote areas, three options are available [6]:

i) Hybrid energy system (HES) can be used to augment power from centralized power plant.

ii) Generate fossil fuel based power (e.g. diesel gensets); and

iii) Generate renewable power using hybrid energy systems.

One possible solution that helps to cancel out the drawbacks of diesel and renewable energy technologies is to employ both types in combination, for minimizing costs and maximizing availability.

IV. Methodology

A. Study Area

The remote rural area for the study was Narendra Nagar block of district Tehri Garhwal of Uttarakhand state, India. The block consists of 15 unelectrified villages [7] with 22 hamlets. There are 775 households with a population of 4755 according to the 1991 census. The area comprises of major hilly and the fertile area under forest with scattered households. The area has been considered by Uttarakhand Renewable Energy Development Agency (UREDA) to be remote and not economically viable for electrification by grid extension. The total literacy rate of the Narendra Nagar block is 52%. The data are available in the published statistics [7]. Data regarding several aspects having an important bearing on rural energy planning are not readily available. Hence, survey was conducted for the household energy needs using multi-stage schedules for the present investigation. This survey was conducted during November 2006-April 2007.

Table 1 List of unelectrified villages & general details

S. N.	Village Name	Latitude N & Longitude E	Population	Households
1.	Laga Mehra	30°07', 78°23'	700	60
2.	Saud	30°11', 78°24'	345	65
3.	Salem Khet	30°21', 78°21'	111	17
4.	Talai lambadi	30°12', 78°23'	200	30
5.	Bandhan	30°11', 78°19'	70	12
6.	Pungarh	30°11', 78°18'	62	11
7.	Bhangla	30°11', 78°27'	300	55
8.	Kakhoor	30°10', 78°27'	340	62
9.	Banskata	30°11', 78°29'	1095	190
Total			**3223**	**502**

During survey, it was found that six villages have already electrified by grid extension. The rest nine villages, which have been considered for the present study as the best candidate for electrification by decentralized hybrid energy system consisting of micro-hydro, biogas, biomass, solar photovoltaic, diesel generator and battery. Table 1. shows the details of cluster of nine villages.

B. Assessment of Energy Potential

The study area, though one of the most backward part of Narendra Nagar block, occupies a unique position as far as natural sources are concerned. The study area has adequate sunshine, low to moderate wind speeds, falling water is available 7-8 months in a year. Biomass potential is available in abundance and the animal population of this area is relatively much greater than in other parts of Narendra Nagar block.

- Micro hydro: Based on the published statistics [7], it has found that out of nine villages, only three villages have micro hydro potential. The total potential at these sites has been estimated as 14.2 kW (~15 kW). In order to estimate the hydroelectric generation that could be supplied, we only considered the March-October period by consulting senior citizens of villages. Table 2. shows the details of micro hydro potential.

Table 2 Estimated energy potential of micro hydro power

S. N.	Village Name	Head (m)	Discharge (m^3/sec)	Power (kW)
1.	Talai lambadi	2 (7.00)	0.05	7 (3.5+3.5)
2.	Pungarh	7.00	0.06	4.20
3.	Banskata	5.00	0.06	3.0
Total Power potential				**14.2**

- Solar Energy: Solar radiation data have been taken from the solar radiation data hand book [Mani et al. 1982] at latitude 30°32' N, Longitude 78°03' E [8]. Table 3. shows the daily and monthly global solar radiation. Total solar energy has been taken as 1854.18 kWh/m^2/year.

Table 3 Hourly and Daily Solar Radiation

S. N.	Month	Daily Total (kWh/m^2)	Monthly Total (kWh/m^2)
1.	January	3.58	110.98
2.	February	4.40	123.20
3.	March	5.47	169.57
4.	April	6.35	190.50
5.	May	6.95	215.45
6.	June	6.06	181.80
7.	July	5.25	162.75
8.	August	4.80	148.80
9.	September	5.32	159.60
10.	October	5.13	159.03
11.	November	4.22	126.60
12.	December	3.53	105.90
Annual Total (kWh/m^2)			**1854.18**

- Biogas Energy: To assess the biogas potential, cattle, buffaloes, horse, goats, and cow/ox have been considered. The data on number of cattle is estimated by consulting *Sarpanch* and senior citizens of the villages.

Based upon the survey, it was found that there are 4564 cattle population at study area. The village wise distribution of livestocks is shown in Table 4.

The biogas production from the dung has been evaluated based on the assumption that 10 kg/day dung will be available from cow/ox, 15 kg/day from buffaloes, 1 kg/day from goat, and 10 kg from horse. The cattle dung

availability in the study area is about 25560 kg/day. The biogas estimation is based on the cattle dung production from different types of animals and assuming that 0.036 m^3 of biogas is generated per kg of cattle dung. Therefore, the biogas availability in the study area is about 644.112 m^3/day out of which biogas 510.8796 m^3/day is used for cooking (Thermal load). The balance 133.2323 m^3/day is available for generation of electricity.

Table 4 Details of Livestocks and Biogas Potential

S. N.	Village Name	Total No. of Cattle	Total Dung/ day (kg)	Total Biogas Generated (m^3/day)
1.	Laga Mehra	546	3060	77.112
2.	Saud	591	3310	83.412
3.	Salem Khet	153	850	21.42
4.	Talai lambadi	273	1530	38.556
5.	Bandhan	109	610	15.372
6.	Pungarh	99	550	13.86
7.	Bhangla	501	2810	70.812
8.	Kakhoor	564	3160	79.632
9.	Banskata	1728	9680	243.936
Total		**4564**	**25560**	**644.112**

- Biomass (fuelwood) Energy: To assess the biomass potential, agricultural and forest waste (fuelwood) have been considered. On the basis of data published [7] from all the nine villages it is estimated that about 1083.35 Tonnes/year fuelwood and 44.71 Tonnes/year of crop residue is available as surplus by taking 2% sustainable yield of fuelwood.

However, in this study only 1% sustainable yield of fuelwood is considered to avoid deforestation and crop residues are left to feed the livestocks. The balance 147.757 Tonnes/year is available for generation of electricity.

C. Demand Assessment

The data for load demand estimation has been collected on the basis of questionnaire. The energy demand was estimated by considering the household load (lighting, T.V., fan, radio/music system), commercial load (lighting for small shops and floor mill), industrial load (saw mill or paddy huller) and community load (primary health centre, street lights, and school lighting). The total energy requirement is estimated as 1271.61 kWh/day in summer and 608.97 kWh/day in winter. A detail of total load of cluster of nine villages with demand side management is shown in Table 5. The general details of villages and electric appliances used for villages are shown in Table 6 and Table 7.

D. Unit cost of resources

The cost of energy generated by a renewable energy resource is obtained by adding the capital recovery cost and operation & maintenance cost per unit of energy. Typical calculations are made on an annual basis and the cost of energy in Rs/kWh is calculated by the following expressions:

$$COE = \frac{ALCC\ (Rs)}{\text{Total annual energy generated (kWh)}} \quad (1)$$

$$ALCC = C_O \times CRF + AC_F + AC_{O\ \&\ M} \quad (2)$$

Where, ALCC = Annualized capital cost (Rs)
COE = Unit cost of energy (Rs/kWh)
CRF = Capital recovery factor
C_0 = Capital cost Rs/kW
AC_F = Annual fuel cost (Rs)
$AC_{O\ \&\ M}$ = Annualized operation and maintenance cost

$$CRF = \left[\frac{d\ (1+d)^n}{(1+d)^n - 1} \right]$$

d = interest rate
n = life time

The unit cost of different resources is shown in Table 8.

Table 8 Sizing & Cost of Energy for different resources

S. N.	Type of Energy Resources	Cost of Energy (Rs/kWh)	Installed capacity
1.	Micro Hydro Generator	1.45	15 kW
2.	Solar Photovoltaic Generator	15.68	23.31 kW
3.	Biogas Generator	3.98	20 kW
4.	Biomass Generator	4.78	34 kW
5.	Diesel Engine Generator	11.0	46 kW
6.	Battery	3.26	106 kWh
7.	Inverter	-	35 kW
8.	Photovoltaic-inverter	17.72	-
9.	Battery-inverter	4.33	-

E. System sizing

A computer program is developed to determine the optimal sizing of system components by minimising the system unit cost. Data input to the program are

i) Hourly load demand data for the design month.

ii) Hourly solar radiation data for the design month.

iii) Unit cost of different components and other required parameters.

The program will use these input to determine optimum size of each system component. The program repeatedly simulates hourly system operation over the month for every component combinations. The feasible solutions are ranked by system unit cost. The optimum sizing of different component is given in Table 8.

V. Optimized Simulations Results

The optimized simulation results for hybrid energy system for a design month August using developed computer algorithm are shown in Table 11, 12, 13 and 14 and monthly results in Table 15. Different values for parameters used for simulation purpose is shown in Table 9 & 10.

The results shown are for specified parameters, which can vary for individual customers, as well as, from area to area. It can be seen that the least economical system is the stand-alone micro hydro generation system (1.45 Rs/kWh) as it has to be run all the time in order to meet the load demand constantly. On the other hand, the most expensive system is the stand-alone solar photovoltaic system (15.68 Rs./kWh). So the stand-alone system will cost more money than it is necessary

Regarding the biogas energy it is clear that potential of biogas is sufficient with second lowest cost of energy.

Table 5 Details of Total Load (in kWh) of Nine Villages

Time Segment (Hours)		Electrical Load (kWh)										Total Electrical Load/Hour (kWh) Summer/ Winter	Thermal Load (kWh)
		Household Load				Commercial Load		Industrial Load	Community Load				Cooking Load
		Lighting Load (11 W)	T.V. (90 W)	Fan (55 W) Summer/ Winter	Radio/ Music System (25 W)	Lights for Small Shops (20 W)	Floor Mill (5 kW)	Saw Mill /Paddy Huller (5 kW)	One Primary Health Centre (20 W)	Street Lights (20 W)	School Lights (20 W)		
0:00-1:00										2.0		2.0	
1:00-2:00										2.0		2.0	
2:00-3:00										2.0		2.0	
3:00-4:00										2.0		2.0	
4:00-5:00		5.522								2.0		7.522	
5:00-6:00		5.522								2.0		7.522	
6:00-7:00		11.044										11.044	
7:00-8:00		11.044	45.18		12.55							68.774	335.05
8:00-9:00			45.18		12.55				0.040		0.72	58.49	
9:00-10:00							5.0	5.0	0.040		0.72	10.76	
10:00-11:00							5.0	5.0	0.040		0.72	10.76	
11:00-12:00				55.22 / 0.0			5.0	5.0	0.040		0.72	65.98 / 10.76	
12:00-13:00			45.18	55.22 / 0.0			5.0	5.0	0.040		0.72	111.16 / 55.94	
13:00-14:00			45.18	55.22 / 0.0								100.40 / 45.18	670.08
14:00-15:00				55.22 / 0.0								55.22 / 0.0	
15:00-16:00				55.22 / 0.0	12.55							67.77 / 12.55	
16:00-17:00				55.22 / 0.0	12.55							67.77 / 12.55	
17:00-18:00		11.044	45.18	55.22 / 0.0		0.34			0.040			111.824 / 56.604	
18:00-19:00		11.044	45.18	55.22 / 0.0		0.34			0.040	2.0		113.824 / 58.604	
19:00-20:00		11.044	45.18	55.22 / 0.0		0.34			0.040	2.0		113.824 / 58.604	
20:00-21:00		11.044	45.18	55.22 / 0.0		0.34			0.040	2.0		113.824 / 58.604	670.08
21:00-22:00		5.522	45.18	55.22 / 0.0						2.0		107.922 / 52.702	
22:00-23:00				55.22 / 0.0						2.0		57.22 / 2.0	
23:00-24:00										2.0		2.0 / 2.0	
Daily Load		**82.83**	**406.62**	**662.64 / 0**	**50.20**	**1.36**	**20.0**	**20.0**	**0.360**	**24.0**	**3.60**	**1271.61 / 608.97**	**1675.21**
Summer	**March Load**	**2567.73**	**12605.22**	**20541.84**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**39419.91**	**51931.51**
Summer	**April Load**	**2484.90**	**12198.60**	**19879.20**	**1506.0**	**40.80**	**600**	**600**	**10.80**	**720**	**108.0**	**38148.30**	**50256.30**
Summer	**May Load**	**2567.73**	**12605.22**	**20541.84**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**39419.91**	**51931.51**
Summer	**June Load**	**2484.90**	**12198.60**	**19879.20**	**1506.0**	**40.80**	**600**	**600**	**10.80**	**720**	**108.0**	**38148.30**	**50256.30**
Summer	**July Load**	**2567.73**	**12605.22**	**20541.84**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**39419.91**	**51931.51**
Summer	**Aug. Load**	**2567.73**	**12605.22**	**20541.84**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**39419.91**	**51931.51**
Summer	**Sep. Load**	**2484.90**	**12198.60**	**19879.20**	**1506.0**	**40.80**	**600**	**600**	**10.80**	**720**	**108.0**	**38148.30**	**50256.30**
Summer	**Oct. Load**	**2567.73**	**12605.22**	**20541.84**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**39419.91**	**51931.51**
Winter	**Nov. Load**	**2484.90**	**12198.60**	**0.0**	**1506.0**	**40.80**	**600**	**600**	**10.80**	**720**	**108.0**	**18269.10**	**50256.30**
Winter	**Dec. Load**	**2567.73**	**12605.22**	**0.0**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**18878.07**	**51931.51**
Winter	**Jan. Load**	**2567.73**	**12605.22**	**0.0**	**1556.2**	**42.16**	**620**	**620**	**11.16**	**744**	**111.6**	**18878.07**	**51931.51**
Winter	**Feb. Load**	**2319.24**	**11385.36**	**0.0**	**1405.6**	**38.08**	**560**	**560**	**10.08**	**672**	**100.8**	**17051.16**	**46905.88**
Yearly Load		**30232.95**	**148416.3**	**162346.8**	**18323.0**	**496.40**	**7300**	**7300**	**131.40**	**8760**	**1314.0**	**384620.85**	**611451.65**

Table 6 General Details of Clusters of Villages

Parameters	Details
Households (HH)	502
Population	3223
Basic School (4 rooms)	3
Junior Basic School Boys (8 rooms)	2
Senior Secondary School Boys (8 rooms)	1
Primary Health Centre (2 rooms)	1
Floor Mill	1
Saw Mill / Paddy huller	1
Fair Price Shops / Control Rate Shops	16 / 1

Table 7 Electric Appliances used for the Villages

Appliances	Quantity/HH or room	Total
CFL for HH Lighting	1 or 2 Points	1004
CFL for small shops	1 Points	17
CFL for health centre	1 Points	2
CFL for street lights	[1 Pole @ Clusters of 5 HH]	100
CFL for school lighting	1 Points	36
Colour T.V. (36 cm)	1	502
Ceiling fan	2	1004
Radio/Music system	1	502

Table 9 Different parameters used for simulation

S. N.	Various Generators	Efficiency
1.	Micro Hydro Generator **(MHG)**	0.60
2.	Solar Photovoltaic Generator **(PVG)**	0.1154
3.	Biogas Generator **(BGG)**	1.0
4.	Biogas Energy System	0.27
5.	Biomass Generator **(BMG)**	1.0
6.	Biomass Energy System	0.21
7.	Diesel Engine Generator **(DEG)**	1.0
8.	Battery Charging efficiency	0.90
9.	Battery Discharging efficiency	1.0
10.	Rectifier / Inverter efficiency	0.95
11.	Charge Controller efficiency	0.90

Table 10 Operating period for different generators

S. N.	Various Generators	Operating Period
1.	**Micro Hydro Generator**	22 hour/day
2.	**Solar Photovoltaic Generator**	-
3.	**Biogas Generator**	10 hour/day
4.	**Biomass Generator**	12 hour/day
5.	**Diesel Engine Generator**	10 hour/day (max)

Table 11 Hourly Simulation Results for Design Month August

Time Segment	E_{LOAD} (kWh)	E_{MHG} (kWh)	E_{BGG} (kWh)	E_{BMG} (kWh)	E_{PVG} (kWh)	$E_{PVG\text{-}INV}$ (kWh)	$E_{NETLOAD}$ (kWh)	$E_{SURPLUS}$ (kWh)	$E_{BATT\text{-}INV}$ (kWh)	$E_{BATT\text{-}LEFT}$ (kWh)	E_{DEG} (kWh)	DEG Status	E_{UNMET} (kWh	E_{DUMP} (kWh	Unit Cost Rs./kWh
First Day Simulation															
0:0-1:0	2.0	-	-	-	0.0	-	2.0	0.0	2.0	103.8947			-	-	4.33
1:0-2:0	2.0	2.0	-	-	0.0	-	0.0	5.3865	0.0	106.0			-	3.2812	1.45
2:0-3:0	2.0	2.0	-	-	0.0	-	0.0	5.3865	0.0	106.0			-	5.3865	1.45
3:0-4:0	2.0	2.0	-	-	0.0	-	0.0	5.3865	0.0	106.0			-	5.3865	1.45
4:0-5:0	7.522	7.522	-	-	0.0	-	0.0	1.1373	0.0	106.0			-	1.1373	1.45
5:0-6:0	7.522	7.522	-	-	0.0	-	0.0	1.1373	0.0	106.0			-	1.1373	1.45
6:0-7:0	11.044	9.0	-	-	2.0979	1.9930	0.0509	0.0	0.0509	105.9464			-	-	4.40
7:0-8:0	68.774	9.0	20.0	-	5.3615	5.0934	34.6805	0.0	34.6805	69.4406			-	-	4.84
8:0-9:0	58.49	9.0	20.0	-	8.8581	8.4151	21.0748	0.0	21.0748	47.2566			-	-	5.69
9:0-10:0	10.76	9.0	-	-	11.4223	1.7600	0.0	7.7514	0.0	55.0080			-	-	4.11
10:0-11:0	10.76	9.0	-	-	13.2872	1.7600	0.0	9.2620	0.0	64.2700			-	-	4.11
11:0-12:0	65.98	9.0	20.0	34.0	14.9189	2.9800	0.0	9.5435	0.0	73.8135			-	-	4.67
12:0-13:0	111.16	9.0	20.0	34.0	14.9189	14.1730	33.9870	0.0	33.9870	38.0377			-	-	5.88
13:0-14:0	100.40	9.0	20.0	34.0	13.2872	12.6228	24.7771	16.3310	0.0	54.3687	24.7771	On-Run	-	-	7.48
14:0-15:0	55.22	9.0	-	34.0	11.4223	10.8512	1.3688	34.3437	0.0	88.7124	1.3688	Run-Off	-	-	6.93
15:0-16:0	67.77	9.0	-	34.0	9.0912	8.6367	16.1334	0.0	16.1334	71.7298	-	-	-	-	5.88
16:0-17:0	67.77	9.0	-	34.0	4.8720	4.8720	19.8980	0.0	19.8980	50.7846	-	-	-	-	5.14
17:0-18:0	111.824	9.0	20.0	34.0	2.0979	1.9930	46.8310	0.0	0.8310	49.9098	46.0	On-Run	-	-	7.16
18:0-19:0	113.824	9.0	20.0	34.0	0.0	-	50.824	0.0	4.824	44.8319	46.0	Run	-	-	6.87
19:0-20:0	113.824	9.0	20.0	34.0	0.0	-	50.824	0.0	4.824	39.7540	46.0	Run	-	-	6.87
20:0-21:0	113.824	9.0	20.0	34.0	0.0	-	50.824	0.0	4.824	34.6761	46.0	Run	-	-	6.87
21:0-22:0	107.922	9.0	20.0	34.0	0.0	-	44.922	0.8295	0.0	35.5056	44.922	Run	-	-	6.94
22:0-23:0	57.22	9.0	-	34.0	0.0	-	14.220	24.4547	0.0	59.9603	14.220	Run	-	-	5.80
23:0-24:0	2.0	-	-	-	0.0	-	2.0	33.8580	0.0	93.8183	2.0	Run-Off	-	-	11.0
Second Day Simulation															
24:0-25:0	2.0	-	-	-	0.0	-	2.0	0.0	2.0	91.7130			-	-	4.33
25:0-26:0	2.0	2.0	-	-	0.0	-	0.0	5.3865	0.0	97.0995			-	-	1.45
26:0-27:0	2.0	2.0	-	-	0.0	-	0.0	5.3865	0.0	102.4860			-	-	1.45
27:0-28:0	2.0	2.0	-	-	0.0	-	0.0	5.3865	0.0	106.0			-	1.8725	1.45
28:0-29:0	7.522	7.522	-	-	0.0	-	0.0	1.1373	0.0	106.0			-	1.1373	1.45
29:0-30:0	7.522	7.522	-	-	0.0	-	0.0	1.1373	0.0	106.0			-	1.1373	1.45
30:0-31:0	11.044	9.0	-	-	2.0979	1.9930	0.0509	0.0	0.0509	105.9464			-	-	4.40
31:0-32:0	68.774	9.0	20.0	-	5.3615	5.0934	34.6805	0.0	34.6805	69.4406			-	-	4.84
32:0-33:0	58.49	9.0	20.0	-	8.8581	8.4151	21.0748	0.0	21.0748	47.2566			-	-	5.69
33:0-34:0	10.76	9.0	-	-	11.4223	1.7600	0.0	7.7514	0.0	55.0080			-	-	4.11
34:0-35:0	10.76	9.0	-	-	13.2872	1.7600	0.0	9.2620	0.0	64.2700			-	-	4.11
35:0-36:0	65.98	9.0	20.0	34.0	14.9189	2.9800	0.0	9.5435	0.0	73.8135			-	-	4.67
36:0-37:0	111.16	9.0	20.0	34.0	14.9189	14.1730	33.9870	0.0	33.9870	38.0377			-	-	5.88
37:0-38:0	100.40	9.0	20.0	34.0	13.2872	12.6228	24.7771	16.3310	0.0	54.3687	24.7771	On-Run	-	-	7.48
38:0-39:0	55.22	9.0	-	34.0	11.4223	10.8512	1.3688	34.3437	0.0	88.7124	1.3688	Run-Off	-	-	6.93
39:0-40:0	67.77	9.0	-	34.0	9.0912	8.6367	16.1334	0.0	16.1334	71.7298	-	-	-	-	5.88
40:0-41:0	67.77	9.0	-	34.0	4.8720	4.8720	19.8980	0.0	19.8980	50.7846	-	-	-	-	5.14
41:0-42:0	111.824	9.0	20.0	34.0	2.0979	1.9930	46.8310	0.0	0.8310	49.9098	46.0	On-Run	-	-	7.16
42:0-43:0	113.824	9.0	20.0	34.0	0.0	-	50.824	0.0	4.824	44.8319	46.0	Run	-	-	6.87
43:0-44:0	113.824	9.0	20.0	34.0	0.0	-	50.824	0.0	4.824	39.7540	46.0	Run	-	-	6.87
44:0-45:0	113.824	9.0	20.0	34.0	0.0	-	50.824	0.0	4.824	34.6761	46.0	Run	-	-	6.87
45:0-46:0	107.922	9.0	20.0	34.0	0.0	-	44.922	0.8295	0.0	35.5056	44.922	Run	-	-	6.94
46:0-47:0	57.22	9.0	-	34.0	0.0	-	14.220	24.4547	0.0	59.9603	14.220	Run	-	-	5.80
47:0-48:0	2.0	-	-	-	0.0	-	2.0	33.8580	0.0	93.8183	2.0	Run-Off	-	-	11.0

Table 12 Simulation Results for Daily and Monthly

Description of Parameters	Daily*	Monthly**	Percentage
Total Load (kWh)	1271.61	39419.91	100.0
Total MHG output (kWh)	198	6138	-
Total BGG output (kWh)	200	6200	-
Total BMG output (kWh)	408	12648	-
Total PVG output (kWh)	111.8918	3468.6458	-
Total REG output (kWh)	917.8918	28454.6458	-
Load by MHG (kWh)	174.044	5395.364	13.6869
Load by BGG (kWh)	200	6200	15.7281
Load by BMG (kWh)	408	12648	32.0853
Load by PVG-INV (kWh)	75.1502	2329.6562	5.9098
Load by REG (kWh)	857.1942	26573.0202	67.4101
Load by DEG (kWh)	271.2879	8409.9249	21.3342
Load by Battery (kWh)	143.1276	4436.9556	11.2557
Total Unmet Energy (kWh)	0.0	0.0	0.0
Total Dump Energy (kWh)	4.1471	128.5601	0.3261

Table 13 Economic Results

Parameters	Rs. / kWh
Optimum Maximum Hourly Unit Cost of HES	11.0
Optimum Minimum Hourly Unit Cost of HES	1.45
Optimum Daily Unit Cost of HES	6.24
Optimum Monthly Unit Cost of HES	6.24

Table 14 Fuel Consumption & Battery Results

Parameters	Daily*	Monthly**
DG Fuel Consumption (Liters)	70.60	2188.84
DG Run Hours (h)	9.0	279.0
DG Start-Stops	2.0	62.0
Battery Initial State (kWh)	93.8183	-
Battery Final State (kWh)	93.8183	-

* Hourly Simulation Results after 24 hours
** Monthly Simulation Results after 744 hours

Table 15 Monthly results

Description of Parameters (kWh)	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
Total Load	18878.07	17051.16	39419.91	38148.3	39419.91	38148.3	39419.91	39419.91	38148.30	39419.91	18269.10	18878.07
Total MHG output	0.0	0.0	6138	5940	6138	5940	6138	6138	5940	6138	0.0	0.0
Total BGG output	6045.0	5600	6200	6000	6200	6000	6200	6200	6000	6200	6000	5580
Total BMG output	11346.0	9520	12648	12240	12648	12240	12648	12648	12240	12648	10200	11594
Total PVG output	2981.42	2793.57	3858.85	4349.80	4884.99	4139.98	4162.35	3468.64	3678.42	3627.60	2923.17	2500.30
Total REG output	20372.42	17913.57	28844.85	28529.80	29870.99	28319.98	29148.35	28454.64	27858.42	28613.60	19123.17	19674.30
Load by MHG	0.0	0.0	5395.36	5221.32	5395.36	5221.32	5395.36	5395.36	5221.32	5395.36	0.0	0.0
Load by BGG	5302.36	5349.23	6200	6000	6200	6000	6200	6200	6000	6200	5731.32	5302.36
Load by BMG	9896.81	7987.05	12648	12240	12648	12240	12648	12648	12240	12648	8557.56	9896.81
Load by PVG-INV	1665.62	1706.49	2495.65	2768.952	3080.75	2702.34	2689.44	2329.65	2407.29	2315.91	1801.81	1669.65
Load by REG	16864.79	15042.78	26739.01	26230.27	27324.11	26163.66	26932.81	26573.02	25868.61	26559.27	16090.69	16868.83
Load by DEG	0.0	0.0	8709.51	8275.551	8139.02	8303.901	8661.45	8409.92	8038.98	8313.81	0.0	0.0
Load by Battery-inv	2013.27	2008.37	3971.36	3642.636	3956.74	3680.715	3825.62	4436.95	4240.67	4546.80	2178.40	2009.23
Unmet Energy (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dump Energy (%)	0.09	0.39	1.43	0.02	0.64	0.96	2.12	0.33	1.02	0.58	0.05	0.03
Diesel Fuel Consumption (L)	0.0	0.0	2262.54	2072	2122.20	2158.89	2250.72	2188.84	2093.72	2165.20	0.0	0.0
DG Run Hours (h)	0.0	0.0	279	240	248	255	279	279	270	279	0.0	0.0
DG Start-Stops	0.0	0.0	31	30	31	30	31	62	60	62	0.0	0.0
Daily Unit Cost	5.65	5.77	6.35	6.44	6.45	6.43	6.40	6.24	6.28	6.22	5.75	5.65
Monthly Unit Cost	5.65	5.77	6.35	6.44	6.45	6.43	6.40	6.24	6.28	6.22	5.75	5.65
Average Annual Unit Cost (Rs./kWh)					**6.14**							

In order to fully utilize the biogas resource, one is required to explore the possibility of generating electricity using biogas engine system in decentralized mode because the cost of generation from the individual resource is Rs. 3.98/kWh followed by biomass energy system (Rs. 4.98/kWh), diesel generator (Rs. 11.0/kWh).

The optimized annual system unit cost is determined by taking the average of optimized unit cost of all months, which comes out to be Rs. 6.14/kWh.

VI. Conclusions

Given the fact that a hybrid energy system consisting two or more energy system has the advantage of stability, supply the power on sustainable basis. The objective of the electrification at the study area can be achieved by making use of solar, micro hydro, biogas, biomass, battery with diesel generator based hybrid energy system. The system was modelled by making use of the computer program based on combined dispatch strategy, developed in C++. Depending upon the variation in discharge and availability of other resources and future increase in demand, the hybrid energy system as indicated above may be able to fulfil the demand of study area in the integrating manner. The local people will be employed to take off the operation and maintenance of the power system as well as to manage the collection of revenues from each household, which may be used for maintaining the sustainability of the system.

References

[1] R. Ramakumar, I. Abouzahr and K. Asenyayi, "A Knowledge-Based approach to the Design of Integrated Renewable Energy Systems," *IEEE Trans. on Energy Conversion,* vol. 7, No. 4, pp. 648-657, 1992..

[2] R. Ramakumar, I. Abouzahr, K. Krishnan and K. Ashenayi, "Design Scenario for integrated Renewable Energy Systems", *IEEE Trans. on Energy Conversion*, Vol. 10, No. 4, pp. 736-746, December 1995.

[3] K. Rajashekara, "Hybrid fuel cell strategies for clean power generation," in *Proc. 2004 of IEEE,* pp. 2077-2083, March 1993..

[4] A. Rosenthal, S. Durand, M. Thomas and H. Post, "Economic analysis of PV hybrid power system: Pinnacles National Monument," in *Proc. of IEEE Photovoltaic Specialists Conf.*, pp. 1269-1272.

[5] W.D. Kellog, M.H. Nehrir, G. Venkataramanan, and V. Gerez, "Generation unit sizing and cost analysis for stand-alone wind, photovoltaic, and hybrid wind/pv systems," *IEEE Transactions on Energy Conversion*, vol. 13, No. 1, pp. 70-75, March 1998.

[6] H. C. de. Coninck, K. J. Dinesh, A. Kets, S. Maithel, P. Mohanty, and H. J. de Vries, "Providing electricity to remote villages-Implementation models for sustainable of India's rural power," *Energy research centers of Netherlands*, ECN Rep. ECN-C-05-037, July 2005.

[7] "Unelectrified villages: surveys, potential sources and electricity demand," Alternate Hydro Energy Centre, I.I.T. Roorkee, Main Project Report, 2005, vol. II (4/12).

[8] *Solar Radiation over India,* 3rd ed., Allied Publishers Private Limited, India, 1982, pp. 302–303.

Biographies

Ajai Gupta was born in Bareilly, India. He received the B. Tech in Electrical Engineering from I.E.T. Rohilkhand University, Bareilly in 2000 and M. Tech in Instrumentation and Control from Aligarh Muslim University, Aligarh in 2004 respectively. Currently he is a Research Scholar at Alternate Hydro Energy Centre, Indian Institute of Technology, Roorkee, India.

Dr. R. P. Saini is serving as an Associate Professor at A.H.E.C, Indian Institute of Technology, Roorkee, India.

Dr. M. P. Sharma has been working as an Associate Professor at A.H.E. C., Indian Institute of Technology, Roorkee, India since the last 20 years

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Control and Instrumentation for Small Wind Turbines

R. Ahshan, *Student Member, IEEE,* M. T. Iqbal, and George K. I. Mann

***Abstract*— Details of control and instrumentation for on-grid and off-grid operation of a small induction generator based wind turbine are presented in this paper. A micro-controller based system controller is used to connect/disconnect the wind power generator to the grid, and also to maintain the grid connection. Several real time situations are considered during the operation of the wind power generation system to test the designed control and instrumentation. The energy generated during the operation in the off-grid mode can be used for space heating or for consumer load which needs stable voltage in spite of the wind or the load variations. To achieve this, a phase control relay based electronic PI controller is designed which maintain the more or less constant voltage at the load terminal. In this research, a dump load is used and its terminal voltage is regulated by controlling the phase control relay. System instrumentation and test results are presented in this paper.**

***Index Terms*— Wind energy, Small wind energy conversion system, Control system, Instrumentation.**

I. Nomenclature

A	=	Current in ampere
AC	=	Alternating current
CB	=	Circuit braker
D_0	=	Relay driving signal for *relay_1*
D_1	=	Relay driving signal for *relay_2*
D_2	=	Signal for normally open solid state relay
DC	=	Direct current
f_{ig}	=	Frequency on generator terminal
f_g	=	Frequency on grid side
PCR	=	Phase control relay
CT	=	Current transformer
I	=	Current flow between the generator and grid
IG	=	Induction generator
IGBT	=	Insulated-gate bipolar transistor
P	=	Power flowing between IG and grid
NO	=	Normally open
PI	=	Proportional-integral
PIC	=	Peripheral interface controller
rms	=	Root mean square
V	=	Voltage in volt
v_{ig}	=	Voltage on generator terminal
v_g	=	Voltage on grid side
V_{ref}	=	Reference voltage for the PI controller
V_{fed}	=	Feedback voltage for the PI controller
WTS	=	Wind turbine simulator
WECS	=	Wind energy conversion system

II. Introduction

A small wind energy conversion system is required to operate in both the grid connected and off-grid mode. The grid connection operation of a micro-controller controlled small induction generator based wind turbine is presented in [1]. However, to achieve optimum operation of a generation system, the grid uncertainty is also required to be considered while the wind turbine is in operation. Control system is important to achieve such a goal. Although the system controller performs the grid connection/disconnection based on wind and grid availability, another controller is required while the system operates in off-grid mode and supplies power to the space heating load or to the consumer load. This paper presents such a control technique and associated instrumentation to regulate the voltage at the load terminal.

Past research on electronic load controller found in [2, 3, 4, 5, 6] are based on different techniques. Firstly, a rectifier chopper feeding a fixed resistive dump load is discussed in [2], where dump power is controlled by varying the duty cycle of the chopper. As choppers receive power from a fixed voltage DC source using a rectifier, therefore this arrangement can not be recommended for the directly grid connected wind turbine system. Secondly, an AC controller with a back-to-back thyristors feeding a fixed resistive dump load where firing angle varies power in the dump load is described in [4, 5, 6]. In a phase controlled thyristor-based load controller [4, 5], the phase angle of back-to-back connected thyristors is delayed from 0 to 180 degree as the consumer load is changed from zero to full load. Due to a delay in firing angle, it demands reactive power loading and injects

harmonics in the system. It also requires complicated driver circuits. Another electronic load controller based on anti parallel insulated-gate bipolar transistor switches is used to control the dump load connection and disconnection in [3]. The technique shown in [6] is the only method applicable for wind energy applications, whereas other methods in [3, 4, 5] are demonstrated for micro/pico hydro applications. This paper proposed control and instrumentation based on phase control relays to regulate the voltage at the dump load terminal. Proposed arrangement is simple because only one signal can operate all three phase control relays for a three phase system, and it does not require any driving circuit.

III. Operational Modes of a Wind Power Generator

One of the important aspects of a grid connected wind turbine system is to run the wind power generation system during the grid absence and maintain the output voltage stable using proper control mechanism. The entire operation of a wind power generation system is divided into two modes.

a) Grid connected mode

b) Off-grid operation mode

Two modes of operation of such a system increases its reliability and provide optimum output in both grid conditions.

A. Grid Connected Mode

In this mode, basically all the available power that can be extracted from the wind is transferred to the grid. This mode depends on the availability of the wind and the grid. The structure of the grid connected mode is shown in Figure 1 which consists of wind turbine simulator that represents wind turbine and gear box, self-excited induction generator, system controller, and soft-starter.

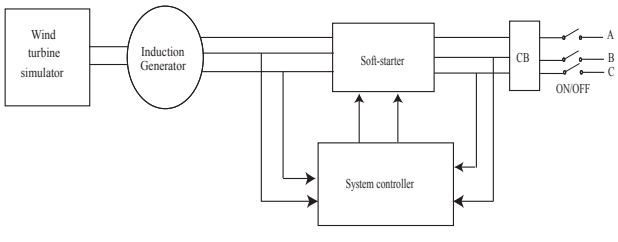


Fig. 1. Block diagram of the grid connected mode

During this mode, the system controller [1] functions while wind is enough to produce power at the grid voltage and frequency with the presence of grid. The system controller always monitors the generator terminal voltage and frequency, as well as the grid voltage and frequency at the available wind speed. While the generator and the grid parameters are matched, the system controller connects the generator to the grid via the soft-starter. Once it is connected to the grid, the controller keeps checking current or power flow between the generator and the grid or the speed of the generator. Depending on these parameters the controller disconnects the system from the grid. The controller detects the generator operating mode that whether it is in motoring mode or in generating mode based on either current or power flow between the wind turbine system and the grid, or the speed of the generator. The control strategy is implemented on a development platform consisting of a wind turbine simulator [7] based on a peripheral interface micro-controller.

B. Off-grid Operation Mode

In this mode, the power generated by the wind power generating system needs to be transferred to the consumer load and the output voltage also requires to be controlled in terms of amplitude. This mode is considered while the grid is absent but wind is available to produce power. The system structure in off-grid mode is shown in figure 2 and it consists of wind power generator, system controller, soft-starter, NO solid state relay, phase control relay based electronic PI controller to regulate the voltage at the dump load terminal and the dump load. The system controller and the phase control relay based electronic PI controller work simultaneously during the operation of this mode . The system controller performs both in grid-connected mode and off-grid mode whereas electronic PI controller works only during the off-grid mode.

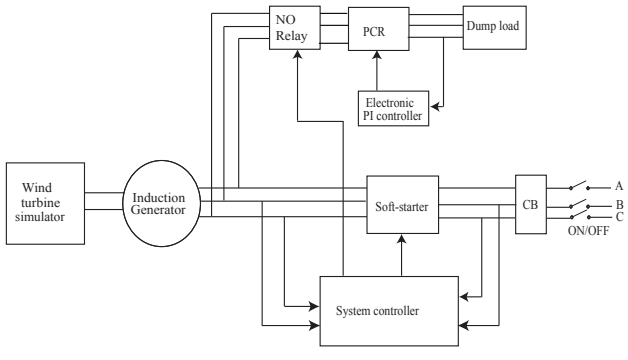


Fig. 2. Block diagram of the off-grid mode of operation

Suppose wind turbine is generating power and after checking the system parameters, the system controller connects the generator to the grid via power resistor based soft-

starter. Once it is connected to the grid, the controller keeps monitoring the voltages at the grid side and either power or current flow between the generator and the grid. In case of the grid failure or in off-grid situations, the grid voltage read by the system controller is zero. Once the system controller receives zero voltage from the grid side, the controller disconnects the wind power generating system from the grid and sends a control value to the NO solid state relay to connect the system to a dump load. After connected to the dump load, electronic PI controller regulates the amplitude of the voltage at the dump load terminal as it is required. At the same time, the system controller keeps checking the generator parameters and also the grid availability. While the generator terminal voltage is too low due to the lack of wind, the system controller sends a control value to the NO solid state relay to disconnect the generating system from the dump load otherwise keeps the system connected to the dump load. When the grid is recovered, the system controller again checks the grid connection conditions. If conditions are met, the system controller disconnects the wind turbine system from the dump load and reconnects the system to the grid. Phase control relay based electronic controller is chosen for voltage regulation because of ease of implementation. A single control value can control the three phase control relays used in three different phases. This control strategy also keeps the wind turbine system free from islanding situation. Islanding of a grid connected wind turbine generator system occurs when a section of the utility system containing such generator is disconnected from the main utility, but the generators still continue to energize the utility lines in the isolated section. Situation may occur in the proposed system while wind is enough to produce power but the grid is absent. The proposed controller has the ability to detect the grid failure and switch over the system operation in stand-alone mode, hence islanding situation will never occur.

IV. Instrumentation for Off-grid Mode of Operation

Block diagram of a low cost instrumentation for measuring the parameters of a small wind turbine system is shown in Figure 3.

The measurement system shown in Figure 3 consists of voltage sensors, current sensor, power measuring unit, frequency to voltage converter, soft-starter, an electronic PI controller based on phase control relay, a technique for acquiring feedback and reference signal for the PI controller. Voltage sensors are placed both at the generator and grid sides to measure the voltages. The current flow between the generator and the grid is measured using a current transformer. In power measuring unit, a four quadratic multiplier is used in order to determine the power flow

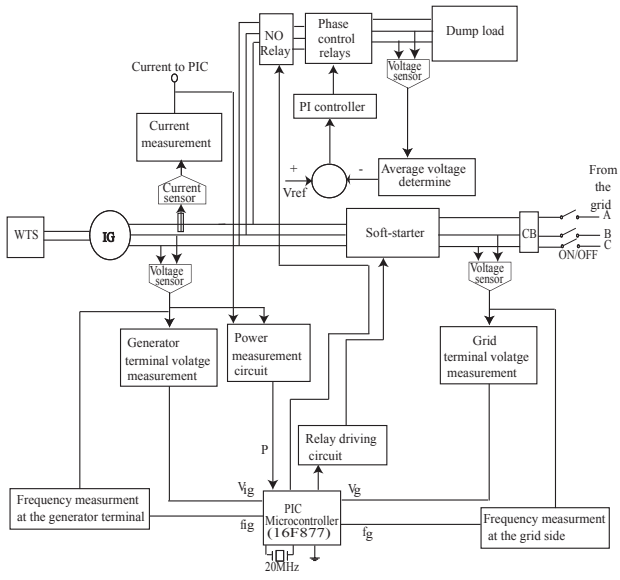


Fig. 3. Block diagram of the instrumentation for off-grid mode of operation

between the generator and grid. Voltage and current are the inputs to the multiplier which are acquired by voltage and current sensor. Frequency to voltage converter circuitry is designed to measure the frequency at both the generator terminal and the grid side. All measuring units are directly or indirectly connected to the PIC micro-controller. Zener diodes are used to protect the micro-controller while the voltage at the analog input of the micro-controller exceeds 5V. An electronic controller (see Figure 7) is designed using operational amplifiers which sends the control variable to the phase control relay. The phase control relay takes 2-10V DC at its input and provides the AC output voltage rated at 120/240V and 25A. The feedback mechanism is performed using voltage sensor, peak detector and low pass filter. Figure 4 shows the designed instrumentation in detail.

V. Control System Design for Off-grid Mode of Operation

Off-grid operation mode of a small wind turbine system is controlled using system controller and phase control relay based electronic PI controller. The main purpose to design control system is to take care and to optimum use of the wind turbine generator output during the grid absence.

Figure 5 shows the block diagram of the control operation in off-grid mode. Suppose the wind power generation system is running in grid connected mode. During operation in this mode, the system controller monitors either current or power flow between the generator and the grid, and the grid status. The grid status is detected by the system

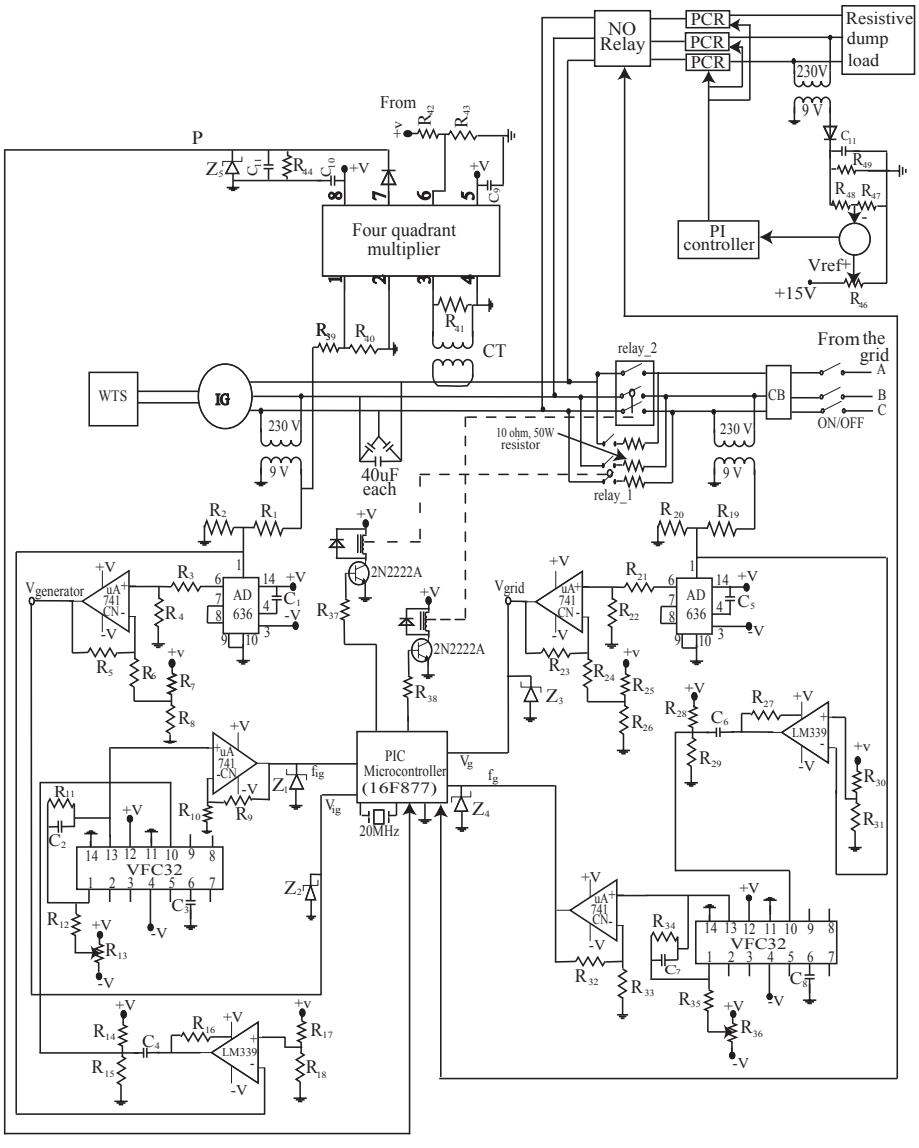


Fig. 4. Instrumentation for off-grid mode of operation and control

controller using grid voltage sensing. If the voltage sensed by the system controller is near or equal to zero, the system controller disconnects the wind turbine system from the grid. After disconnection from the grid, the system controller sends a signal D_2 to the NO solid state relay to initiate the connection between the generator and the dump load. The PI controller is designed to maintain rated line-to-line/phase voltage at the load terminal and it works as follows. The voltage at the output of the voltage sensor due to the rated line-to-line voltage at the load terminal is taken as the reference voltage. If the feedback voltage is less than or equal to the reference voltage then no control action is required to chop the voltage at the load terminal. The load terminal voltage at which the feedback voltage is less or equal to the reference voltage is basically less or equal to the rated line-to-line/phase voltage. If the feedback voltage is greater than the reference voltage, the load terminal voltage will be higher than the rated value which needs to be chopped to keep the rated voltage at the load terminal. In such case the PI controller sends a control value to trigger the phase control relay and regulates the voltage at the load terminal. The input of the phase control relay varies from 2-10V, upon which the output voltage is regulated. At the same time, the system controller checks

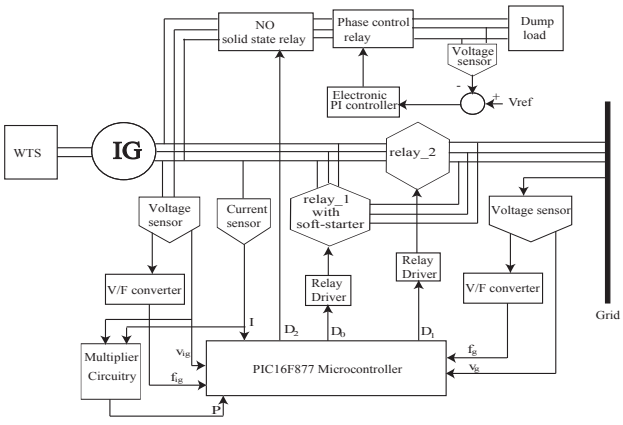


Fig. 5. Block diagram of the control operation during off-grid mode

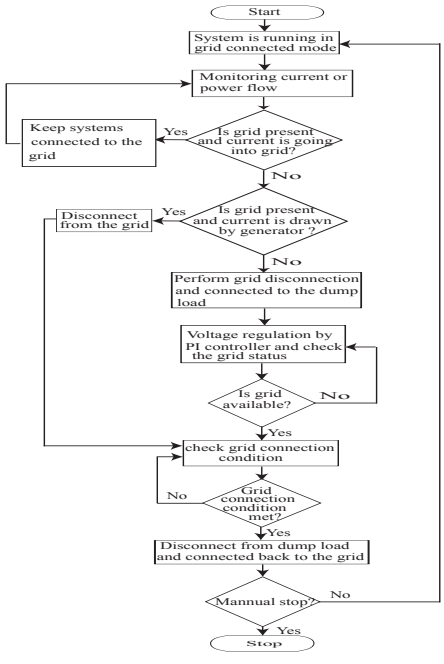


Fig. 6. Flow chart of the controller operation in off-grid mode

the grid status and the grid connection conditions. If the grid recovers and grid connection conditions are met, the system controller sends a signal (D_2 = low) to the solid state relay to disconnect the system from the dump load and connects the wind turbine system back to the grid, otherwise the system controller keeps checking the conditions.The described control operation performs according to the flow chart shown in Figure 6. Figure 7 shows the op-amp based PI controller to operate phase control relays.

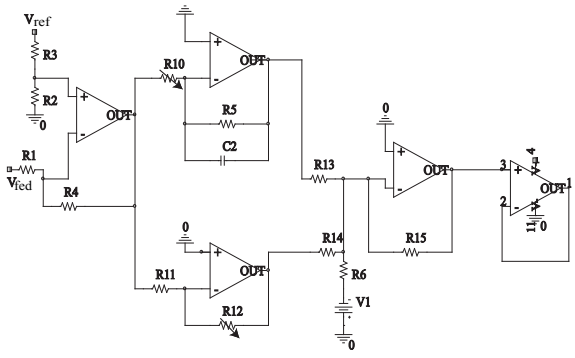


Fig. 7. Op-amp based PI controller design

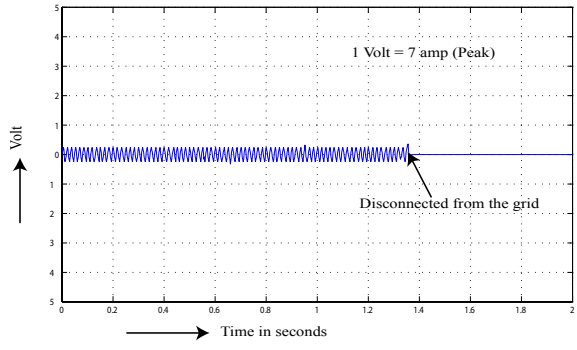


Fig. 8. Current flow between the generator and the grid

VI. Test Results

Both the system controller and PI regulator tests were performed together during the off-grid operation. The system controller was assisted to connect/disconnect the system to the grid, grid status checking and connect the system with the dump load. Figure 8 shows the current flow when the system was connected to the grid and power is going into the grid. At t=1.38 seconds the system is disconnected from the grid and current becomes zero because the system controller identifies the grid is unavailable. Figure 9 shows the load terminal voltage while the system is connected with the resistive dump load. The system is connected to the dump load at t = 8 seconds while the load was 215 ohm per phase. At t = 70 seconds, the load was changed to 140 ohm per phase which means the load was increased. As a result the terminal voltage is changed in a small amount and remains constant until the next load change. Again the load was reduced to 215 ohm which causes the terminal voltage changes by a small amount, and later kept more or less constant. The lower plot of Figure 9 shows the alternating voltage at the load terminal which is acquired by 100X probe and oscilloscope. The voltage during t = 5 to 70 seconds is about 1.69V peak which is equivalent

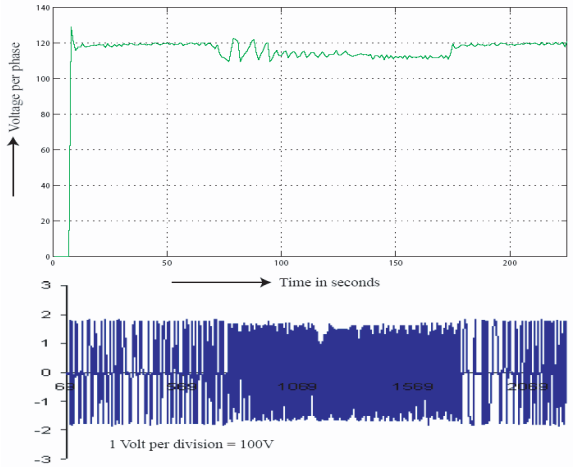


Fig. 9. Load terminal voltage due to the load variation during off-grid operation

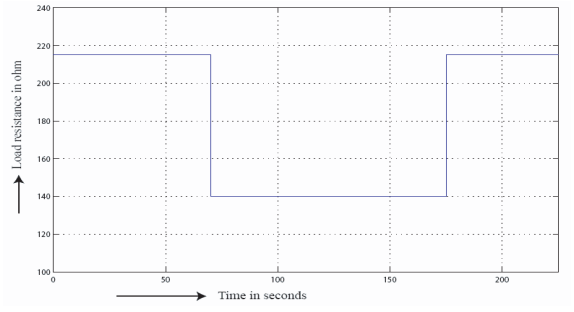


Fig. 10. Change in the load during off-grid operation

to (1.69/1.414 =1.195x100=119.5) 119.5V in each phase. Due to the change in load, although the terminal voltage is dropped by a small amount, the PI controller was able to maintain more or less constant voltage at the load terminal. The changes in load is shown in Figure 10. Figure 11 shows the PI regulator performance while there is a step change in wind speed. The load terminal voltage is constant from *t* = 0 to 65 seconds while the wind speed is 7.8 m/sec. The wind speed step change is shown in figure 12. At t = 65 seconds, the wind speed is changed to 8.15 m/sec. Due to the change in the wind speed, the load terminal voltage also increases. However, the PI regulator tries to keep the load terminal voltage more or less constant at about 120V.

VII. Conclusions

This paper describes the operation, required instrumentation, and control technique of a grid connected small WECS while grid is absent. A phase control relay based electronic controller test results are also presented. The test results show that the controller is able to regulate the voltage at the load terminal even in the load or in the wind speed variations. The proposed control strategy and developed instrumentation are simple and cost effective which is significant for a small wind turbine system.

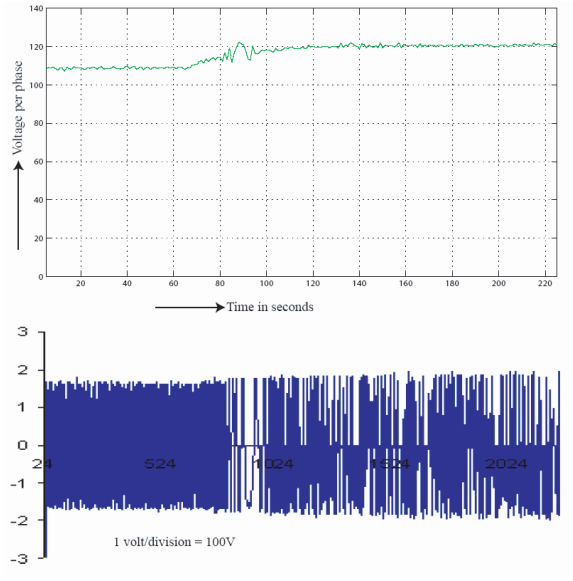


Fig. 11. Load terminal voltage due to the wind speed variation during off-grid operation

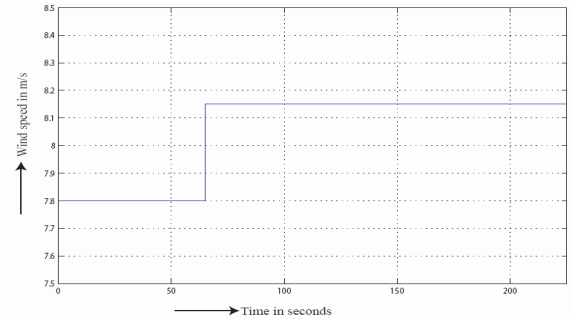


Fig. 12. Wind speed variation during off-grid operation

VIII. Acknowledgement

This work is supported by a research grant from the National Science and Engineering Research Council (NSERC) of Canada, Atlantic Innovation Funds (AIF) Canada, and Memorial University of Newfoundland.

References

[1] R. Ahshan, M. T. Iqbal, and George K. I. Mann, "Controller for a Small Induction Generator Based Wind Turbine", *Applied Energy Journal*, Vol. 85, Issue 4, pp. 218-227, April 2008.

[2] Ammasaigounden N. and M. Subbiah, "Chopper-controlled wind-driven self-excited induction generators", *IEEE Transactions on Aerospace and Electronic Systems*, AES-25(2), March 1989.

[3] M. Juan, Ramirez, and M. Emmanuel Torres, "An electronic load controller for the self-excited induction generator", *IEEE Transaction on Energy Conversion*, 22(2), June 2007.

[4] B. Singh, S. S. Murthy, and S. Gupta, "Analysis and implementation of an electronic load controller for a self excited induction generator", *IEEE Proceedings in Gener. Transm. Distrib.*, Vol. 151, January 2004.

[5] B. Singh, S. S. Murthy, and S. Gupta, "Analysis and design of electronic load controller for self-excited induction generators", *IEEE Transactions on Energy Conversion*, Vol. 21, March 2006.

[6] Tarek Ahmed, Koki Ogura, Koji Soshin, Eiji Hiraki, and Mutsuo Nakaoka, "Small- scale wind turbine coupled single-phase self-excited induction generator with svc for isolated renewable enery utilization", *Technical report*, The graduate School of Science and Engineering, Yamaguchi University, Japan.

[7] R. Ahshan, M. T. Iqbal and George K. I. Mann, "Small Induction Generator Based Wind Turbine Simulator", *IEEE 16th, NECEC conference*, St. John's, NL, 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Characteristics of the Induced Currents in Horizontal Conductors Due to a Nearby Lightning Strike

Md. Osman Goni
Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology (KUET), Bangladesh (e-mail: osman@ieee.org)

Abstract—
A nearby lightning strike can induce significant currents in long horizontal and tall down conductors. Although the magnitude of the current in this case is much smaller than that encountered during a direct strike, the probability of occurrence and the frequency content is higher. In view of this, appropriate knowledge of the characteristics of such induced currents is relevant for the interpretation of recorded currents. Considering these, the present paper discusses a modeling procedure that permits simulation of lightning-induced voltages or currents on overhead lines due to a nearby lightning strikes. The hypothesis of perfect conducting ground, generally adopted in studies on the subject, is discussed in order to better assess the validity of the simulation results. In this paper, homogeneous non-perfect ground is also investigated for the influence of lightning-induced voltage. The procedure for analyses of the voltages induced on an overhead line by a nearby lightning return stroke with a striking point in an unequal distance from the line terminations. The analysis shows that lightning induced voltages depend on the soil conductivity.

Keywords— **FDTD method, horizontal conductor, ground conductivity, induced voltage, nearby lightning strike.**

I. INTRODUCTION

A NEARBY lightning strike can induce appreciable currents in both horizontal and vertical conductors. The magnitudes of such induced currents are definitely much lower than those experienced during a direct hit. However, their frequency of occurrence is comparatively higher. Accurate knowledge of the characteristics of induced currents would help in the characterization and classification of currents recorded on instrumented conductors. Such knowledge would also be useful for the study of the electromagnetic noise/interferences caused by the induced currents on the electrical and electronic systems in the vicinity and for systems mounted on the conductors (towers). For a rough estimation of the number of strikes in the surrounding area, the information on the annual frequency of induction due to a strike in the vicinity can be used in conjunction with the number of direct hits. In view of these facts, investigations on the characteristics of the induced effects seem to be essential.

There is a large amount of literature on the problem of induced currents on conductors of electrical distribution lines and telecommunication lines [1], [2]. Detailed studies on the lightning-induced disturbances in buried electrical cables have also been carried out [3], [4]. Similarly, the induction effect in the protection systems as well as the electrical network of a building has also been considered [5], [6]. A Vertical conductor model has also been investigated [7]–[10]. However, the characteristics of the induced voltages on horizontal and vertical conductors due to lightning strikes in the vicinity and with different ground conductivities seems to be less studied. The present paper evaluates the basic characteristics of the induced currents in horizontal overhead lines due to a lightning hit in the vicinity under different values of ground conductivities.

II. METHOD OF ANALYSIS

Numerical electromagnetic analysis is becoming a powerful approach to analyze a transient which is hard to solve by a conventional circuit-theory based approach such as the EMTP [11]. It follows from a solution of Maxwell's equations for boundary conditions of the EM field at the surface of the conductor and the earth. However, it is still based on some idealistic hypotheses, such as homogeneous earth and ideal contact between the conductor and the soil. Additionally, only a few papers consider nonlinear phenomena [12].

Unfortunately, there is no systematically developed and reliable set of experimental data available that would serve as a standard, so we consider here the EM model as the basis for comparison.

Numerical electromagnetic analyses based on the FDTD method are effective to analyze the transient response of a large solid conductor or electrode. The accuracy of this method, applied to such an analysis, has been fully investigated in comparison with an experiment and shown to be satisfactory [13]. As this method requires long computation time and large memory capacity, the analysis is restricted to rather small spaces.

The FDTD method employs a simple way to discretize a differential form of Maxwell's equations. In the Cartesian coordinate system, it generally requires the entire space of interest to be divided into small rectangular cells and calculates the electric and magnetic fields of the cells using the disretized Maxwell's equations. As the material constant of each cell can be specified arbitrarily, a complex inhomogeneous medium can be easily analyzed.

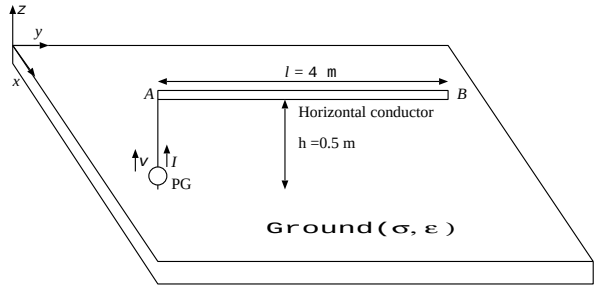


Fig. 1. Horizontal conductor arrangement (source connected at end A while the other end B is left open or terminated).

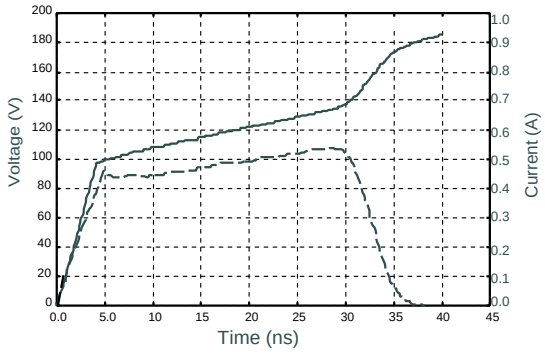


Fig. 2. Voltage and current waveforms when far-end B is left open (solid line for voltage and broken line represents current).

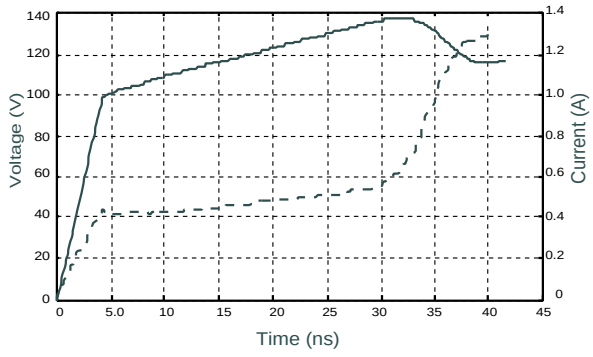


Fig. 3. Voltage and current waveforms when far-end B is terminated directly to the ground (solid line denotes voltage and broken line for current).

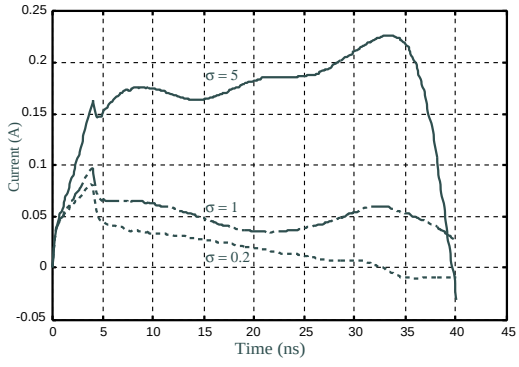


Fig. 4. Current waveforms computed at the source position of the horizontal conductor model of Fig. 1. The end B is terminated but the ground is characterized by different soil conditions.

To analyze fields in an open space, an absorbing boundary has to be set on each plane which limits the space to be analyzed, so as to avoid reflection there. In the present analysis, the second-order Mur's method [14] is employed to represent absorbing planes.

So far in most of FDTD analyses of transient and steady-state voltages, large solid electrodes [13], [15], which can be decomposed into small cubic cells, have been chosen and thin-wire electrodes have been dealt with. This is because an equivalent radius of a thin wire in a lossy medium has already been developed [16], [17]. In the present paper, an equivalent radius for a thin wire in lossy medium is utilized with the help of the concept proposed for an aerial thin wire [16]. The validity is already tested by comparing grounding-resistance values obtained through FDTD simulations on simple buried structures with theoretical values [17], [18].

A. Models for Analysis

As an electromagnetic field produced by lightning is basically responsible for the current induction, a model to be employed for study must be based on the electromagnetic model. Thus, the present paper employs the electromagnetic model, which ensures reliable description of the associated field problem and has been successfully employed in the literature for the estimation of currents and fields in the vicinity [19], [20], [21].

In measuring a transient response of a horizontal electrode, a horizontal current lead wire and a horizontal voltage reference wire have been used [6], [13], [22], [23], although it is desirable to place the horizontal conductors perpendicular to one another in order to reduce undesired inductions. Recently, Tsumura et al. [18] recommended that the perpendicular arrangement of the voltage reference wire is an appropriate one. However, the difference in the evaluated voltage peaks due to wire arrangements is only 6%.

Figure 1 shows a representative arrangement of the horizontal conductor system in which AB is considered to be a horizontal copper conductor of 4 m in length. The near end A is connected to a pulse generator which is placed at the bottom of the ground plane. The length of the vertical current lead wire is 45 cm. The remote end B is either terminated or left open.

The conductor system is excited by a current pulse generator (PG) placed at the bottom of terminated end A via a current lead wire which has an internal impedance of 50 Ω in series. The arbitrary voltage source produces a steep-front wave having a risetime of 4.1 ns to 119 V. The voltage waveform is sustained another 40 ns with a slow rise of the voltage to 180 V. Then it goes to zero [16]. The PG was modeled as a z-directional voltage source, of which the waveform was given by a piecewise linear approximation of its open voltage as in Fig. 2. The

source waveform is assigned in such a way as to allow the propagation time through the entire horizontal and other associated conductor system. As the time taken for the reflection from the open end to the measuring position is observed to be (450 cm*2)/(30 cm/ns) = 30 ns.

For the present FDTD simulation, the conductor system shown in Fig. 1 is surrounded by a large rectangular analysis space of 2 m, 6 m, and 2 m in the x, y and z directions respectively, with space length $\Delta s = 5$ cm. An earth is placed at the bottom of the analysis space with a thickness of 10 cm and a resistivity $\rho = 1.69 \times 10^{-8}\,\Omega - \mathrm{m}$. The gap length is maintained as the space length Δs of the conductor system at which a voltage probe or current probe is placed to record the voltage and current. The time step for the simulation was determined by (14) of [16] with $\alpha = 0.001$, and all the six boundaries of the cell were treated as second-order Liao's absorbing boundaries. The radius of the horizontal thin wire was taken into account by the method discussed in the previous paper and $0.23\Delta s = 1.15$ cm of radius was chosen accordingly [16], [17].

The FDTD method is normally a time-consuming method. However, the progress of computers in terms of speed and memory is considerable, and even a personal computer can be used for the FDTD calculation here. In fact, the simulations presented in this paper were performed by a personal computer with Intel Pentium 4, 2.80 GHz CPU and 512 MB RAM. Responses are calculated up to 40 ns for the reduced-scale model (2 m × 6 m × 2 m) and up to 9 μs for the actual model (100 m × 2100 m × 100 m model in Fig. 8) with a time increment of 0.096 ns and 9 ns respectively. Therefore, the computation time for the above two different scale models are about 1 min and 3 min respectively, regardless of ground parameters.

B. Analyzed Results

Computed voltage and current are shown in Fig. 2 as are computed at the injected point of PG for the horizontal conductor model of Fig. 1. Fig. 3 shows the results for the horizontal conductor model in which the far end B is also terminated to the ground. In both cases, the reflection of voltage and current waveforms are observed to be exactly at 30 ns because the model is characterized by a lossless and uniform line where the traveling wave propagates through the conductor at the velocity of light.

To this point, we have presented the analysis of current and voltage waveforms for a ground having infinite conductivity. The effect of finite ground on the resultant overvoltage also needs to be analyzed. Figure 4 shows the resultant current recorded at the injection point influenced by the finite ground conductivities showing different soil conditions and with a relative permittivity of 10.

From Figs. 3 and 4, the current waveforms show that as conductivity increases, current increases. Time-to-crest is the same in every case and the current drops at the lower soil conductivity (namely, 1 mS/m or less) and continues decreasing but increasing gradually at higher conductivity (above 5 mS/m) until reflection from the short-circuited end reaches the measuring point. Hence, the high resistivity of the soil blocks the current flowing into the soil and forces the current to move toward the terminal end of the electrode.

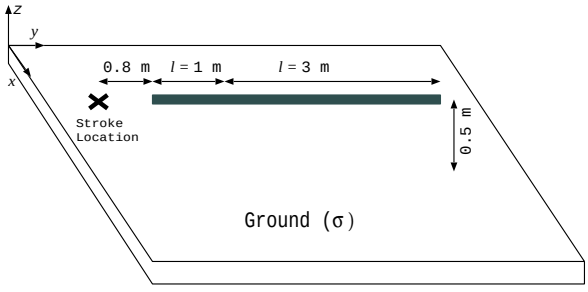


Fig. 5. Horizontal conductor arrangement (source is connected at 0.8 m from the left end simulating a nearby lightning strike).

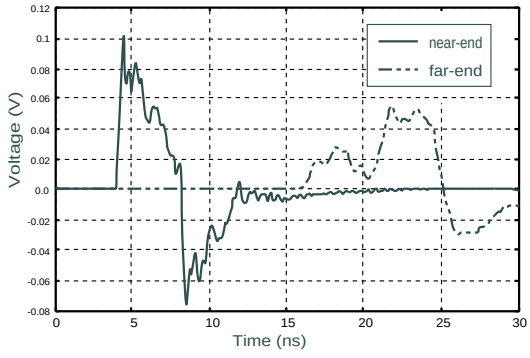


Fig. 6. Near-end and far-end induced voltages on the horizontal conductor model of Fig. 5, computed at the perfectly conducting ground.

III. INDUCED VOLTAGE DUE TO NEARBY LIGHTNING STRIKE

In this section, reduced scale model of a horizontal conductor system, as shown in Fig. 5, is presented to analyze the induced effects due to a nearby lightning strike. The same step-like z-directional current pulse considered for the model of Fig. 1 is also applied here at 50 cm above the ground surface instead of placing to the ground (i.e., at the same height as the horizontal conductor) and 0.8 m from the left end of the horizontal conductor and simulates as indirect lightning hit. Unlike the experimental setup by Pokharel et.al. [24], we did not consider an elevated lightning channel as we only considered the effect of a first stroke rather than subsequent strokes and the return stroke which could be initiated from the elevated lightning channel. The same analysis space is chosen with initially a copper plate of thickness 10 cm simulating perfectly conducting ground.

The induced-voltages are computed at the ground level along with the horizontal conductor and at 0.8 m and

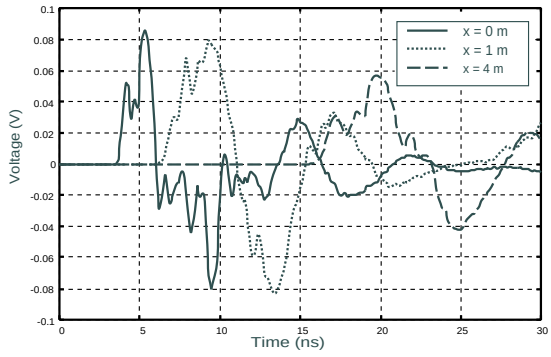


Fig. 7. Induced voltages at three different positions along the horizontal conductor with respect to the copper ground (lightning impulse injected at 0.8 m left of the horizontal conductor).

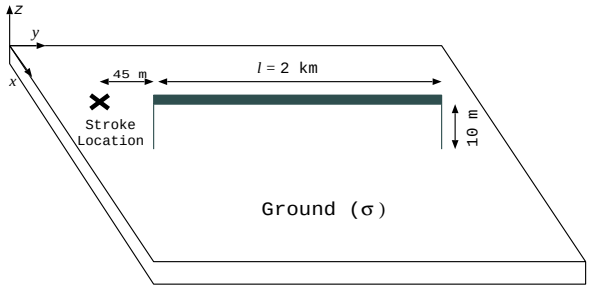


Fig. 8. Large-scale horizontal conductor arrangement (stroke location is at the 45 m from the left end simulating nearby lightning strike).

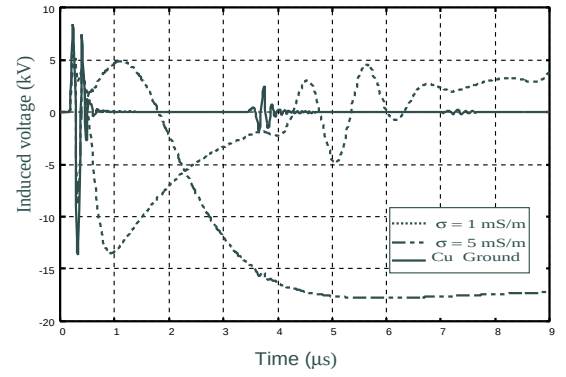
4.8 m from the injection point. In this paper, these voltages are treated as near-end and far-end voltages respectively. The two voltage waveforms of Fig. 6 are in the same z-directional parameter as the injected voltage and measured at the bottoms of the vertically terminated ends with respect to the copper ground. Figure 7 shows the voltages induced on the horizontal conductor which are computed along this conductor with respect to the auxiliary voltage measuring wire in three different locations. Both Figs. 6 and 7 also demonstrate that the peaks of these induced-voltages decrease as we move far away from the stroke location.

TABLE I

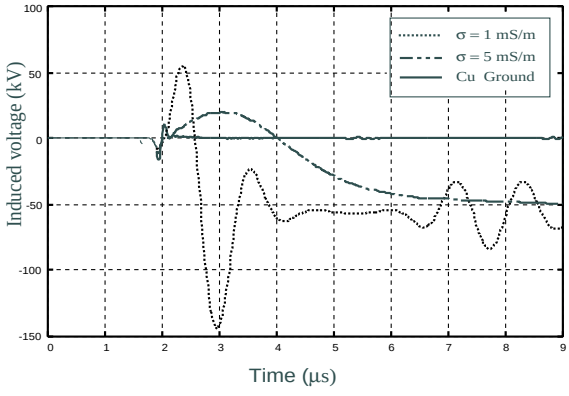
EFFECT OF THE GROUND CONDUCTIVITIES ON THE INDUCED VOLTAGE ON THE HORIZONTAL CONDUCTOR AT DIFFERENT DISTANCES FROM THE STRIKE POINT (REDUCED-SCALE MODEL)

Distance	σ=5.8E7 S/m	σ=5 mS/m	σ=1 mS/m	σ=0.2 mS/m
(meter)	(mV)	(mV)	(mV)	(mV)
0.8	53	53	53	53
1.8	30.5	28.2	28.2	28.2
4.8	26.9	9.35	7.53	4.2

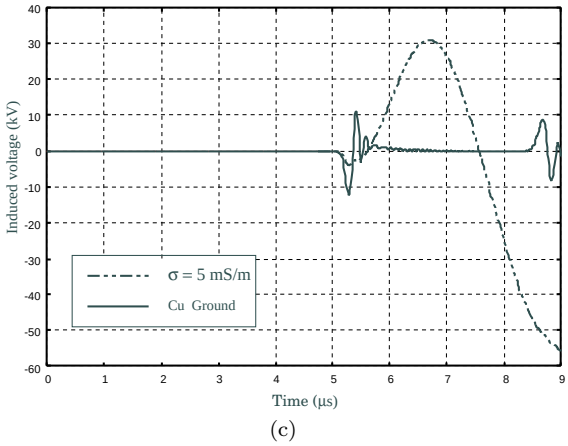
Table I shows the effect of the ground conductivities on the induced voltage s at different horizontal distances from the strike location. The waveforms of those induced voltages are observed to be similar to Fig. 7 for



(a)



(b)



(c)

Fig. 9. Induced voltage at a different distances from the left end of the horizontal conductor model of Fig. 8; (a) l = 10 m, (b) l = 500 m and (c) l = 1500 m).

the model of Fig. 5 with varying soil conditions. Near the strike point, the magnitude of induced voltage does not have significant dependence on the ground conductivity. But the magnitude of the induced voltage obviously depends on the horizontal distance from the strike point regardless of the ground conductivity. Pokharel et. al.[24] shows that the induced voltage will be lower for a perfect conductor than for a finite ground. This is because of the location of simulated strike is at the ground and thus the ground conductivity affects not only the shape of induced voltage but also its magnitude. In this paper,

we consider a strike location parallel to the horizontal conductor and at a height of 50 cm from the ground surface. This type of lightning may be attributed to a downward leader and strikes in an open field where personnel or other small objects (e.g., umbrella, golf club, fishing rod, etc.) can be exposed to a direct hit. At the same time, the induced voltage or current in the vicinity of the lightning path cannot be ignored and should be considered as serious issues like a direct hit.

Due to the computation time and large memory storage, a small-scale model has been chosen for the analysis of induced effects due to indirect hits. It is also necessary to investigate those effects with a full-sized analysis space in order to show the length dependent parameters. Non-linear effects such as soil ionization have not been included in this investigation. Fig. 8 describes the full-scaled model of a horizontal conductor of 2 km in length and radius of 1.15 m. The rectangular analysis space of 100 m, 2100 m, and 100 m in the x, y and z directions respectively, with space length $\Delta s = 5$ m has been simulated. The radius of the wire and space length of the simulated area are chosen according to the guidelines of [16] and also to save memory requirements and computation time in a personal computer. The ground thickness was 10 m and it was modeled with different ground conductivities. The step-like current pulse having a magnitude of 60 kA and a maximum time derivative of 600 kA/μs has been considered to simulate the indirect lightning stroke current. The source is injected at 45 m from the horizontal conductor and at 10 m above the ground surface.

Fig. 9 shows the induced voltages at different distances in between the near-end to far-end of the horizontal conductor with respect to auxiliary voltage measuring wire. The simulation has been carried out by the FDTD method, postulating homogeneous ground with conductivities of 1 mS/m and 5 mS/m and relative permittivity of 12 corresponding to different soil conditions [17]. The responses can be calculated up to 9 μs with a time increment of 9.62 ns because of temporary storage limitations. At the closer-end, the voltage waveforms started with positive peaks but at the distant-end the waveforms are negative because the polarity of the horizontal electric field is dependent on the distance from the striking point [25]. In Fig. 9(c), the waveform for $\sigma = 1$ mS/m is not shown because the positive peak is so high that it could not be drawn in the same scale. The magnitude of the first peak is given in the Table II for comparison.

In such a horizontal conductor and striking point arrangement, where the striking point is on the extension of the line, the induced-voltage increases with increasing ground conductivity which is different from the result obtained in [25]. This is due to the arrangement of the lightning source in this present research which is simulated above the ground surface rather than at the ground surface. Table II summarizes the above results.

TABLE II

Effect of the ground conductivities on the induced voltage on the horizontal conductor at different distances from the strike point (large-scale model)

Distance	σ=5.8E7 S/m	σ=5 mS/m	σ=1 mS/m	σ=0.2 mS/m
(meter)	(kV)	(kV)	(kV)	(kV)
10	8.04	6.4	5.25	4.3
500	-15.8	-11	-8.36	-6.86
1500	-12.3	-4.05	-2.08	-1.07

IV. CONCLUSIONS

The lightning-induced voltage on an overhead horizontal wire, with lossy and lossless ground, associated with a first stroke above the earth surface and at a distance of about 2 km from the injection point, is studied based on FDTD simulations. The following conclusions are made.

(1) By comparing the current waveforms calculated at the excitation point, the induced voltage or current increases with increasing ground conductivity corresponding to different soil conditions.

(2) In a nearby lightning strike, the induced voltages on the horizontal overhead wire are greatly affected by the distance from the striking point. The lightning-induced voltage at the near-end is higher than at the far-end.

(3) It is also shown that the ground conductivity influences the induced voltage. The lightning-induced voltage becomes larger with the increase of the ground conductivity. This is valid when the injection is above the ground surface rather than at the ground.

V. ACKNOWLEDGMENT

The authors are indebted to the CASR, Khulna University of Engineering and Technology for the research grant to carry out this research work.

References

[1] A. Borghetti, J.A. Gutierrez, C.A. Nucci, M. Paolone, E. Petrache, and F. Rachidi, "Lightning-induced voltages on complex distribution systems:Models, advanced software tools and experimental validation," *J. Electrostat.,,* vol. 60, pp. 163–174, March 2004.

[2] M.G. Sorwar, H. Ahmed, and M.M. Ali "Analysis of transients in overhead telecommunication subscriber line due to nearby lightning return stroke," *in Proc. IEEE Int. Symp. Electromagn. Compat.* vol. 2, pp. 1083–1088, Aug. 1998.

[3] L. Grcev and F. Dawalibi, "An electromagnetic model for transients in grounding systems," *IEEE Trans. Power Delivery,* vol. PWRD-5(4), pp. 1773–1781, 1990.

[4] Y. Liu, M. Zitnik and R. Thottappillil, "A time domain transmission line model of grounding systems," *Proc. Int. Symp. High Voltage Engineering,* pp. 154–157, 2001.

[5] Y. Liu, M. Zitnik and R. Thottappillil, "An improved transmission-line model of grounding systems," *IEEE Trans. Electromagn. Compat.,* vol. 4, no. 3, pp. 348–355, Aug. 2001.

[6] S. Sekioka, H. Hayashida, T. Hara, and A. Ametani, "Measurements of grounding resistances for high impulse current," *Proc. Inst. Elect. Eng. Gen. Transm. Ditrib.,* vol. 145, no. 6, pp. 693–699, Nov. 1998.

[7] M. O. Goni, and H. Takahashi, "Thin wire representation of the vertical conductor in surge simulation," *The Applied Computational Electromagn. Society (ACES) Journal*, vol. 18, no. 1a, pp. 41–47, Mar. 2004.

[8] M. O. Goni, P. T. Cheng and H. Takahashi, "Theoretical and experimental investigations of the surge response of a vertical conductor," in *Proc. IEEE Power Engineering Soceity Int'n Conf.*, vol. 2, pp. 699–704, 2002.

[9] M. O. Goni, and H. Takahashi, "Theoretical and experimental investigations of the surge response of a vertical conductor," *The Applied Computational Electromagn. Society (ACES) Journal*, vol. 18, no. 1, pp. 41–47, Mar. 2003.

[10] M. O. Goni, M. F. Hossain, M. M. Rahman, M. S. Yusuf, E. Kaneko, H. Takahashi, "Simulation and experimental analyses of electromagnetic transient behaviors of lightning surge on vertical conductors," *IEEE Trans. Power Delivery*, vol. 21, No. 4, October 2006.

[11] Working group of IEEJ, "Numerical electromagnetic analysis and its application to power system transients," *IEEJ WG Report,* 2003–2007.

[12] T. Mozumi, Y.Baba, M.Ishii, N.Nagaoka and A.Amatani "Numerical electromagnetic field analysis of archorn voltages during a back-flashover on a 500 kV twin circuit line," *IEEE Power Delivery,*, vol. 18, no. 1, pp. 207–213, 2003.

[13] K. Tanabe, "Novel method for analyzing dynamic behavior of grounding systems based on the FD-TD method," *IEEE Power Eng. Rev.,*, vol. 21, no. 9, pp. 55–57, Sep. 2001.

[14] G. Mur, "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equation," *IEEE Trans. Electromagn. Compat.*, vol. EMC-23, no. 4, pp. 377–382, 1981.

[15] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equation in isotropic media," *IEEE Trans. Antennas and Propagation,* vol.AP-14(4), pp. 302–307, 1966.

[16] T. Noda and S. Yokoyama, "Thin wire representation in finite difference time domain surge simulation," *IEEE Trans. Power Delivery*, vol. 17, no. 3, pp. 840–847, 2002.

[17] Y.Baba, N. Nagaoka, and A. Ametani, "Numerical analysis of groundng resistance of buried thin wires by the FDTD Method," *Int. Conf. on Power Syst. Trans(IPST)*, New Orleans, USA, 2003.

[18] M. Tsumura, Y. Baba, N. Nagaoka and A. Ametani, "FDTD simulation of a horizontal grounding electrode and modeling of its equivalent circuit," *IEEE Trans. Electromagn. Compat.,* vol. 48, no. 4, pp. 817–825, Nov. 2006.

[19] Y. Baba and M. Ishii, "Numerical electromagnetic field analysis of lightning current in tall structures," *IEEE Trans. PWRD,* vol. 16, no. 2, pp. 324–328, Apr. 2001.

[20] Y. Baba and M. Ishii, "Characteristics of electromagnetic return-stroke models," *IEEE Trans. Electromagn. Compat.,* vol. 45, no. 1, pp. 129–134, Feb. 2003.

[21] B. Kordi, R. Moini, W. Janischewskyj, A.M. Hussein, V.O. Shostac, and V.A. Rakov, "Application of the antenna theory model to a tall tower struck by ligtning," *J. Geophys. Res.,* vol. 108, pp. ACL 7/1–ACL7/9, no. D17 4542, 2003.

[22] A. Ametani, N. Nagaoka, T. Sonoda, and S. Sekioka, "Experimental investigation of transient voltage and current characteristics on a grounding mesh," *presented at the Int. Conf. Power Syst. Trans.,* New Orleans, LA, Sep.–Oct. 2003.

[23] K. Tanabe, A. Asakawa, T. Noda, M. Sakae, M. Wada and H. Sugimoto, "Varifying the novel method for analyzing transient grounding resistance based on the FD-TD method through comparison with experimental results," *CRIEPI Report*, no.99043, 2000. (in Japenese)

[24] R.K. Pokharel, M.Ishii, and Y.Baba, "Numerical electromagnetic analysis of lightning-induced voltage over ground of finite conductivity," *IEEE Trans. Electromagn. Compat.,* vol. 45, no. 4, pp. 651–656, Nov. 2003.

[25] M.Ishii, K.Michishita, and Y.Hongo, "Experimental study of lightning- induced voltage on an overhead wire over lossy ground," *IEEE Trans. Electromagn. Compat.,* vol. 41, no. 1, pp. 39–45, Feb. 1999.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Self-Excited Single-Phase and Three-Phase Induction Generators in Remote Areas

M. H. Haque, *Senior Member, IEEE*
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore 639798

***Abstract*--Self-excited induction generators are increasingly being used in remote areas to generate electrical power from both conventional and nonconventional energy sources. This paper describes a method of evaluating the steady state characteristics of a single-phase as well as a three-phase induction generator for stand-alone operation. The steady state problem is formulated in a very general way so that the same solution technique can be applied to both single-phase and three-phase generators. A numerical based routine is then used to solve the formulated problem and that eliminates the additional algebraic calculations needed to explicitly express the equations in terms of actual unknowns. The proposed method is then applied to evaluate the steady state characteristics of a number of single-phase and three-phase induction generators. The simulation results obtained by the proposed method are also compared with the corresponding experimental values and are found to be in very good agreement.**

***Index Terms*-- Self-excited induction generators, stand alone operation, renewable energy sources, micro-grid.**

I. Introduction

It is an established fact that an ordinary induction motor can operate as a generator when it is driven by a prime mover and a capacitor bank of appropriate size is connected across its stator terminals. Such a machine is called self-excited induction generator (SEIG). The purpose of using the capacitor bank is to provide adequate reactive power needed for excitation of the generator. Alternatively, the generator can be connected to a grid supply to draw reactive power from the grid. The voltage and frequency of a grid connected induction generator are fixed and determined by the grid. Thus, the analysis of a grid connected generator becomes much easier. On the other hand, the voltage and frequency of a SEIG are not fixed but depend on many factors, such as speed, magnetization characteristic, excitation capacitor, load, etc. and that is why the analysis of a SEIG is much more difficult than that of a grid connected generator.

The increasing concern of the environment, especially the greenhouse effects, has motivated the world towards exploring the use of renewable energy sources and reduces dependency on fossil fuels. Most of the renewable energy sources, such as wind, mini-hydro, etc. are usually available in remote areas. This leads to build renewable energy sources based small power plants in remote areas where the traditional grid supply may be well out of reach in developing countries. Such plants can form a micro-grid to supply power to isolated communities. In addition, recent developments in transportation and telecommunication systems initiated people of developed countries to move away from large urban concentration and live in less populated areas. This again leads to build small power plants in remote areas for isolated communities.

A self-excited induction generator is found to be very suitable to generate electrical power in remote areas from renewable energy sources, such as wind and mini-hydro turbines [1], [2]. A SEIG has many advantageous features over its counterpart synchronous generator [3]. These features are low cost, high reliability, maintenance and operational simplicity, rugged construction, brushless operation, protection against overloads and short circuits, etc. Even though a SEIG is very suitable for wind and mini-hydro plants, it can also efficiently be used with prime movers driven by other energy sources, such as diesel, biogas, natural gas, gasoline, etc.

The steady state and dynamic performances of a SEIG are investigated by a large number of researchers and are well summarized in recent articles [3]-[5]. In most cases, three-phase generators are used. However, loads in remote areas are mainly single-phase residential type. For such a case, generation of single-phase power is more appropriate. Many researchers used three-phase induction machines to generate single-phase power using a single excitation capacitor [6]-[8]. Alternatively, a single-phase induction machine can be used to generate a single-phase power [9]-[12]. Most of the single-phase induction machine has two windings (main and auxiliary) but it is not necessary to use both the windings for generator operation. For a single-winding operation, both the excitation capacitor and the loads are connected to the main winding and the auxiliary winding is kept open. On the other hand, the excitation capacitor can be connected to the auxiliary winding and the loads can be connected to the main winding to share the current burden by both the windings [11], [12].

Determination of steady state performance of a single-phase as well as a three-phase induction generator usually requires the solution of two simultaneous nonlinear equations [8], [10], [13]-[15]. Derivation of the equations involves lengthy and tedious algebraic manipulations. In addition, the equations are usually solved by the Newton Raphson method that requires partial derivatives of the equations. However, most of the

algebraic manipulations needed in deriving the equations and their partial derivatives can be avoided if a numerical based routine is employed in solving the equations.

This paper describes a simple method of evaluating the steady state characteristics of self-excited single-phase and three-phase induction generators. The problem is formulated in a general way so that the same solution technique can be applied to both the single-phase and the three-phase generators. The formulated problem is then solved using a numerical based routine to avoid lengthy step-by-step algebraic derivations. The effectiveness of the proposed method is then evaluated on a number of three-phase and single-phase induction motors operated as generators. The simulation results obtained by the proposed method are also compared with the corresponding experimental values.

II. Mathematical Model

In a SEIG, the excitation capacitor reactance and other inductive reactances (magnetizing reactance, load reactance, stator and rotor leakage reactances, etc.) form a resonance circuit to generate voltage. This is essential for generator operation and for such an operation, the slip is negative and thus the resistance of the rotor circuit is also negative. In other words, the rotor circuit delivers active power which is taken from the prime mover. In fact, the negative rotor circuit resistance and other positive resistances (load resistance, core loss resistance, stator resistance, etc.) form another special resonance circuit to maintain active power balance in the circuit. The above principles are fully exploited in analyzing the steady-state performance of an induction generator through its equivalent circuit. The equivalent circuit of a three-phase and a single-phase SEIG as well as the respective equations that dictate the steady state performance of the generators is described in the following.

A. Equivalent Circuit of a Three-Phase SEIG

The per-phase equivalent circuit of a three-phase SEIG with an excitation capacitor and an R-L load is shown in Fig. 1 where R_1, X_1, R_2, X_2, R_c and X_m represent the stator resistance, stator leakage reactance, rotor resistance, rotor leakage reactance, core loss resistance and magnetizing reactance, respectively, of the generator. F and ω represent the per unit (pu) frequency and speed, respectively. X_C and $Z_L\angle\theta = (R_L+jX_L)$ represent the excitation capacitor reactance and the load impedance, respectively. Note that the circuit of Fig. 1 is normalized to the base frequency by dividing all parameters and voltages by F [1]. The magnetizing reactance X_m of the generator is considered as variable and it depends on the core flux or the ratio of air-gap voltage to frequency (V_g/F). In this study, V_g/F is expressed by the following third polynomial of X_m [16].

$$\frac{V_g}{F} = \sum_{i=0}^{3} k_i X_m^i \tag{1}$$

The coefficients *k*'s of the polynomial can be determined from synchronous speed test data [17]. The circuit of Fig. 1 can be represented by three series impedances ($\bar{Z}_1$, $\bar{Z}_2$ and $\bar{Z}_3$) as shown in Fig. 2. The impedances are given by

$$\bar{Z}_1 = (R_1/F + jX_1) \tag{2.1}$$

$$\bar{Z}_2 = \left(\frac{1}{R_c/F} + \frac{1}{jX_m} + \frac{1}{R_2/(F-\omega) + jX_2}\right)^{-1} \tag{2.2}$$

$$\bar{Z}_3 = \left(\frac{1}{-jX_C/F^2} + \frac{1}{R_L/F + jX_L}\right)^{-1} \tag{2.3}$$

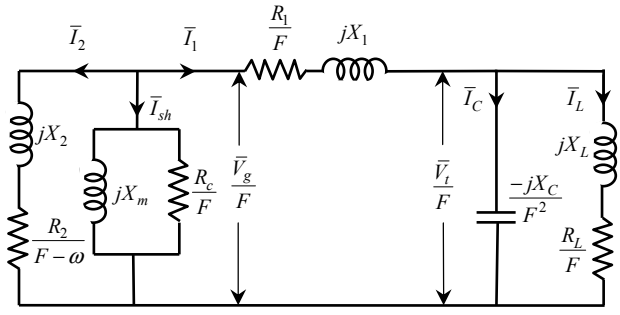


Fig. 1 Per-phase equivalent circuit of a three-phase SEIG.

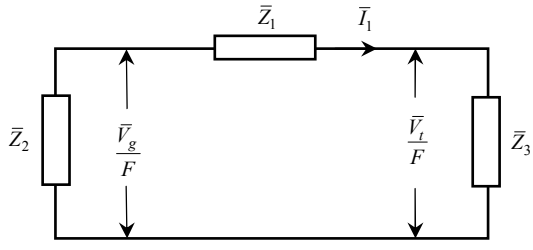


Fig. 2 Simplified representation of Fig. 1.

The loop equation in Fig. 2 is

$$\bar{I}_1 \bar{Z}_{loop} = 0 \tag{3}$$

Here $\bar{Z}_{loop} = (\bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3)$. Since the stator current $\bar{I}_1$ is not zero under normal operating condition, the loop impedance $\bar{Z}_{loop}$ must be zero.

B. Equivalent Circuit of a Single-Phase SEIG

The equivalent circuit of a single-phase SEIG (with an excitation capacitor and an R-L load) is shown in Fig. 3 [12]. Note that both the excitation capacitor and the loads are connected to the main winding of the generator. The definition of various parameters of the generator in Fig. 3 is exactly the same as that for a three-phase generator. In this case, the core loss resistance R_c is neglected. For a single-phase generator, the ratio of forward air-gap voltage to frequency (V_g/F) can also be expressed by a similar polynomial of (1). The circuit of Fig. 3 can be represented by four series impedances ($\bar{Z}_1$, $\bar{Z}_{2b}$, $\bar{Z}_{2f}$ and $\bar{Z}_3$) as shown in Fig. 4. The impedances are given by

$$\bar{Z}_1 = R_1/F + jX_1 \tag{4.1}$$

$$\bar{Z}_{2b} = \left(\frac{1}{jX_m/2} + \frac{1}{R_2/2(F+\omega) + jX_2/2} \right)^{-1} \quad (4.2)$$

$$\bar{Z}_{2f} = \left(\frac{1}{jX_m/2} + \frac{1}{R_2/2(F-\omega) + jX_2/2} \right)^{-1} \quad (4.3)$$

$$\bar{Z}_3 = \left(\frac{1}{-jX_C/F^2} + \frac{1}{R_L/F + jX_L} \right)^{-1} \quad (4.4)$$

Again the loop equation in Fig. 4 is

$$\bar{I}_1 \bar{Z}_{loop} = 0 \quad (5)$$

Here $\bar{Z}_{loop} = (\bar{Z}_1 + \bar{Z}_f + \bar{Z}_b + \bar{Z}_{ext})$. As mentioned before, the stator (or main winding) current $\bar{I}_1$ under normal operating condition is not zero and thus the loop impedance $\bar{Z}_{loop}$ must be zero.

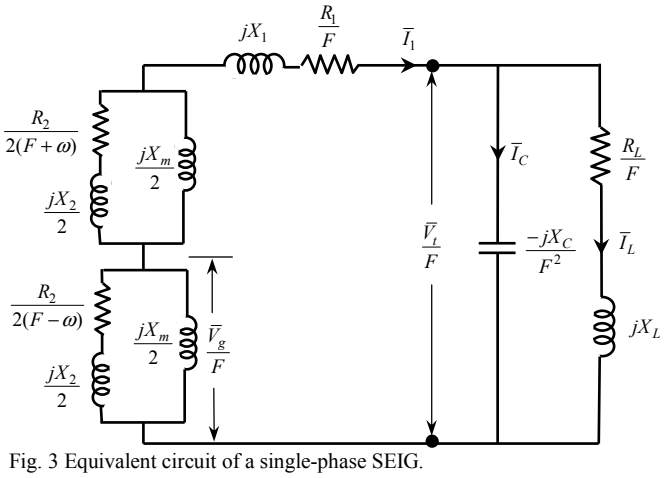


Fig. 3 Equivalent circuit of a single-phase SEIG.

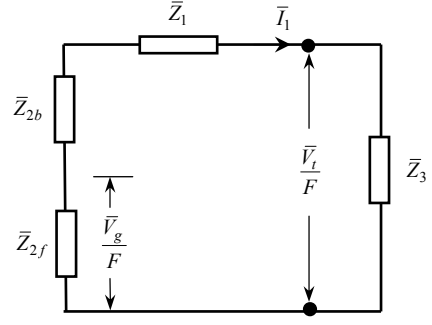


Fig. 4 Simplified representation of Fig. 3.

III. Solution Technique

For both the single-phase and the three-phase induction generators, the ultimate equation that needs to be solved to obtain the steady state performance is

$$\bar{Z}_{loop} = 0 \quad (6)$$

By separating the real and imaginary parts of the above complex equation, the following two scalar equations can be obtained.

$$g_1 = real(\bar{Z}_{loop}) = 0 \quad (7.1)$$

$$g_2 = imag(\bar{Z}_{loop}) = 0 \quad (7.2)$$

In general, (7.1) and (7.2) are solved to find the values of X_m and F for given values of ω, X_C and Z_L. When X_m and F are considered as independent variables, the above equations in general form can be written as

$$G(x) = 0 \quad (8)$$

Here $G = [g_1 \;\; g_2]^T$ and $x = [X_m \; F]^T$. In most of the previous methods, (7.1) and (7.2) are expressed in terms of X_m and F through some lengthy and tedious algebraic manipulations [8], [10], [12]-[15]. The equations are then solved by the Newton Raphson method that requires partial derivatives of the equations. In this study, (8) is solved using a numerical based routine "fsolve" given in the optimization toolbox of MATLAB. The routine does not require expressing the equations in terms of X_m and F (explicitly) and thus simplifies the problem formulation. In addition, the expressions for partial derivatives of the equations are not needed.

By knowing the values of X_m and F from (8), the air-gap voltage V_g can be evaluated from (1). Once V_g is known, the steady sate performance of the generators can easily be determined through their respective equivalent circuits [12]-[15].

IV. Results and Discussions

The proposed method of evaluating the steady state characteristics of a single-phase and a three-phase SEIG is tested on the following five induction motors operated as generators.

a. Three-phase, delta-connected, 220-V, 50-Hz, 1.5-kW, 4-pole, squirrel-cage induction machine
b. Three-phase, delta-connected, 220-V, 50-Hz, 1.5-kW, 4-pole, wound-rotor induction machine
c. Three-phase, delta-connected, 220-V, 50-Hz, 1.0-hp, 4-pole, squirrel-cage induction machine
d. Single-phase, 220-V, 50-Hz, 1.5-hp, 4-pole, induction machine
e. Single-phase, 220-V, 50-Hz, 1.0-hp, 4-pole, induction machine

It is assumed that the generators are driven at a constant speed of 1.0 pu (or 1,500 rpm). Assuming a constant speed is not surprising because in most cases the generators operate with some forms of speed control turbines or prime movers. In the laboratory, three-phase, 4-pole synchronous motors are used to run the generators at a constant speed of 1,500 rpm. Results of only one of the three-phase generators and one of the single-phase generators are briefly described in the following.

A. Three-Phase SEIG

The experimental setup of a three-phase SEIG is shown in Fig. 5 where the generator is driven by a synchronous motor. A number of small switched capacitor banks are used to vary the excitation capacitor of the generator. Various quantities of the generator are measured by using a multi-instrument MIC. It is a versatile microprocessor-based measuring unit providing measurement of all electrical quantities (voltage,

current, frequency, active, reactive & apparent power, etc.) of a single-phase as well as a three-phase system. The constant parameters of the generator (generator-*a*) are given in the Appendix.

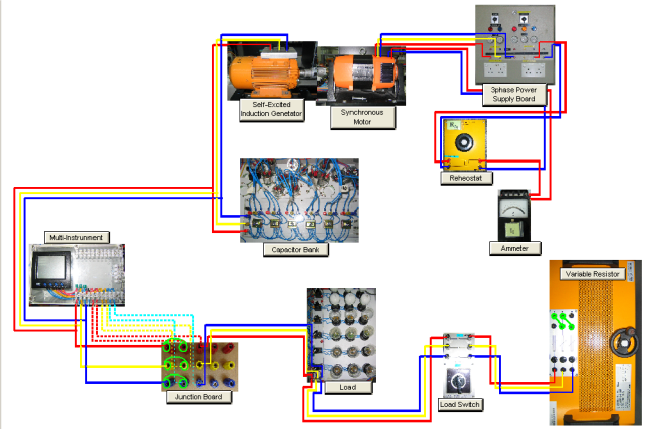


Fig. 5 Experimental setup of a three-phase SEIG.

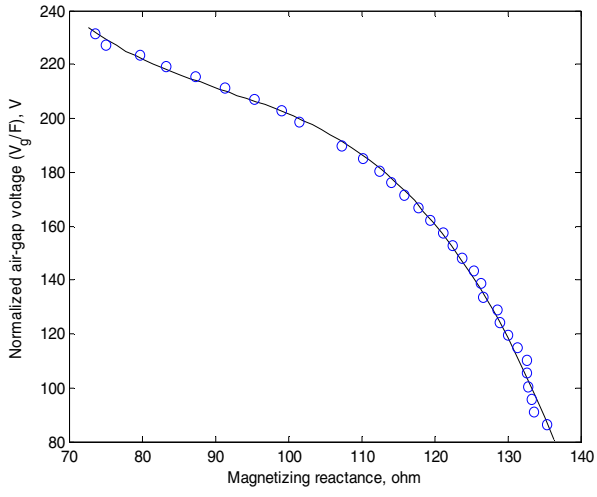


Fig. 6 Magnetization characteristic of the three-phase generator.
'—' obtained from (1); 'o' obtained from synchronous speed test data

First the magnetization characteristic (i.e. V_g/F vs. X_m curve) of the generator is evaluated from the synchronous speed test data. The "polyfit" routine given in MATLAB is used to find the coefficients of (1) which are also given in the Appendix. Figure 6 shows the magnetization characteristic of the generator and it indicates that the results obtained through (1) (shown by a solid line) are in very good agreement with the corresponding synchronous speed test data (shown by a symbol 'o'). The load characteristics (load voltage vs. load power curve) of the generator are then evaluated for three different values of excitation capacitors (33 µF, 36 µF and 39 µF per phase) and are shown in Fig. 7. For simplicity, the load power factor is considered as unity. Figure 7 indicates that the load voltage initially decreases with load until the maximum power is reached. When the load impedance is decreased further (beyond the maximum power), both the load voltage and load power decrease (see Fig. 7) and such an operation is called an unusual or unstable operation [18]. However, the upper part of the characteristics (when load voltage decreases with load) is called normal operation or stable operating region. It can also be noticed in the figure that the maximum power delivery capacity of the generator depends on the value of the excitation capacitor used and it increases with the increase of excitation capacitor. However, use of higher capacitors may cause over voltage at no-load or light load condition. Figure 7 also indicates that the experimental results are very close to the corresponding simulation results.

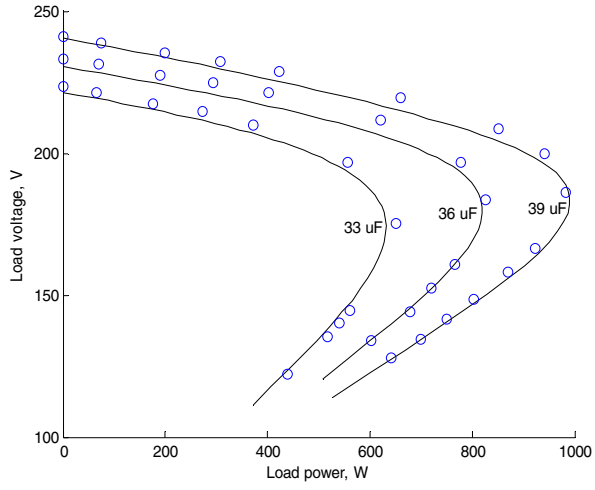


Fig. 7 Load characteristic of the three-phase generator.
'—' simulation results; 'o' experimental results

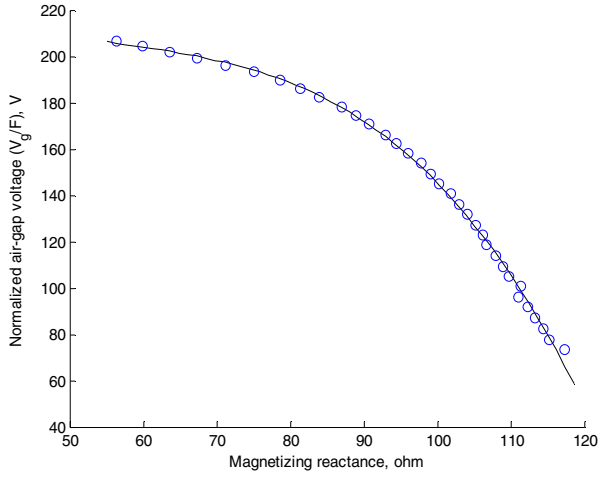


Fig. 8 Magnetization characteristic of the single-phase generator.
'—' obtained from (1); 'o' obtained from synchronous speed test data

B. Single-Phase SEIG

The fixed parameters of a single-phase induction generator (generator-*e*) including the coefficients of its magnetization characteristic are given in the Appendix. Figure 8 shows the magnetization curve of the generator and it again indicates that the results obtained through (1) are very close to the corresponding synchronous speed test data. The load characteristics of the generator are then evaluated for 86.5 µF and 94 µF of excitation capacitors and are shown in Fig. 9.

Again the load power factor is considered as unity. The pattern of the load characteristics (for the single-phase generator) is found to be very similar to that of the three-phase generator. Again, the experimental results in Fig. 9 are observed to be very close to the corresponding simulation results in both the stable and the unstable operating regions.

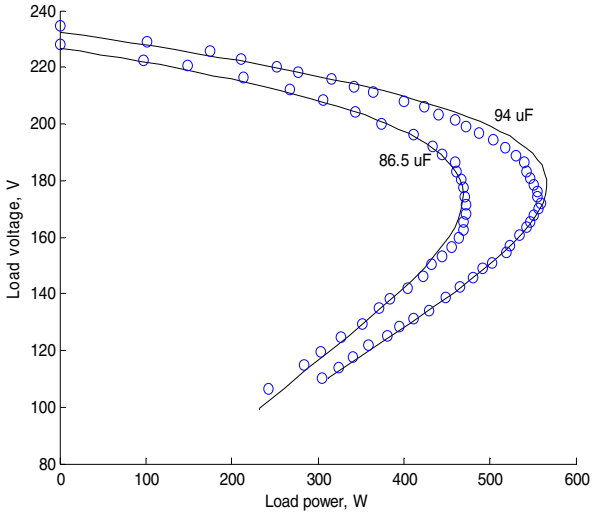


Fig. 9 Load characteristic of the single-phase generator.
'—' simulation results; 'o' experimental results

V. Conclusions

This paper formulated the steady state problem of a single-phase and a three-phase self-excited induction generator in a very general way so that the same solution technique can be applied. A numerical based routine "fsolve' given in the optimization toolbox of MATLAB is then used in solving the formulated problem to determine the values of the magnetizing reactance of the generator and the frequency of the induced voltage. The above results are then used, in conjunction with the magnetization curve, to obtain the steady sate performance of the generators through their respective equivalent circuits. The effectiveness of the proposed method is tested on two single-phase and three three-phase induction generators. Simulation results indicated that the load characteristic of a single-phase generator is very similar to that of a three-phase generator. In both cases, the maximum power delivery capacity depends on the value of the excitation capacitor used. It is also observed that the simulation results found by the proposed method are very close to the corresponding actual values obtained through experimental setups in a laboratory.

VI. Acknowledgement

The author would like to thanks a group of 20 second year students for building the necessary hardware and collecting some of the experimental data during their five-week Design and Innovation Project (DIP) supervised by the author.

VII. References

[1] M.G. Simoes and F.A. Farret, "Renewable energy systems – design and analysis with induction generators", CRC Press, New York, USA, 2004.
[2] F.A. Farret and M.G. Simoes, "Integration of alternative sources of energy", IEEE Press, New Jersey, USA, 2006.
[3] G.K. Singh, "Self-excited induction generator research – A survey", Electric Power Systems Research, Vol. 69, 2004, pp, 107-114.
[4] R.C. Bansal, "Three-phase self-excited induction generators: An overview", IEEE Trans. on EC, Vol. 20, No. 2, 2005, pp. 292-299.
[5] R.C. Bansal, T.S Bhatti and D.P. Kothari, "Bibliography on the application of induction generators in nonconventional energy systems", IEEE Trans. on EC, Vol. 18, No. 3, 2003, pp. 433-439.
[6] S.S. Murthy, B. Singh, S. Gupta and B.M. Gulati, "General steady-state analysis of three-phase self-excited induction generator feeding three-phase unbalanced load/single-pahse load for stand-alone applications", IEE Proc. Gener. Transm. Distrib., Vol. 150, No. 1, 2003, pp. 49-55.
[7] T.F. Chan and L.L. Lai, "Steady-state analysis and performance of a single-phase self-regulated self-excited induction generator", IEE Proc. Gener. Transm. Distrib., Vol. 149, No. 2, 2002, pp. 233-241.
[8] Y.H.A. Rahim, "Excitation of isolated three-phase induction generator by a single capacitor", IEE Proc.-B, Vol. 140, No. 1, 1993, pp. 44-50.
[9] O. Ojo and I. Bhat, "An analysis of single-phase self-excited induction generators: Model development and steady state calculations", IEEE Trans. on EC, Vol. 10, No. 2, 1995, pp. 254-260.
[10] Y.H.A. Rahim, A.I. Alolah and R.I. Al-Mudaiheem, "Performance of single phase induction generators", IEEE Trans. on EC, Vol. 8, No. 3, 1993, pp. 368-395.
[11] B. Singh and L.B. Shiplkar, "Steady-state analysis of single-phase self-excited induction generator", IEE Proc. Gener. Transm. Distrib., Vol. 146, No. 5, 1999, pp. 421-427.
[12] S.S. Murthy, 'A novel self-excited self-regulated single-phase induction generator", IEEE Trans. on EC, Vol. 8, No. 3, 1993, pp. 377-381.
[13] S.P. Singh, B. Singh and M.P. Jain, "Comparative study on the performance of a commercially designed induction generator with inductor motors operating as self excited induction generators", IEE Proc.-C, Vol. 140, No. 5, 1993, pp. 374-380.
[14] S.S. Murthy, O.P Malik and A.K. Tandor, "Analysis of self-excited induction generators", IEE Proc. Pt-C, Vol. 129, No. 6, 1982, pp. 260-265.
[15] S.P. Singh, B. Singh and M.P. Jain, "Performance characteristics and optimal utilization of a cage machine as capacitor excited induction generator", IEEE Trans. on EC, Vol. 5, No. 4, 1990, pp. 679-685.
[16] A.L. Alolah and M.A. Alkanhal, "Optimization-based steady state analysis of three phase self-excited induction generator", IEEE Trans. on EC, Vol. 15, No. 1, 2000, pp. 61-65.
[17] B. Singh, S.S. Murthy and S. Gupta, "Analysis and design of STATCOM-based voltage regulator for self-excited induction generators", IEEE Trans. on EC, Vol. 19, No. 4, 2004, pp. 783-790.
[18] P. Kundur, "Power system stability and control", McGraw-Hill, New York, USA, 1993.

VIII. Appendix

Data of the three-phase generator (generator-a)

$R_1 = 5.033\ \Omega$, $R_2 = 4.668\ \Omega$, $R_c = 4.138\ k\Omega$, $X_1 = 5.605\ \Omega$, $X_2 = 5.605\ \Omega$.
$k_0 = 1019.2$, $k_1 = -24.77$, $k_2 = 0.2615$, $k_3 = -9.56\times10^{-4}$.

Data of the single-phase generator (generator-e)

$R_1 = 3.0\ \Omega$, $R_2 = 3.714\ \Omega$, $X_1 = 3.262\ \Omega$, $X_2 = 3.262\ \Omega$
$k_0 = 331.47$, $k_1 = -5.62$, $k_2 = 8.925\times10^{-2}$, $k_3 = -5.165\times10^{-4}$.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effect of Lightning Return Stroke Current Parameter's on the Components of Lightning Generated Vertical Electric Field over Finitely Conducting Earth

M. Z. I. Sarkar[1], M. A. I. Sarker, and M. M. Ali[2]

Department of Electrical and Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh.
[1]E-mail: mzi_ruet@yahoo.com
[2]Email:mmali57@yahoo.com

Abstract - The scope of this paper is to investigate the effects of lightning induced return stroke current parameters such as peak value, velocity of propagation and cloud height, and the distance of observation point from lightning strike point on the electrostatic, induction and radiation components of lightning generated vertical electric field in time domain. Modified dipole technique has been used in this paper to perform these investigations on and above the finitely conducting earth's surface. The use of modified dipole technique has overcome the limitations of conventionally used Sommerfeld Integral and Wave-tilt function methods for lightning induced field analyses. It has been observed that all the components of vertical electric field are highly affected by the peak value of lightning induced return stroke current and the distance of observation point from lightning strike point. The effects of velocity of propagation of lightning current on the electrostatic and radiation components are greater than that of induction component. The radiation component is not affected by the cloud height.

Index Terms — Modified dipole technique, Lightning generated vertical electric field, Finitely conducting earth, Lightning current parameters.

I. Introduction

The lightning return stroke contains a large current which results in surges in power devices and upward streamers from transmission line and telecommunication towers. The surge voltages induced on the overhead telecommunication and power lines due to the induction of lightning induced electromagnetic fields may range from few hundreds of volts to few kilovolts. The nearby lightning activity is found to affect the operation of electrical/electronic/telecommunication equipment most severely because of the discharge of lightning return strokes. So, the analysis of lightning induced electric fields is an important factor in selecting the insulation level of the conductors and in the optimal design of effective and economic protective devices for adequate protection of electrical, electronic and telecommunication systems. As a consequence, the evaluation of lightning induced electric field has been the subject of theoretical and experimental studies for the last few decades [1-4].

Previously published theoretical works on lightning-induced overvoltages and overcurrents calculation can be classified into two categories. In the first category [2, 3], the earth is considered a medium of infinite conductivity. But the conductivity of earth is a finite quantity and varies from $10^{-3}\Omega^{-1}$/m to $3 \times 10^{-2}\Omega^{-1}$/m depending on the nature of soil [5]. So the fields calculated previously taking earth as the medium of infinite conductivity are not adequate to explain the practical problems. In the second category [6-10], the effect of finite conductivity of earth is considered to evaluate the electric field. The authors of the second category are classified into two groups. The authors of the first group used "Wave-tilt function method" and the second group used "Sommerfeld Integral method" to evaluate the electric fields [7, 9]. However, when the lightning strike point is close to the observation point and the earth's conductivity is lower than $10^{-2}\Omega^{-1}$/m, then both the methods are inapplicable to determine electric fields [7, 9]. Existence of Sommerfeld Integral in the Sommerfeld Integral method makes it complicated and time consuming [9]. On the other hand, Wave-tilt function method is applicable only to explain the effect of earth's conductivity on the radiation component of total electric field [7]. Here, a straightforward and simple approach has been used to investigate the effect of lightning induced return stroke current parameters on the lightning induced electric fields on and above the finitely conducting earth's surface by removing the earth and modeling the reflection from earth by an additional source.

Lightning induced electric field consists of two major components such as horizontal component and vertical component. Again vertical electric field has three components such as electrostatic, induction and radiation. In this paper, Modified dipole technique [10, 11], which is straightforward and simple, and overcomes the limitations of both Sommerfeld Integral and Wave-tilt function

methods, has been used to investigate the effect of lightning induced return stroke current parameters on the components of lightning generated vertical electric field on and above the finitely conducting earth's surface in time domain.

The remainder of the paper is organized as follows. In Section II, the modified dipole model and lightning induced return stroke current have been discussed. For a typical lightning induced return stroke current, estimation of vertical electric field and its components has also been discussed in this section. Section III explains the numerical results of this paper. Finally Section IV contains the conclusion of this work.

II. Theory

An arbitrarily oriented charge S(s), located above a finitely conducting earth's surface must induce a surface charge on the air-earth interface to satisfy the boundary condition. Assuming $\rho_{s,}$ as the induced charge on the air-finitely conducting earth interface, the modified dipole model [10,11], as shown in Fig.1 can be used for estimating electric field at any point P_g in air. In this case, the image source is not just opposite to that of the original source as we see in the case where the second medium is infinitely conducting earth.

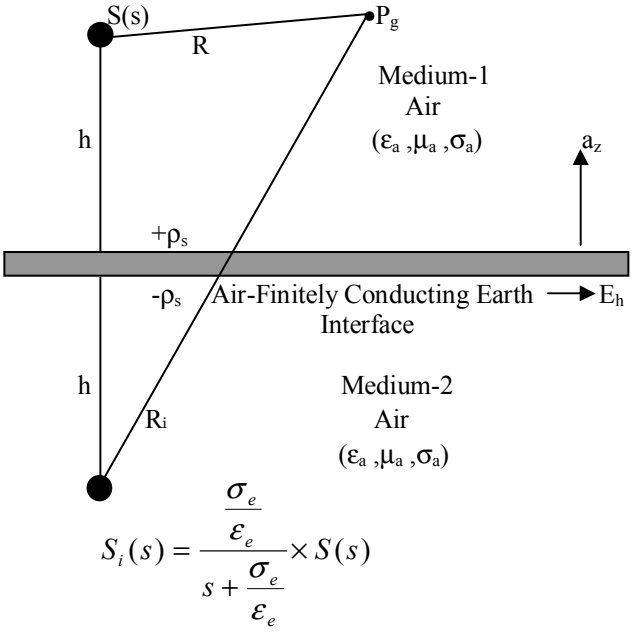


Fig. 1 Modified dipole model for estimating electric field at any point P_g in air due to a charge S(s) on and above the air-finitely conducting earth interface.

To investigate the effect of lightning induced return stroke current parameters on the components of lightning generated vertical electric field, a simple model of lightning induced return stroke current having the waveform shown in Fig.2 is assumed to propagate up and upright channel from the finitely conducting earth at a velocity of v m/μsec.

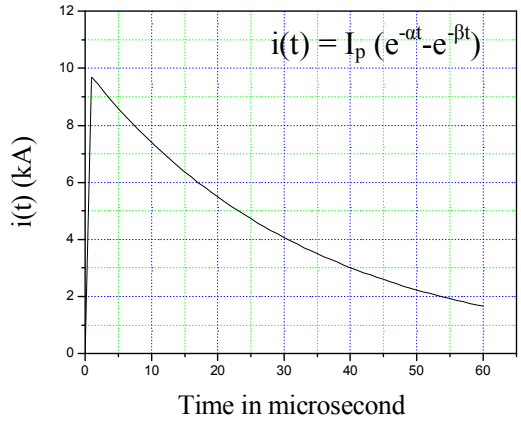


Fig. 2 Lightning discharge model- typical return stroke current.

The generalized expression of image source $I_i(s)$, placed at a depth h below the air-finitely conducting earth interface and capable of inducing a surface charge $\rho_{si} = -\rho_s$ on the interface can be estimated from the following equation [11].

$$I_i(s) = -\frac{\frac{\sigma_e}{\varepsilon_e}}{\left(s+\frac{\sigma_e}{\varepsilon_e}\right)} I(s) \qquad (1)$$

where I(s) and $I_i(s)$ are the laplace transform of source and image current, respectively.

The image current $i_i(t)$, is then obtained from the inverse laplace transform of (1) as —

$$i_i(t) = -I_p\left(e^{-\alpha t} - e^{-\beta t}\right) + \frac{\sigma_e I_p}{\varepsilon_e}\left[\frac{\left(e^{-\alpha t} - e^{-\beta t}\right)}{\left(\frac{\sigma_e}{\varepsilon_e}-\alpha\right)\left(\frac{\sigma_e}{\varepsilon_e}-\beta\right)}\right. + \left.\frac{\sigma_e I_p}{\varepsilon_e}\frac{(\alpha-\beta)e^{-\frac{\sigma_e}{\varepsilon_e}} - e^{-\alpha t} + e^{-\beta t}}{\left(\frac{\sigma_e}{\varepsilon_e}-\alpha\right)\left(\frac{\sigma_e}{\varepsilon_e}-\beta\right)}\right] \qquad (2)$$

Equation (2) obviously satisfy the two extreme cases where σ_e is either 0 or ∞. The expression of image current consists of two terms, the first term is the image of original current and the second term is the effect of earth properties. Figure 3 shows the waveforms of source current and image current.

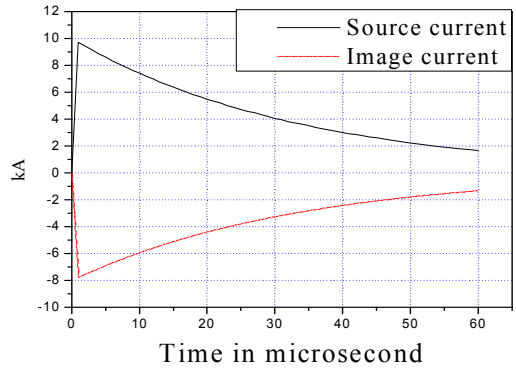


Fig. 3 Typical return stroke current and its image considering the effect of earth properties.

The vertical electric field at any point on and above the finitely conducting earth surface can be calculated by solving the time varying Maxwell's equations. The equation of vertical electric field at point $P_g(r, \phi, z)$ due to a small current element of length dz and carrying a current i(t) at a height z above the ground surface can be expressed as follows:

$$dE_z = \frac{dz}{4\mu\varepsilon_0}\left[\frac{2(z-z')^2 - r^2}{R^5}\int_0^t i(z, t-R/c)dt + \frac{2(z-z')}{cR^4} i(z, t-R/c) - \frac{r^2}{c^2R^3}\times\frac{\delta i(z, t-R/c)}{\delta t}\right] \quad (3)$$

where
E_z = Vertical electric field.

The vertical electric field due to image source $i_i(t)$ can be determined from equation (3) replacing R by R_i, z' by $-z'$ and i(t) by $i_i(t)$. The vertical electric field at point $P_g(r, \phi, z)$ for the small segment of channel is the vector sum of the fields induced by the source dipole and image dipole and the total vertical electric field can then be obtained from the integration of this field over the whole channel.

In equation (3), the first, second and third terms are called electrostatic, induction and radiation fields, respectively.

III. Numerical Results

Numerical results of this research are based on the following parameters unless stated in otherwise: conductivity of earth = 0.001mho/m, relative permittivity of earth = 10, distance of observation point from lightning strike point = 0.2km, cloud height = 4km, peak value, velocity of propagation, front time and tail time of lightning induced return stroke current are 15kA, 100m/μsec, 0.6μsec and 24μsec, respectively.

Figure 4 shows the effect of peak value of lightning induced return stroke current on the electrostatic component of vertical electric field at an altitude of 10 meters from the earth's surface. The field is found to be negative and increases in magnitude with the increase in peak value of lightning induced return stroke current.

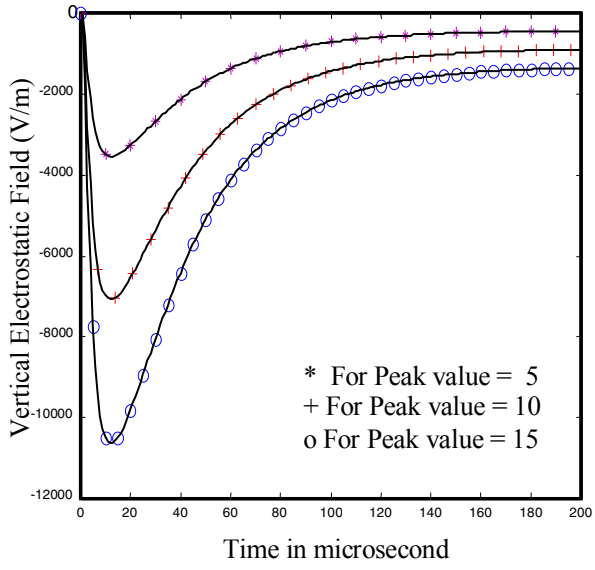


Fig. 4 Effect of peak value of lightning current on the electrostatic component of vertical electric field at an altitude of 10 meters from the earth's surface.

The effect of peak value of lightning current on the induction component of vertical electric field at an altitude of 10 meters from the earth's surface is shown in Fig.5. Initially the field is negative and increases in magnitude with the increase in peak value of lightning current, after a certain period of time t_c (here we call it critical time), it changes its polarity from negative to positive and finally becomes zero. The effect of peak value of lightning current on the field remains the same even when the field becomes positive from negative.

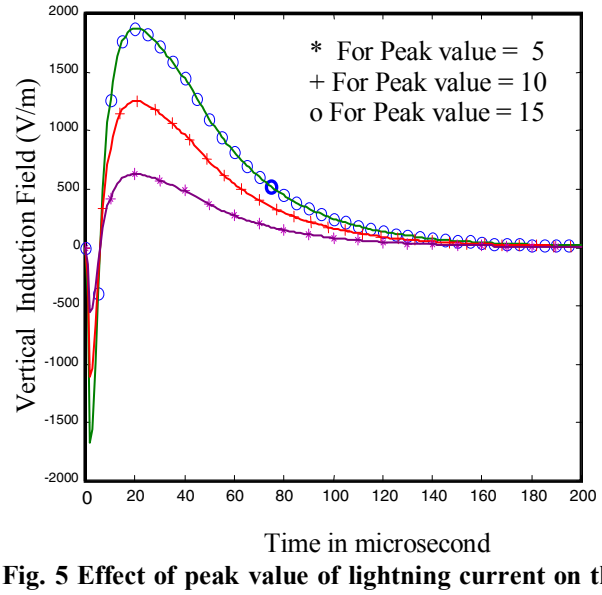


Fig. 5 Effect of peak value of lightning current on the induction component of vertical electric field at an altitude of 10 meters from the earth's surface.

Figure 6 shows the effect of peak value of lightning induced return stroke current on the radiation component of vertical electric field at an altitude of 10 meters from the earth's surface. The field is found to be negative and increases in magnitude with the increase in peak value of lightning induced return stroke current.

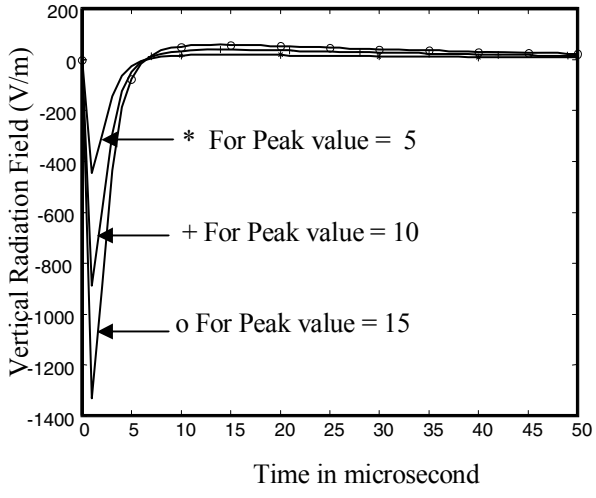


Fig. 6 Effect of peak value of lightning current on the radiation component of vertical electric field at an altitude of 10 meters from the earth's surface.

The effect of velocity of propagation of lightning induced return stroke current on the electrostatic component of vertical electric field at an altitude of 10 meters from the earth's surface is shown in Fig.7. The field is found to be negative and highly affected by the velocity of propagation of lightning induced return stroke current. The intensity of the field decreases in magnitude with the increase in velocity of propagation of lightning current. With the increase in velocity of lightning current, the charge induced at any observation point due to the small segment of lightning channel decreases and, which in turn reduces the induced electric field at that point.

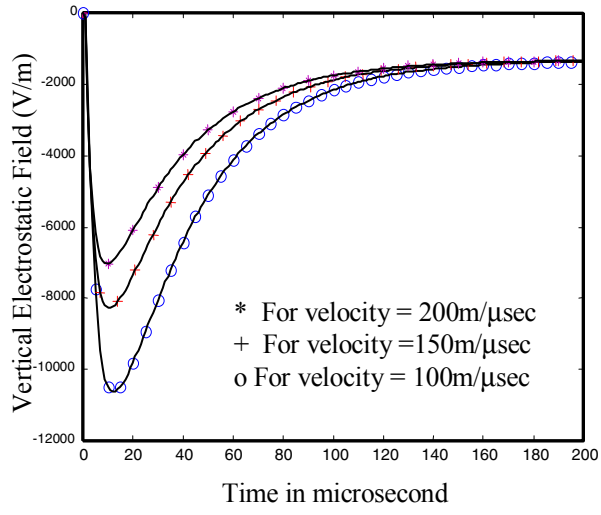


Fig. 7 Effect of velocity of propagation of lightning current on the electrostatic component of vertical electric field at an altitude of 10 meters from the earth's surface.

The effect of velocity of propagation of lightning induced return stroke current on the induction component of vertical electric field at an altitude of 10 meters from the earth's surface is shown in Fig.8. It has been observed that the induction component of vertical electric field is slightly affected by the velocity of lightning current. Initially the field is negative and after a certain period of time, it changes its polarity from negative to positive.

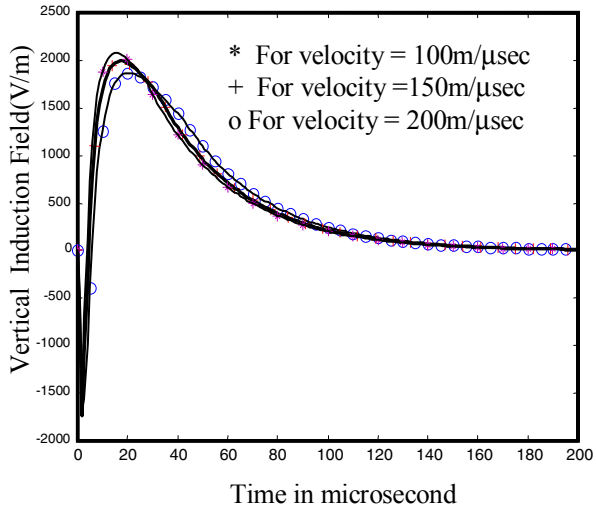


Fig. 8 Effect of velocity of propagation of lightning current on the induction component of vertical electric field at an altitude of 10 meters from the earth's surface.

Figure 9 shows the effect of velocity of propagation of lightning induced return stroke current on the radiation component of vertical electric field at an altitude of 10 meters from the earth's surface. The field is negative and decreases in magnitude with the increase in velocity of propagation of lightning induced return stroke current.

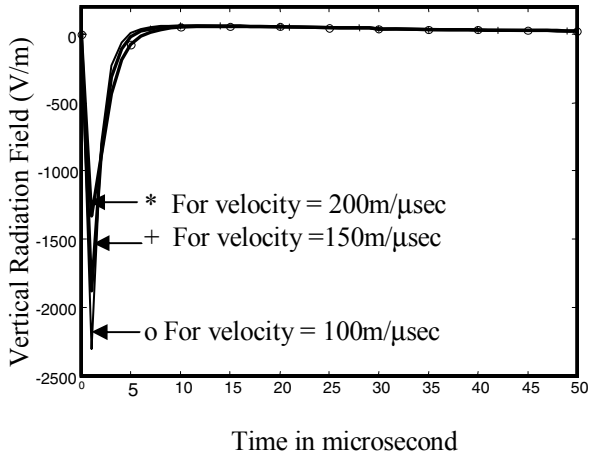


Fig. 9 Effect of velocity of propagation of lightning current on the radiation component of vertical electric field at an altitude of 10 meters from the earth's surface.

Effect of cloud height on the electrostatic component of vertical electric field at an altitude of 10 meters from the

earth's surface is shown in Fig.10. The field is found to be negative and slightly affected by the cloud height. Initially the field is not affected by the cloud height, after a certain period of time; the intensity of the field slightly decreases in magnitude with the increase in cloud height.

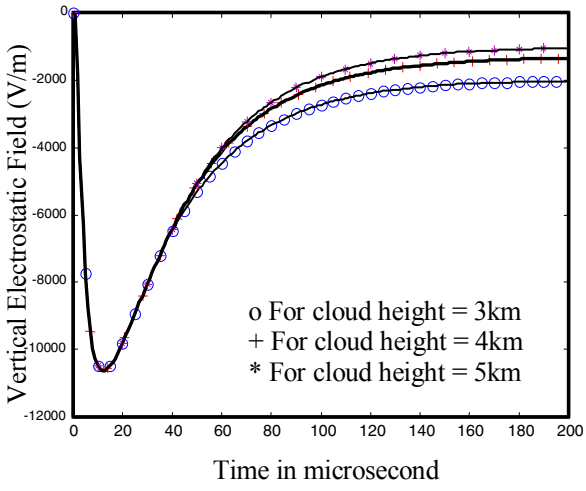


Fig. 10 Effect of cloud height on the electrostatic component of vertical electric field at an altitude of 10 meters from the earth's surface.

Figure 11 shows the effect of cloud height on the induction component of vertical electric field at an altitude of 10 meters from the earth's surface. Initially the field is negative and remains almost constant with the change in cloud height, after a certain period of time it changes its polarity from negative to positive and slightly decreases with the increase in cloud height.

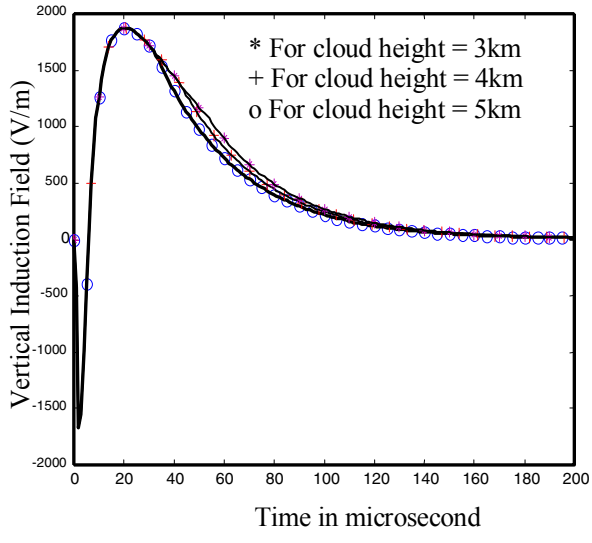


Fig. 11 Effect of cloud height on the induction component of vertical electric field at an altitude of 10 meters from the earth's surface.

The effect of cloud height on the radiation component of vertical electric field at an altitude of 10 meters from the earth's surface is shown in Fig.12. It has been observed that the radiation component of vertical electric field is not affected by the cloud height.

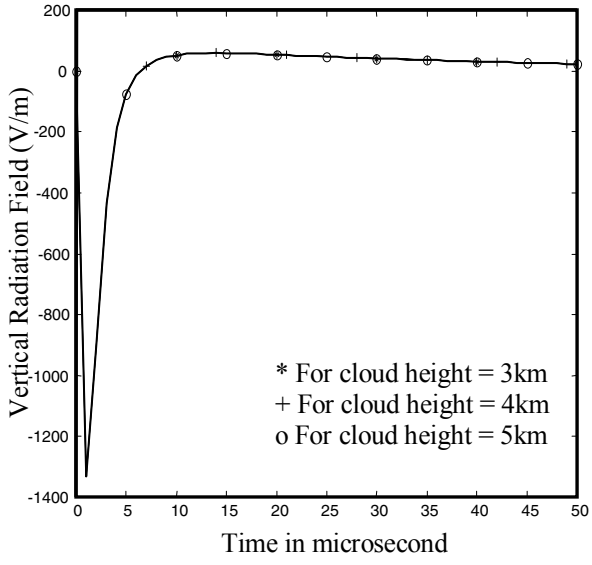


Fig. 12 Effect of cloud height on the radiation component of vertical electric field at an altitude of 10 meters from the earth's surface.

Figure 13 shows the effect of distance of observation point from lightning strike point on the electrostatic component of vertical electric field on the earth's surface. On the earth's surface, the field is negative and decreases in magnitude with the increase in distance of observation point from the lightning strike point.

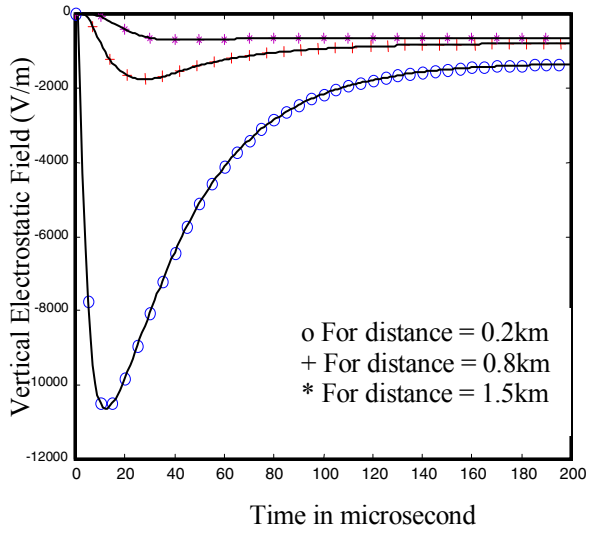


Fig. 13 Effect of distance of observation point from lightning strike point on the electrostatic component of vertical electric field on the earth's surface.

The effect of distance of observation point from lightning strike point on the induction component of vertical electric field on the earth's surface is shown in Fig.14. At close distance, initially the field is negative and decreases in magnitude with the increase in distance of observation point from lightning strike point. After a certain period of time, t_c, it changes its polarity from negative to positive and finally becomes zero.

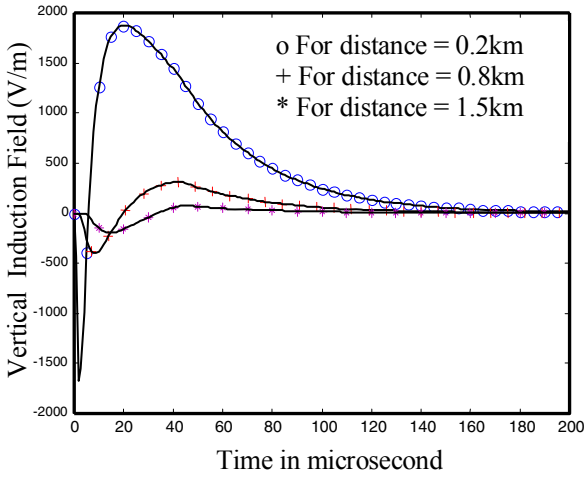


Fig. 14 Effect of distance of observation point from lightning strike point on the induction component of vertical electric field on the earth's surface.

Figure 15 shows the effect of distance of observation point from lightning strike point on the radiation component of vertical electric field on the earth's surface. On the earth's surface, the radiation field is negative and decreases in magnitude with the increase in distance of observation point from lightning strike point.

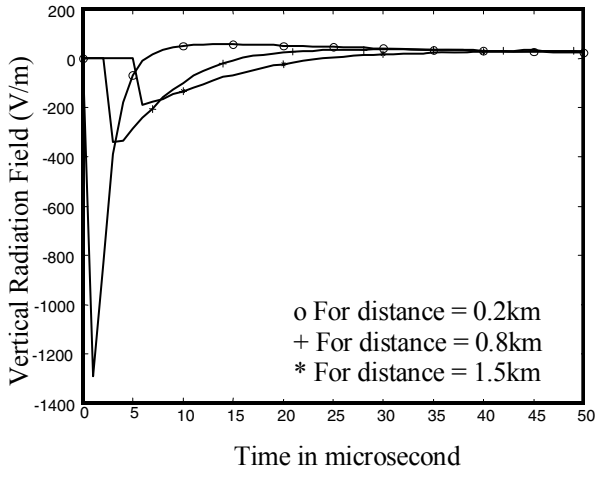


Fig. 15 Effect of distance of observation point from lightning strike point on the radiation component of vertical electric field on the earth's surface.

IV. Conclusion

Modified Dipole Technique has been used in this paper to investigate the effect of lightning induced return stroke current parameters on the components of lightning generated vertical electric field in time domain. It has been observed that the components of vertical electric field are greatly affected by the peak value of lightning induced return stroke current and the distance of observation point from lightning strike point. The electrostatic and radiation fields are negative and increase in magnitude with the increase in peak value of lightning current. The effects of distance of observation point on the electrostatic and radiation components are opposite to the effect of peak value of lightning current. At close distance, initially the induction field is negative and increases in magnitude with the increase in peak value of lightning current. After a certain period of time t_c (here we call it critical time), it changes its polarity from negative to positive and finally becomes zero. The effect of peak value of lightning current on the induction field remains the same even when the field becomes positive from negative. The effects of velocity of propagation of lightning current on the electrostatic and radiation components are greater than that of induction component. The electrostatic and induction components are slightly affected by the cloud height and there is no effect of cloud height on the radiation component.

References

[1] Condettina Buccella and Saverio Cristina "Frequency Analysis of the Induced Effects Due to the Lightning Stroke Radiated Electromagnetic Field", IEEE Transaction on Electromagnetic Compatibility, vol.34, No 3, August 1992.

[2] M. A. Uman, D. k. Mclain, and E.P. Krider, "The Electromagnetic Radiation from a Finite Antenna", Am. J. Phys., vol.43, 1975, pp. 33-38.

[3] M. J. Master, M.A. Uman, Y.T.Lin, and R.B. Standler, "Calculation of Lightning Return Stroke Electric and Magnetic Fields above Ground", J. of Geophys, Res., vol.86.1981, pp.12, 127-12,132.

[4] P. Chowdhuri "Parametric effects on the induced voltages on overhead lines by lightning strokers to nearby ground." Center of Electric Power, Tennessee Technological University Cookeville, TN 38505.

[5] Magdy F. Iskander, *Electromagnetic fields and Waves*, 1992 by Prentice-Hall Inc. A Simon & Schuster Company Englewood Cliffs, New Jersey 07632.

[6] M. J. Master and M. A. Uman, "Lightning induced Voltages on power Lines: Theory", IEEE Trans. On power apparatus and system, vol. PAS-103, 1984, pp. 2502-2518.

[7] A. Zeddam and P. Degauque, " Current and Voltage Induced on the Telecommunication Cables by a Lightning Stroke", Electromagnetics, vol. 7, 1987, pp.541-564.

[8] M. Ishii, K. Michishita, Y. Hongo, and S. Oguma, "Lightning –Induced Voltages on an Overhead Wire Dependent on Ground Conductivity", IEEE Trans. On Power Delivery, vol.9, 1994, pp. 109-118.

[9] K.A. Norton," The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere Part-II",proc.IRe vol.25.1937,p.1203.

[10] M. Z. I. Sarkar and M. M. Ali, "Analysis of Lightning Induced Electric Field and its Components on and above the Lossy Earth's Surface by Modified Dipole Technique", Proc. of IEEE TENCON'02, Vol. 3, pp. 1893-1896, 28-31 Oct. 2002.

[11] M. M. Ali, M. Z. I. Sarkar, M. Y. Hussain and H. Ahmad, "Calculation of Lightning Induced Electric Field Over Lossy Ground with Modified Dipole Technique", ELEKTRIKA, Vol. 6, No. 1, December 2004, pp. 23-27.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Fuzzy Based Micro Calcification segmentation

J. Mohanalin, P. K. Kalra, and Nirmal Kumar

Electrical department, Indian Institute of Technology kanpur
Kanpur,India
E-mail: mohanlee@iitk.ac.in

***Abstract*:** Breast cancer is one of the leading causes of women death in the world. Since the causes are unknown, breast cancer cannot be prevented. Micro calcifications are the earliest signs of breast cancer and their detection is one of the most important research areas now. A novel approach for image segmentation of denser mammography images is introduced, for more accurate detection of microcalcifications clusters. In original mammographic images obtained by X-ray radiography, most of the information is hidden to the human observer. The fundamental idea of the proposed approach is to fuzzify the original image of a mammogram in order to make the difference between the backgrounds and object more. Then we enhance the region of interest and simultaneously suppress the tissues along with background thereby we do segmentation. The advantage of the proposed method is its ability to detect micro calcifications in very dense breast mammograms and also not to loose any vital data .Lot of clinical mammograms was used to test the fidelity of this algorithm. The experiments suggest that the micro calcifications are not only accurately detected but also enhance the visibility of micro calcifications.

I. Introduction

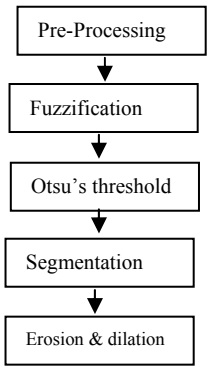
Breast cancer is the second-leading cause of cancer death in women. The reasons behind the cancer are still not clear to prevent it. Early detection is the key for survival of breast cancer patients. Even though we have lot of advanced imaging modalities, Mammography stands tall when it comes to micro calcification detection. Mammogram picks the micro calcification better than its competitors. Although research in computer-aided mammography is there for decades, automated interpretation of micro calcifications (MCs) is still not an easy task. It is mainly due to their fuzzy nature, low contrast and low distinguish ability from their surroundings. MCs are very small. The sizes of MCs are in the range of 0.1–1.0 mm, and the average is about 0.3 mm. They are present with various sizes, shapes, and distributions; so template matching is impossible. They are low contrast so that the intensity difference between suspicious areas and their surrounding tissues can be negligible. MCs are closely connected to surrounding tissues, and simple segmentation algorithms cannot work well. The reasons are the MCs are extremely small and can be easily confused with artifacts. They lead to potential misidentification. One more Issue in handling the mammogram is the denser mammograms. High breast density increases breast cancer risk and the difficulty of reading mammograms. Breast density decreases with age and increases with postmenopausal hormone therapy use. The risk for breast cancer is four to six times higher in women with dense breasts. Breast density may also decrease the sensitivity and, thus, the accuracy of mammography. Radio graphically dense breast tissue may obscure tumors, which increases the difficulty of detecting breast cancer. Studies show that the sensitivity of mammography increases with age especially in postmenopausal women whose breasts are less dense. The interplay of breast density, age, and hormone therapy use on the accuracy of mammography is uncertain. It is a challenging task to extract the Micro calcification alone from a denser mammogram. The main idea of segmenting the MCs from the denser mammograms comes from the fact that MCs are brighter than its tissue. Lot of work has been done using this fact. Some of the works related to MC segmentation are as follows. Chan et al. [2] described an algorithm for micro calcification segmentation. The difference between the enhanced image and suppressed image is calculated. A local thresholding technique is then used to segment the image. Strickleand et.al[6] described a computer-aided mammographic algorithm. Brightness, compactness, and statistics measures are applied in a decision tree to characterize the candidates. Cross-correlation coefficient is then used to measure the presence of micro calcification. Cheng et al. [1] proposed an approach to detecting micro calcifications based on fuzzy logic. The image is fuzzified using a Pi function. But it lacks robustness as it may crumble when MCs lie behind the tissue structures, since they employ complicated tissue structural removal step. Our work is motivated from Cheng's algorithm. A technique called difference-image has been proposed by Dengler et al. [5]. First the difference image is obtained from two weighted Gaussian functions of the original image. The filters are chosen using an expected size for MCs and the distance between MCs. The parameters of Gaussian function is selected such as the First filter works as High pass filter and second one Low pass filter. Next this difference image is thresholded. Kovalerchuk et al. used a fuzzy logic approach to formalize description to an intermediate case

between lobulated and microlobulated masses. Fuzzy membership function was derived with the help of radiologists which might be tricky issue.
Rest of the paper is organized as follows: Section 2) Details about Database utilized. Section 3) complete algorithm Section 4) explains Pre-Processing Section 5) portraits use of Gaussian membership function in Fuzzy based Enhancement 6) Validation report has been included.

II. Mammography Database and tools:

Some of The data related to the breast cancer have been imported from Lawrence Livermore National Laboratory All films were digitized at 35 microns per pixel, and with 12 bits of grayscale per pixel. Image data is stored in the ICS format. It is a database of digitized film screen mammograms with associated ground truth and other information. The Data contains mammograms obtained from Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital and Washington University of St. Louis School of Medicine. We also procured the MIAS data from its website. All the data were digitized to 200 micron pixel edge and clipped/padded so that every image is 1024 pixels x 1024 pixels. We used Matlab 2006a to process the images.

III. Algorithm for segmentation:



IV. Pre-processing:

Generally the mammogram can be classified into 3 important regions, namely 1) Image background 2) Tissue background 3) Object background. Image background is other than the Breast region which is of no use. But the background may have labels which can produce artifacts on the final segmented image. The purpose of pre-processing is to remove noise and artifacts contained within the mammogram caused by labels and markers thereby increasing region homogeneity, reliability and robustness. Most often Mammograms include identification labels or markers. The problem with the presence of label is, it gives imprecise segmentation of the breast region or in other words makes the algorithm to fail by introducing artifacts or false positives. We use an artifact suppression algorithm based on area morphology [10] to remove labels from the background region of mammograms. An example of the algorithm is shown in Figure 1.

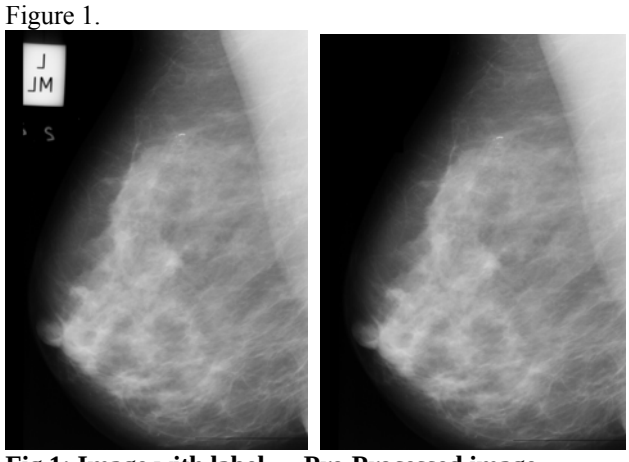


Fig 1: Image with label Pre-Processed image

Once the labels have been removed our next target will be extract the Breast region alone from the Image background. Generally researchers assume that the Intensity level of MCs higher than the average of the entire mammograms. This is a valid assumption as the MCs are small and brighter region. So to extract the Breast region we take the average of the image and remove the region lesser than the average. Entropy based techniques have been used to maximize the information between the tissue background and object background [1].

$$H_b(t) = \sum_{i=k}^{t} P_b \log P_b$$

P_b – Probability distribution of background
k- Average value of intensity
t- Threshold value chosen [varying from k to max value]

$$H_o(t) = \sum_{i=t+1}^{N} P_o \log P_o$$

N-max intensity value of image
P_o-Probability distribution of object
$H_b(t)$ &$H_o(t)$ are the background and object entropies calculated according to [1].

$$t^* = \max\left\{\arg\left(\mathrm{H_b}(\mathrm{t}) + \mathrm{H_o}(\mathrm{t})\right)\right\}$$

This is the optimal threshold value calculate by maximizing the information between background and object.

V. Image fuzzification and enhancement:

Fuzzy set theory has revolutionized the image processing. Mammographic images are extremely fuzzy in nature such as indistinct borders, ill-defined shapes, and different densities. Fuzzy set theory perfectly fits in solving the mammograms as if it has been invented for mammograms. Mammograms are highly fuzzy in nature, having complicated structures like fatty acids and tissues running

everywhere making the life harder for radiologist. Due to this reason, fuzzy logic would be a better choice to deal the fuzziness of mammograms than traditional methods. First, the intensities of an image are transformed to an interval [0, 1] using a Gaussian membership function.

Even though several membership functions are available we used Gaussian Membership function because of its simplicity, robustness. A sigmoid function is not capable of modeling a range such as a class interval. A triangular function will not ensure that all inputs are fuzzified in some class. These are the reasons why Gaussian function was used.

The function is, used to locate the intensities of microcalcifications. The selection of a cross-over point could be viewed as an object-background classification problem, and the thresholding techniques, can be applied. Since we assume that MCs have intensity more than the tissue, it is obvious that the fuzzy region of the function must be the range from the mean intensity to the maximum intensity of the image. In other words, the proposed approach gives maximum membership value to the intensities of microcalcifications higher than the optimal threshold values of breast tissues. It implies that the intensities of microcalcifications are located somewhere between the mean intensity and the maximum intensity of an image. The proposed enhancement technique employs fuzzy set theory to increase the contrast of microcalcifications and statistics have been not taken into account anywhere in our work which increases the robustness of the algorithm.

The Gaussian function can be written as

$$\mu_{m,n} = \exp\left(-\frac{\left|\left(x - MaxN\right)\right|^2}{f_h^{\,2}} \right)$$

x- Pixel value at (m,n) position

MaxN-Maximum Intensity value

f^2_h-is the width of the Gaussian membership function. It acts as scaling function here. f^2_h can be calculated by the following method. The bandwidth of the Gaussian function can be found out by the following equation.

$$b = \max\left\{\left(t^* - k\right), \left(MaxN - t^*\right)\right\}$$

The region of image below to the optimal threshold value will be given least membership function and the region which is greater than the optimal threshold value will be given more belongingness. This can be done by properly setting the width (scale) of the Gaussian membership function. The width (fh) of the tissue background region will be set as difference between t* and bandwidth **b** of that image, While width of the tissue region will be set as the half the difference found in previous step. This is a pixel by pixel process and not traditional windowing technique. By this way we suppress more the region which doesn't belong to the ROI and enhance the remaining. The enhanced image can be obtained by following formula.

Enhanced image= μ *N

μ is the fuzzied image and N is the maximum value of our image.

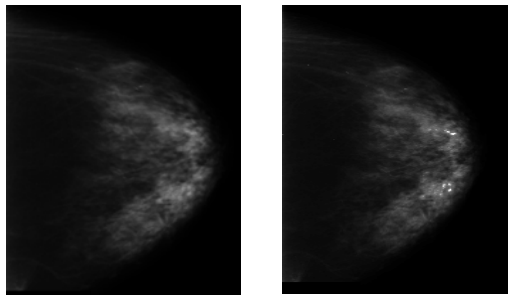


Fig 2: Original image and over laid enhanced image

For enhancing the image we haven't taken the local statistics of the image which makes our algorithm special. Also this technique is highly reliable even if the mammogram is denser. Some of the MIAS database is much denser such that many algorithms will fail. But our algorithm has segmented without much trouble. Now the image contains ROI along with some tissues left. This can be simply viewed as two class problem, as we have only MCs and some tissues. This can be solved by using simple Otsu's technique [9]. Entire image can be split into 2 classes C0 and C1. The threshold is calculated by taking the ratio between "between-class variance", and the total variance. Let

σ^2_B, σ^2_T be the between-class variance, and the total variance, respectively.

$$\eta = \sigma^2_B / \sigma^2_T$$

The threshold between the two classes can be found out by t= Arg min η. This step isolates eventually all the tissues.

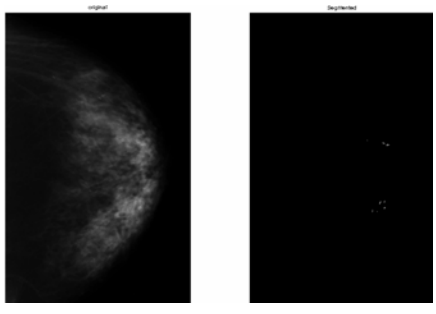


Fig 3: original image and segmented image

Binary morphological operators are employed to reconstruct the shapes of the microcalcifications and to remove isolated pixels. Generally MCs appear as clusters. Some of the isolated pixels might be false positives. Erosion is the morphological operator used to remove the isolated spots. Traditionally 3 different levels of [0 1 0; 1 1 0; 0 0 0], [0 1 0; 1 1 0; 0 0 0]; [0 0 0; 0 1 1; 0 1 0] have been used parallel to suppress the isolated pixels. But erosion operation causes loss of some vital information. This will be regained by using morphological

operation called dilation. This can be done by using [0 0 0;0 0 1;0 1 1],[0 0 0;0 1 0;1 1 1] [0 1 0;1 1 0;0 0 0] parallel to fill the holes created by the erosion operation.

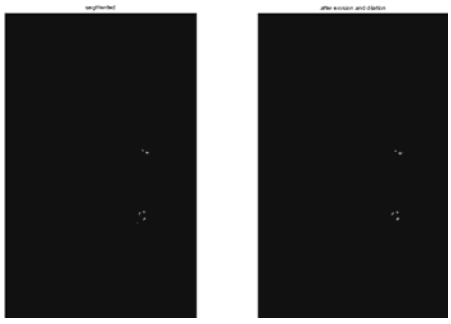


Fig 4: Segmented image and Morphological processed image

The figure displayed has been artificially increased the brightness to get a clear visualization .Our Algorithm is robust enough to deal with the denser image to precisely segment the MCs. Following diagram shows a dense mammogram and its segmented results.

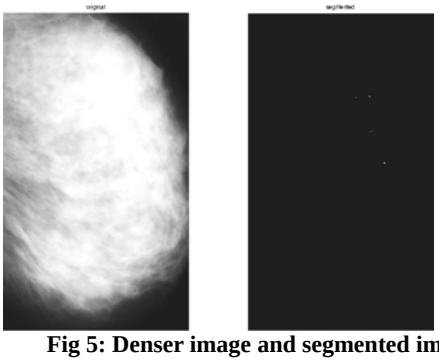


Fig 5: Denser image and segmented image

VI. Validation of results

Free response operating characteristic (FROC) is the method to evaluate the credibility of the mammogram segmentation algorithms. We conducted FROC analysis for totally 45 images carefully selected from group of denser images. Out of those 33 were from UCSF and 12 were MIAS. All images selected were denser images, however the image displayed is less dense and it is selected for display purpose only for better clarity. The FROC test is conducted for 3 different threshold levels. We got 85.8% sensitivity and 80.4% specificity when MIAS database included. While considering the UCSF data alone the sensitivity and specificity increased to 90.04% and 87.5%. This is due to the fact that, MIAS database carries more challenging data because it caries lot of unwanted tissues and flesh part included in it which has the intensity level more than the MCs. So basically the assumption which we took may get disturbed. To avoid this problem we included that pre-processing step.

VII. Conclusion:

We have proposed a new algorithm for effectively segment the Microcalcifications. We have shown the obtained results so far in this experiment. The sensitivity and specificity of chan's algorithm were 0.85 and 1.5 respectively. The results suggest that our algorithm performs better than some existing algorithms like Chan's algorithm. The Algorithm works perfect for denser mammograms.

References

[1]. Heng-Da Cheng, Yui Man Lui, and Rita I. Freimanis, " A Novel Approach to Microcalcification Detection Using Fuzzy Logic technique", IEEE transactions on medical imaging, vol. 17, no. 3, june 1998

[2]. H. P. Chan et al, "Image feature analysis and computer-aided diagnosis in digital radiography—I. Automated detected of microcalcifications in mammography," Med. Phys., vol. 14, no. 4, pp. 538–548, 1987.

[3]. J. Kittler and J. Illingworth, "Minimum error thresholding," Pattern Recogn., vol. 19, no. 1, pp. 41–47, 1986.

[4]. X. Li, Z. Zhao, and H. D. Cheng, "Fuzzy entropy threshold approach to breast cancer detection," Inform. Sci., Applicat.: An Int. J., vol. 4, no. 1, pp. 49–56, 1995.

[5]. J. Dengler, S. Behrens, and J. F. Desaga, "Segmentation of microcalcifications in mammograms," IEEE Trans. Med. Imag., vol. 12, pp. 634–642, Dec. 1993.

[6]. R. N. Strickland and H. I. Hahn, "Wavelet Transforms for Detecting Microcalcifications in Mammograms", IEEE Trans. on Medical Imaging, vol. 15, no. 2, pp. 218-229, April 1996.

[7]. H. Chan, K. et.al "Image feature analysis and computer-aided diagnosis in digital radiography, Automated detection of microcalcifications in mammography", Medical Physics, pp. 538-547, vol. 14, no. 4, Jul/Aug 1987.

[8]. M. N. Gaurcan, et.al "Detection of microcalcifications in mammograms using nonlinear subband decomposition and outlier labeling", in Proceedings of SPIE Visual Communications and Image Processing Conference, pp. 909-918, 8-14 February, 1997, San Jose, CA.

[9]. N. Otsu, "A threshold selection method from grey-level histograms", IEEE Trans. Syst., Man, Cybern., vol. SMC-8, pp. 62–66, 1978.

[10] M. Wirth, D. Nikitenko, J. Lyon "Segmentation of the Breast Region in Mammograms using a Rule-Based Fuzzy Reasoning Algorithm", ICGST-GVIP Journal, Volume 5, Issue2, Jan. 2005.

[11] D. Nesbitt, "Detection of Microcalcifications in Digitized Mammogram Film Images using Wavelet Enhancement and Local Adaptive False Positive Suppression", 1995 IEEE

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Compression of ECG Signal Based on Its Deviation From a Reference Signal Using Discrete Cosine Transform

Mohammad Saiful Alam and Newaz Muhammad Syfur Rahim

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh.
E-mail: saiful209@yahoo.com, newazrahim@eee.buet.ac.bd

Abstract - In this paper a discrete cosine transform (DCT) based ECG signal compression method is proposed. DCT is applied on the residual beats of ECG signal obtained after subtracting the period normalized and dc removed beats from a reference beat. The quantization and the threshold vectors are optimized to minimize the entropy of the quantized coefficients for a target distortion. The quantized coefficients are lossless encoded for further compression. Simulation result shows that the proposed method ensures higher compression ratio than the conventional methods published in the literature.

I. Introduction

A typical ECG monitoring device generates large volume of digital data from 7.5 KB (sampling frequency 125 Hz, quantizer resolution 8 bit) to 45 KB (500 Hz, 12 bit) per minute per sensor. Also there may be up to 12 different sensors placed over different positions of the body capturing different signals. As a result it becomes essential to compress this massive ECG data maintaining the signal quality since it is a major requirement of storage and transmission systems having limited memory and bandwidth. A great effort has been given for many years for finding out better methods of ECG compression. We can classify the methods into three major groups [1]: direct data compression (DDC), transformation compression (TC), and parameter extraction compression (PEC). DDC methods are based on their detection of redundancies on direct analysis of the actual signal samples to provide the compression. The DDC schemes include AZTEC, TP, CORTES, Fan, SAPA, SAIES, SLOPE algorithm, CORNER algorithm. In TC methods, the original samples are subjected to a (linear) transformation and the compression is performed in the new domain. Fourier Transform, Walsh Transform, Cosine Transform, KLT and Wavelet Transform are examples of transformation compression. While, PEC is an irreversible process with which a particular characteristic or parameter of the signal is extracted. The extracted parameter (e.g. measurement of the probability distribution) is subsequently utilized for the classification based on a priori knowledge of the signal features. ANN, Long-Term Prediction, Vector Quantization etc are included in PEC.

The distortion of the compressed signal is generally measured by Percent Root-Mean- Square Difference (PRD) which is defined as

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}\left(x(n)-\tilde{x}(n)\right)^2}{\sum_{n=1}^{N}x^2(n)}}\times 100 \qquad (1)$$

where $x(n)$ and $\tilde{x}(n)$ are the samples of original and reconstructed signals respectively. Another measure PRD2 is sometimes used, as PRD becomes low when the signal has some dc offset part and defined as

$$PRD2 = \sqrt{\frac{\sum_{n=1}^{N}\left(x(n)-\tilde{x}(n)\right)^2}{\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^2}}\times 100 \qquad (2)$$

where $\bar{x}$ is the average or dc value of the original signal. For the evaluation of amount of compression, we often use the terms compression ratio (CR) and compressed data rate (CDR). CR is the ratio of total number of bits of original signal to that of compressed signal where CDR is the ratio of number of bits of compressed signal to the time duration (in sec) of the signal. The CR of the previous methods lies approximately in the range of 2 to 50, but when the methods try to get good CR, the PRD becomes poor.

In the field of data compression Discrete Cosine Transform (DCT) is the most frequently used transform based method for its better decorrelation and energy compaction properties. For example DCT is being used for ECG, image, video and audio compression.

In this paper, we presented a DCT based compression method of ECG signal. The main reason behind our choice of DCT is, being a transformation method, it is insensitive to noise exists in the signal. The signal energy concentrated in a few transform coefficients helped us to improve the compression.

II. Compression Algorithm

As Discrete Cosine Transform (DCT) and Inverse Discrete Cosine Transform (IDCT) are used in this method their definitions are given at first. The most common DCT definition of a 1-D sequence $b[n]$ of length N is

$$B[m] = \left(\frac{2}{N}\right)^{1/2} c_m \sum_{n=0}^{N-1} b[n] \cos\left[\frac{(2n+1)m\pi}{2N}\right], \quad (3)$$

m =0, 1,..., N-1

Similarly, the inverse transformation IDCT is defined as

$$b[n] = \left(\frac{2}{N}\right)^{1/2} \sum_{m=0}^{N-1} c_m B[m] \cos\left[\frac{(2n+1)m\pi}{2N}\right], \quad (4)$$

n=0,1,…,N-1

In both equations (3) and (4) c_m is defined as

$$c_m = \begin{cases} (1/2)^{1/2} & for \quad m = 0 \\ 1 & for \quad m \neq 0 \end{cases}$$

The coefficient $B[0]$, which is directly related to the average value of the time-domain block, is often called the *DC coefficient*, and the remaining coefficients are called *AC coefficients*.

The proposed compression algorithm can be divided into pre-processing stage and encoding stage. The pre-processing stage is also divided into the first and second pre-processing stages. Figures 1 to 3 shows their block diagrams.

In our method of ECG compression, we followed the idea of defining optimum quantization and threshold vector for the quantization process of DCT coefficients which was used by Batista, Melcher and Carvalho in their "optimized quantization of DCT coefficients" method [2]. This kind of optimization is generally employed in JPEG image compression.

A. The First Pre-processing Stage

Figure 1 shows the first pre-processing stage where the reference signal or beat is generated. At first, the ECG signal is partitioned into its periods or beats. A beat is defined here as the signal between two R waves. As ECG signal is quasiperiodic, the lengths of the partitioned blocks (beats) are not equal. The dc components of the beats are also different. These partitioned beats are then period normalized. The dc variations are also removed. For this purpose the partitioned blocks are DCT transformed. The dc coefficients are removed by assigning zeros. Then the first 270 DCT coefficients are taken from the DCT coefficient blocks. If any DCT coefficient block contains less than 270 coefficients, then the blank coefficients are filled with zeros. Then the blocks of 270 DCT coefficients are IDCT transformed. These operations make the beats of equal length (270 samples, which is the standard period) with zero dc component. After averaging and rounding these normalized periods, we get the reference signal $\boldsymbol{R}$.

The duration of the segmented beats are stored in the vector $\boldsymbol{p}$. The removed dc coefficients are divided by 10

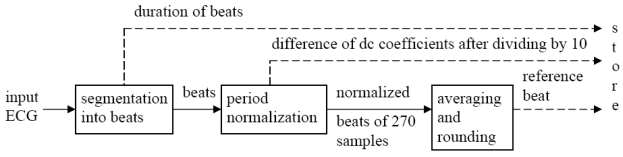


Fig. 1 The first pre-processing stage

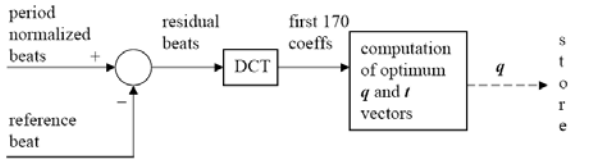


Fig. 2 The second pre-processing stage

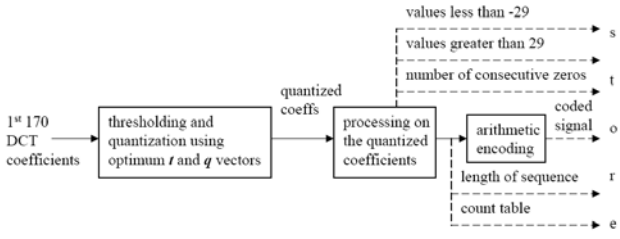


Fig. 3 The encoding stage

and rounded to the nearest integer and their differences are stored in the vector $\boldsymbol{m}$.

Figures 4 to 7 are presented here to explain how the residual beats are created from segmented beats. Here the segmented beats are taken from the first two minutes of record 100/MLII from MIT-BIH Arrhythmia Database.

B. The Second Pre-processing Stage

Figure 2 shows the second pre-processing stage where the optimum quantization and threshold vectors are generated. The reference beat is subtracted from the period normalized beats to get residual beats of equal length (270 samples).

DCT is then applied to these residual beats. The first 170 DCT coefficients are stored for the encoding stage.

The first 170 DCT coefficients are considered for quantization and thresholding operation and rest of the coefficients are omitted. Let N_b be the total number of beats, B_i be the first 170 DCT coefficients corresponding to the of i^{th} residual beat.

$\boldsymbol{B}$ is to be thresholded and quantized using the optimum threshold vector $\boldsymbol{t}$ and quantization vector $\boldsymbol{q}$. Thresholding and quantization are necessary for entropy minimization and reduction of bit numbers allocated for the quantized values. Thresholding generates a lot of zeros and quantization limits the values in a shorter range. To get the optimum $\boldsymbol{t}$ and $\boldsymbol{q}$ vectors, the following tasks are done. For all the n^{th} coefficients, $q[n]$ = 1, 2, 3, …, 32 and $t[n]$ = $q[n]/2$, $q[n]/2+0.25$, $q[n]/2+0.5$, …, 32 are tested for quantization and thresholding operation, where n = 1, 2, 3, … , 170.

Let $\hat{\boldsymbol{B}}$ be the quantized coefficients after thresholding and quantization.

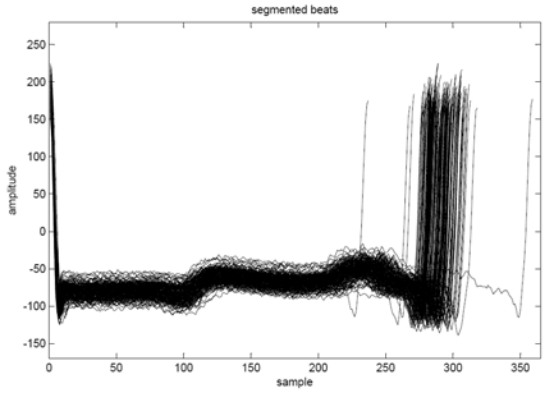


Fig. 4 The partitioned unnormalized beats of ECG signal

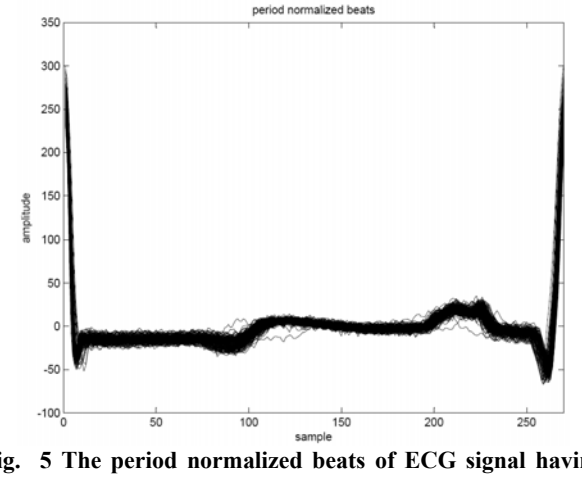


Fig. 5 The period normalized beats of ECG signal having zero dc component

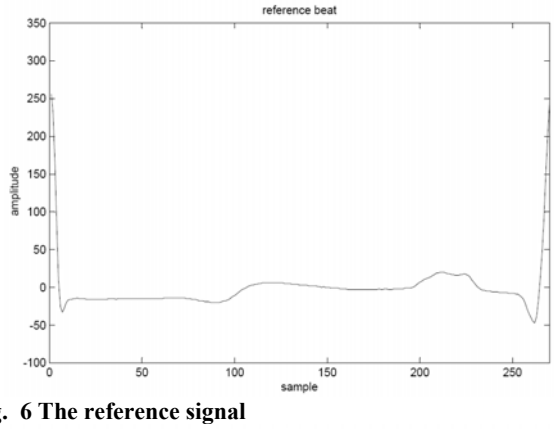


Fig. 6 The reference signal

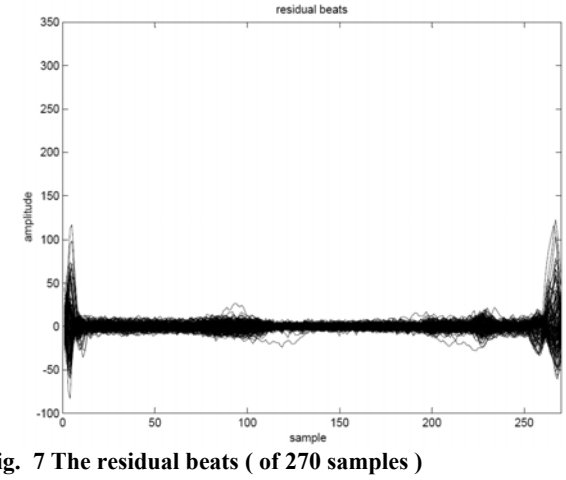


Fig. 7 The residual beats (of 270 samples)

$$\hat{B}_i[n] = \begin{cases} 0, & \text{if } |B_i[n]| < t[n] \\ \text{round}(B_i[n]/q[n]), & \text{otherwise} \end{cases} \quad (5)$$

n =1, 2, 3, …, 170 and i=1, 2, 3,…, N_b

The mean squared error introduced by the quantization of coefficient number n is given by,

$$D_n(q[n],t[n]) = \frac{1}{N_b}\sum_{i=1}^{N_b}(B_i[n]-q[n]\hat{B}_i[n])^2 \quad (6)$$

n = 1, 2, 3, …… , 170.

Considering that, due to the quantization and thresholding process, a value v arises $N_n(v)$ times in the coefficient number n of the N_b quantized blocks. Then the entropy $H_n(q[n],t[n])$ of the coefficient number n, measured over all quantized DCT blocks is given by,

$$H_n(q[n],t[n]) = -\sum_v p_n(v)\log_2 p_n(v) \quad (7)$$

where $p_n(v) = \frac{N_n(v)}{N_b}$. Then for a target distortion, we get the desired optimum $q[n]$ and $t[n]$ by minimization of Lagrangian J_n,

$$J_n = H_n(q[n],t[n]) + \lambda_n D_n(q[n],t[n]), \quad (8)$$

n = 1, 2, 3, ….., 170

where λ is the Lagrange multiplier. The value of λ is given by the negative slope of the line joining the two points on the convex hull that supports all the points of D-H at target distortion. Figure 8 shows how λ is found from the D-H curve. Figure 9 shows the ***q*** and ***t*** vectors for PRD = 2.5%.

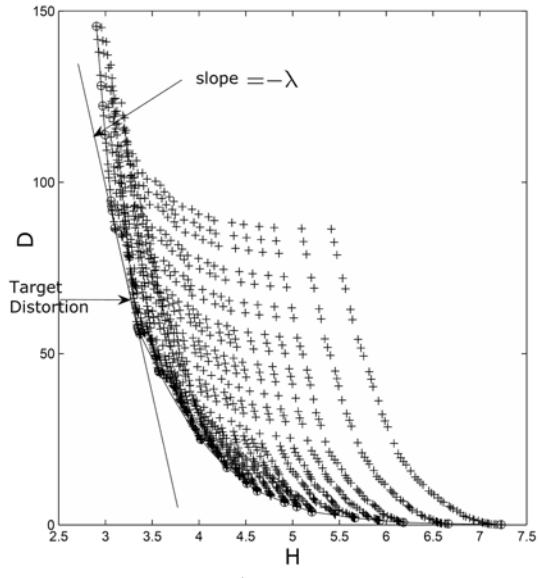


Fig. 8 Finding the value of λ from the points of D-H

C. Encoding Stage

After defining the optimum quantization and threshold vectors ***q*** and ***t***, the quantization of the first 170 DCT coefficients ***B*** is performed using ***q*** and ***t***, and the quantized coefficients $\hat{\boldsymbol{B}}$ is produced. Some simple processing is done on the quantized coefficient blocks and the vector ***s*** is generated containing values -31 to 31. ***s*** is then lossless compressed by arithmetic encoding and ***comp*** is produced which is a stream of binary numbers.

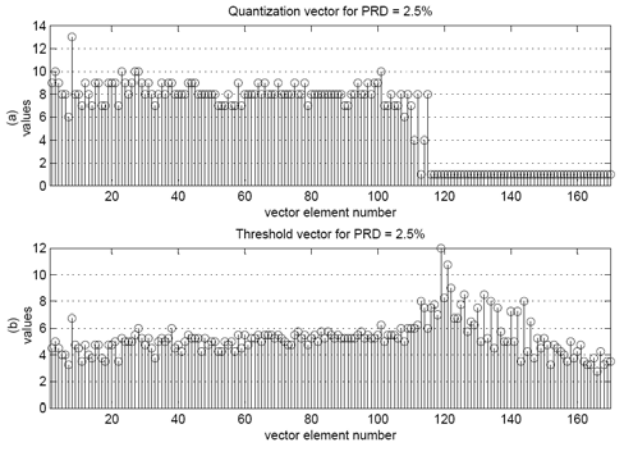


Fig. 9 (a) Quantization vector and (b) threshold vector for PRD = 2.5%, record 100/MLII

Two overheads are also generated before arithmetic coding. They are *count table* of the values and *length of sequence* (i.e, length of the ***s*** vector).

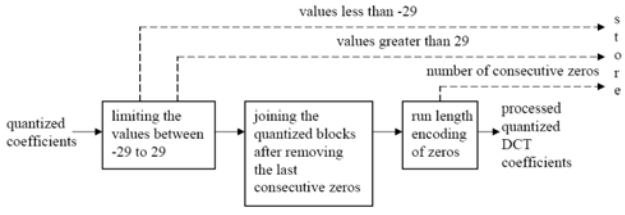


Fig. 10 The processing on the quantized coefficients

The processing on the quantized DCT coefficients are shown in Figure 10 and explained here. Maximum values in $\hat{\boldsymbol{B}}$ are between -29 to 29. For a value v<-29, v is replaced by -30 in $\hat{\boldsymbol{B}}$, and the value –(v+30) is stored in the vector ***u1***. Similarly a value v>29 is replaced by 30 in $\hat{\boldsymbol{B}}$, and (v-30) is stored in vector ***u2***.

After every last nonzero quantized DCT coefficient of $\hat{B}_i$, a value -31 takes place which indicates the end of the i^{th} block. Omitting the last consecutive zeros (zeros after -31), all the blocks are joined together in a single vector ***s***.

There are many zeros in ***s*** for threshold and quantization process. These zeros are reduced by run length encoding. If there are more than 9 consecutive zeros, then a value 31 takes place for them in the ***s*** vector, and the value (number of consecutive zeros -10) is stored in a vector ***r***.

All the values of ***u1***, ***u2*** and ***r*** vectors are positive numbers. The vector ***u*** is rewritten in the vector ***u1′***, where 5 bits are allocated for each vector element. So the values of ***u1*** are broken down into 5 bits as follows. A value $v \geq 31$ is represented as a sequence of c numbers of 31, where c is the greatest integer such that $31 \times c \leq v$, followed by the value equal to $v - 31 \times c$. For example, the value 66 in ***u1*** would be represented as (31, 31, 4) in ***u1′***. Similarly the vectors ***u2*** and ***r*** are rewritten in the vectors ***u2′*** and ***r′*** so that each element can be expressed by 5 bits.

III. Bit Allocation and Compression Measure

The compressed signal is composed of the vectors ***comp***, ***u1′***, ***u2′***, ***r′***, ***m***, ***p***, reference signal ***R***, quantization vector ***q***, dc coefficient corresponding to the first beat, count table of the symbols for arithmetic coding, length of the sequence ***s*** which is to be coded by arithmetic encoding.

As the elements of ***q*** are limited to integer values between 1 and 32, allocating 5 bits for each element, ***q*** takes 170×5=850 bits in the compressed signal. Using 16 bits to represent the number of occurrences of the symbols in ***s***, the count table requires (31+1+31)×16=1008 bits. Allocating 10 bits for each element, ***R*** vector takes 270×10=2700 bits. 10 and 30 bits are allocated for first dc coefficient and length of sequence respectively. Thus 4598 bits are fixed irrespective to the duration of the ECG signal. If the length of the ECG signal is large, then these fixed bits can not influence the compression ratio. The compression ratio actually depends on the vectors ***comp***, ***u1′***, ***u2′***, ***r′***, ***m*** and ***p*** only.

The vector ***m*** contains the information about the dc coefficients. 8 bits are allocated for each element of ***m*** where the values may range from -128 to 127.

The vector ***p*** contains the beat durations. 10 bits are allocated for each element of ***p*** so that the values may range from 0 to 1023.

5 bits are allocated for each element of the vectors ***u1′***, ***u2′*** and ***r′*** to range the values from 0 to 31. Finally the Compression Ratio (CR) and Compressed Data Rate (CDR) are calculated as follows:

$$\mathrm{CR} = \frac{b_{org\ ECG}}{b_{comp} + b_m + b_p + b_{r'} + b_{u1'} + b_{u2'}} \qquad (9)$$

and the compressed data rate can be measured as

$$\mathrm{CDR} = \frac{b_{comp} + b_m + b_p + b_{r'} + b_{u1'} + b_{u2'}}{t_{org\ ECG}} \qquad (10)$$

where,

$b_{org\ ECG}$ = length (ECG signal) × quantizer resolution
b_{comp} = length (***comp***) × 1
b_m = length (***m***) × 8
b_p = length (***p***) × 10
$b_{r'}$ = length (***r′***) × 5
$b_{u1'}$ = length (***u1′***) × 5
$b_{u2'}$ = length (***u2′***) × 5
$t_{org\ ECG}$ = length (ECG signal) / sampling frquency

IV. Decompression Algorithm

Decompression is performed in the reverse order.
Using the count table and length of sequence, the arithmetic coded signal ***comp*** is arithmetic decoded and the vector ***s*** is reproduced.

From the vectors ***u1′***, ***u2′*** and ***r′***, ***u1***, ***u2*** and ***r*** vectors are reproduced respectively.

In the place of 31, consecutive zeros are replaced. The numbers of consecutive zeros are found from the vector ***r*** where the numbers are added with 10.

The values greater than 29 and less than -29 in ***s*** are represented by 30 and -30 respectively, and are recovered with the help of the vectors ***u1*** and ***u2***.

The last consecutive zeros of each quantized DCT coefficient block were replaced by -31. So these values (-31) are replaced by zeros such that the length of each quantized DCT block becomes 170. Then these blocks are separated and $\hat{\boldsymbol{B}}$ is reproduced.

The quantized coefficients of $\hat{\boldsymbol{B}}$ are multiplied by the quantization vector ***q*** and we get ***C*** .

$$C_i[n] = \hat{B}_i[n] \times q[n] \qquad (11)$$

$$n = 1, 2, \ldots, 170 \quad i = 1, 2, \ldots, N_b$$

The coefficients of ***C*** are increased to 270 where the first 170 coefficients are unchanged and the remaining coefficients are zeros.

C is then IDCT transformed and the residual beats are created. These residual beats are added with the reference signal ***R*** and we get the period normalized beats of 270 samples.

Again these normalized beats are DCT transformed. The dc coefficients are replaced that are reproduced from the vector ***m***. The numbers of coefficients are changed according to the vector ***p***. Then these DCT coefficient blocks are IDCT transformed and finally the decoded / reconstructed signal is generated.

V. Results of Simulation

To evaluate the performance of the proposed method, we used the beats of the first two minutes of MLII channel of 40 records of the MIT-BIH Arrhythmia Database. These ECG signals are sampled at 360 Hz and quantizar resolution is 11bits/sample. The records are 100, 101, 103, 105, 106, 107, 111, 112, 113, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 200, 202, 205, 208, 209, 210, 212, 213, 214, 215, 217, 219, 220, 221, 222, 223, 230, 231, 232, 233 and 234. We determined the optimum ***q*** and ***t*** vectors for each test signal and for target PRD levels of 1.5%, 2.0%, 2.5% and 3.0% and PRD2 levels of 2.0%, 3.5%, 5.0% and 6.5% and then compressed the signals. Figure 11 summarizes the results.

A single cycle of record 100/MLII and the reconstructed signal for PRD equal to 2.5% and CR equal to 13.1 is presented in Figure 12. It shows that the important characteristics of the original signal have been preserved in very high precision in the reconstructed signal.

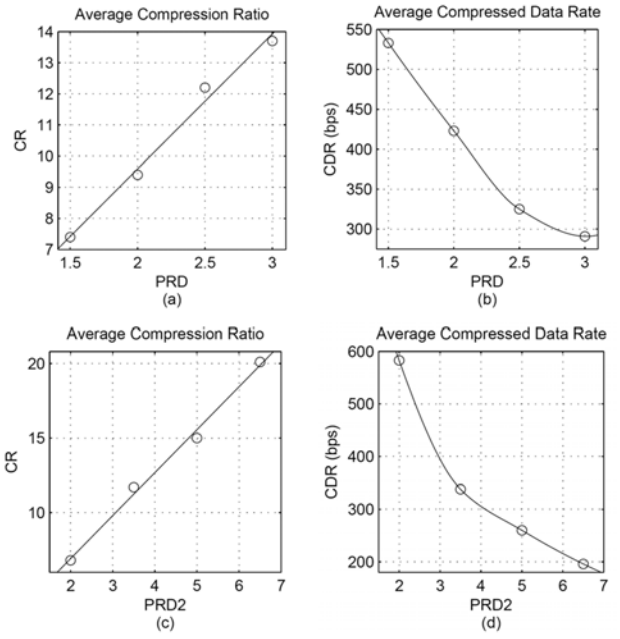


Fig. 11 (a) Average Compression Ratio and (b) Average Compressed Data Rate at PRD 1.5%, 2.0%, 2.5% and 3.0%, (c) Average Compression Ratio and (d) Average Compressed Data Rate at PRD2 2.0%, 3.5%, 5.0% and 6.5%.

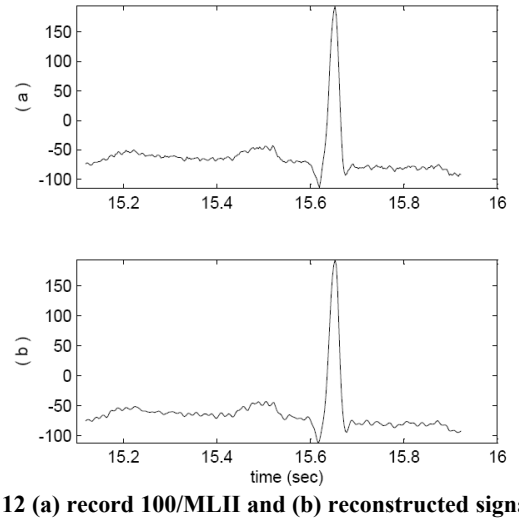


Fig. 12 (a) record 100/MLII and (b) reconstructed signal for PRD =2.5%, CR =13.1

Figures 13 and 14 shows 7 sec segment of ECG signal taken from record 124/MLII and 222/MLII respectively and their reconstructed signal at PRD2 = 6.5%. In both cases the compression and signal quality are satisfactory. Figure 14 indicates very good reconstruction in presence of noise in the original signal.

Table 1 presents the comparison of the performance of the proposed method with some well known ECG compressors. Fixing the CRs, the PRDs of those methods and the proposed method are compared in this table. Numbers marked by * are PRD2 values.

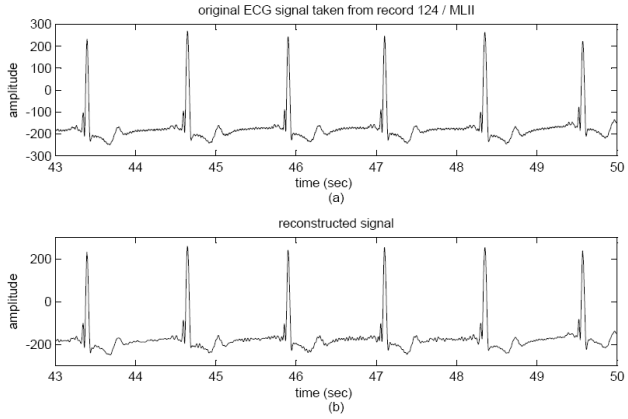


Fig. 13 (a) 7 sec segment of record 124/MLII and (b) reconstructed signal for PRD2 = 6.5%, PRD = 2.7%, CR = 40.4, CDR = 98 bps

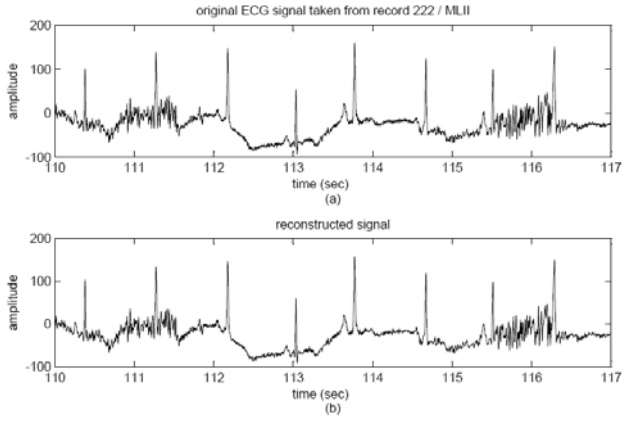


Fig. 14 (a) 7 sec segment of record 222/MLII and (b) reconstructed signal for PRD2 = 6.5%, PRD = 5.1%, CR = 11.1, CDR = 357 bps

VI. Conclusion

The average CR and CDR achieved by running our compressor on 40 records is very much convincing. It can be seen from the Table 1 that our proposed method can compress ECG data much better than the mentioned methods. The quantization method adopted here is very much realistic which is mainly used in JPEG image compression. Both the quantization method and consideration of large number of DCT coefficients (first 170 from 270) helped us to obtain better CR and CDR at relatively lower distortion. Although the method performs better, the main disadvantage is, it cannot be implemented for online ECG data compression. This is an offline method. As a result the compression of a long duration of signal needs much time. Also if any beat is undetected, a large error occurs in that particular beat in the reconstructed signal. However, we can further extend this type of compression method on 2D DCT to get better performance.

Table 1: Comparison of the proposed method with other methods for fixed compression ratio.

(From literature) Compressor	PRD (%)		Compression Ratio
	Compressor	Proposed Compressor	
Optimized quantization of DCT coefficients [2]	1.5 2.0 2.5 3.0	1.2 1.6 2.0 2.3	6.2 7.9 9.3 10.9
Mean-shape VQ [3]	4.1	2.8	13.1
Wavelet compression by SPIHT[4]	1.2 3.0	0.8 2.1	4.0 10.0
Peak selection and DCT [5]	3.0	1.0	5.3
Sub-band compressor/ F16B FIR filter [6]	2.8*	2.2*	7.3
AZTEC [7]	15.5*	2.0*	6.9
SAPA [7]	9.6*	2.0*	6.9
LTP [7]	7.3*	2.0*	6.9
$ASEC_{PRD}$ [7]	4.0*	2.0*	6.9
Gold washing adaptive VQ / WT [8]	3.3* 6.3* 8.2*	1.4* 2.8* 3.8*	4.6 9.4 12.4

References

[1] S. M. Jalaleddine, C. G. Hutchens, R. D. Strattan and W. A. Coberly, "ECG data compression techniques-A unified approach," IEEE Transactions on Biomedical Engineering, vol. 37, no. 4, pp. 329–343, 1990.

[2] L. V. Batista, E. U. Melcher and L. C. Carvalho, "Compression of ECG signals by optimized quantization of discrete cosine transform coefficients," Medical Engineering & Physics,Elsevier, vol. 23, pp. 127–134, 2001.

[3] J. C. Barreras and J. L. Ginori, "Mean-shape vector quantizer for ECG signal compression," IEEE Transactions on Biomedical Engineering, vol. 46, no. 1, pp. 62–70, 1999.

[4] Z. Lu, D. Kim and W. Pearlman, "Wavelet compression of ECG signals by set partitioning in hierarchical trees algorithm," IEEE Transactions on Biomedical Engineering, vol. 47, no. 7, pp. 849–856, 2000.

[5] L. V. Batista, E. U. Melcher and L. C. Carvalho, "An ECG compression method using peak selection and discrete cosine transform (in Portuguese)," Brazilian Journal of Biomedical Engineering, vol. 16, no. 1, pp. 39–48, 2000.

[6] J. Husoy and T. Gjerde, "Computationally efficient sub-band coding of ECG signals," Medical Engineering and Physics, vol. 18, no. 2, pp. 132–142, 1996.

[7] Y. Zigel, A. Cohen, A. Abu-Ful, A. Wagshal and A. Katz, "Analysis by synthesis ECG signal compression," Computers in Cardiology, vol. 24, pp. 279–292, 1997

[8] S. Miaou and H. Yen, "Quality driven gold washing adaptive vector quantization and its application to ECG data compression," IEEE Transactions on Biomedical Engineering, vol. 47, no. 2, pp. 209–218, 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effects of White Matter on EEG of Multi-layered Spherical Head Models

Md Rezaul Bashar[1,3], Yan Li[1], and Peng Wen[2]

[1]Department of Mathematics and Computing, [2]Faculty of Surveying and Engineering, Centre for Systems Biology, University of Southern Queensland, QLD 4350, Australia
[3]Department of Information and Communication Engineering, Islamic University, Kushtia, Bangladesh
E-mail: {bashar, liyan, pengwen}@usq.edu.au

Abstract - Biological tissues are multi-compartmental inhomogeneous media composed of different cellular and subcellular domains. Human head, a multi-compartmental inhomogeneous medium, is composed of scalp, skull, cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), and other subcellular domains. Among these domains, skull and WM show the complicated anisotropy tissue property because of their physiological structure. However, many researchers model human head excluding WM. This research investigates the necessity of WM using four- and five- layered spherical head models. Four-layered head model excludes WM while five-layered model includes it. The piecewise homogeneous forward model using finite element method is implemented to measure electroencephalogram (EEG) on head surface for both head models. Analyzing these EEGs, this research finds the necessity of using WM to make an accurate human head model.

I. Introduction

Electroencephalogram (EEG) measured from the scalp using electrodes is originated by the neural activity of the brain. This neural activity is modelled with a distribution of current sources. The method of estimating this current distribution is known as EEG source localization or source analysis [1][2][3]. The EEG source analysis is an important tool for diagnosing neurological disorders (such as epilepsy), origin of evoked potentials and in research into cognitive brain functions [1][2]. It is measured by solving the forward and inverse problems [1]. The forward problem is solved by calculating the electrode potentials on scalp for a given source configuration. Inverse problem is solved by iterative solving of forward problem to estimate the current source inside the brain with known electrode potentials [1] [4]. Here is a serious problem that a small error in forward model leads a severe error in inverse model. Therefore, solving forward problem plays a vital rule for accurate source analysis. The solution of forward problem also requires a head or volume conductor model. Volume conductor model ranges from simple homogeneously conducting sphere to complex numerical model consisting millions of elements, each with a different conductivity value [3], for example, the homogeneous sphere model, three-or four-layered concentric model [4], isotropic or anisotropic multi-layered concentric [4], and realistic head models [5]. Zhang *et al.* [6] mentioned that using the homogeneous model to approximate the three-layered concentric model leads to error. Their results recommend that it is better to avoid oversimplified models. The three-layered spherical head model ignores the cerebrospinal fluid (CSF) compartment between the cortex and skull. However, most recently, Wendel *et al.* [7] proved the significant of CSF to construct an accurate head model. In general, these head models assume that the conductivity within a spherical compartment is constant. However, in reality, conductivities at each point in the head are unique, even though they may be of the same tissue type. Moreover, the skull and white matter (WM) of brain compartment exhibit direction related (radial or tangential) conductivities. Previous head construction methods described in different literature neglected WM for anisotropy consideration. However, Hallez *et al.* [8] has shown that neglecting WM anisotropy in the spherical head model arise a 15mm dipole localization error.

Neuronal source localizations based on EEG require a mathematical model in order to compute the electrical potential distribution resulting from dipole or distributed source inside the head. A critical component of source reconstruction is the numerical approximation method used to reach an accurate solution of the associated forward problem. The forward problem requires boundary element method (BEM) or finite element method (FEM) to approximate the solution in addition to source and volume conductor model [5]. In clinical practice, simple volume conductors are still in use, however, recent advances in head modelling techniques make the use of high resolution volume conductor by means of FEM [3]. Only the FEM is able to treat head modelling, electrical properties of each tissue and anisotropy information to include [3][5].

In this paper, we examine how WM affects the scalp potentials to study the effects of WM on constructing an accurate human head model. We generate four-layered spherical head models consisting of scalp, skull, CSF and brain. Again, we generate a five-layered head model consisting gray matter (GM) and WM as separate layer while other layers (scalp, skull and CSF) are identical. We tessellate both head models into approximately the same number of elements. We measure EEG for both head models using FEM with the same electric source. We

analyse both EEG models by means of relative difference measure (RDM) and magnification (MAG) error measurements to exhibit the influence of WM.

II. Problem Statement

Conductivity assignment is a critical problem for anisotropic head model generation. To assign proper conductivity to brain tissue, it is to be separated into GM and WM because of their tissue property. To classify brain into GM and WM, we construct two types of head models: i) with the concept of fractional anisotropy (Head model A) and ii) with the concept of effective medium approach (Head model B). In addition, we also mention Head model C to present five-layered model.

A. Head Model A

Li *et al.* [9] classified WM and non-WM tissues from brain tissue layer using fractional anisotropy (FA) based images captured from diffusion tensor magnetic resonance image (DT-MRI) [10]. FA is a technique to measure the extent of the anisotropy property for each voxel (element). Let us suppose that λ_1, λ_2, and λ_3 ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) are the three eigen values of diffusion tensor matrix and λ is the average eigen value. Then FA is defined as [9]:

$$FA = \frac{\sqrt{3}}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda)^2 + (\lambda_2 - \lambda)^2 + (\lambda_3 - \lambda)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (1)$$

The FA is in the range from 0 to 1. A fully anisotropic tissue has a factor of FA=1, and an isotropic tissue has a factor FA=0. Kim *et al.* [10] separated WM from non-WM tissues considering WM tissues have larger FA values than non-WM tissues. Applying the FA technique, WM and non-WM tissues are classified according to equation (2) [11]. However, in this research, it is assumed that if FA is less than threshold, the tissue is considered as non-WM; otherwise WM. Non-WM tissues are considered as GM all over this study. Tissue conductivities are randomly selected using the *tissue_type* and isotropic or anisotropic conductivity values are assigned to classified tissues mentioned by the literature [5]. The *tissue_type* is defined as

$$tissue\_type = \begin{cases} 0, & FA < threshold \\ 1, & FA \geq threshold \end{cases} \quad (2)$$

B. Head model B

In reality, due to the complicated structure and direction of nerve bundles, the anisotropy ratio is not constant every where in the brain [1]. One can estimate diffusion tensor from diffusion weighted MRI (DW-MRI) in each element to determine the direction of nerve bundles. Analysing the diffusion tensor at each element, anisotropic ratio (AR) is defined as the ratio of the largest eigen value to the mean of the two other eigen values [1].

$$AR = \frac{d_1}{mean(d_2, d_3)} \quad (3)$$

where d_1 is the largest eigen value and d_2, d_3 are two smallest eigen values of the diffusion tensor at a specific element. The values of AR lie between 1 and 10. An element is said to be isotropic where the ratio of d_1 and the mean of d_2 and d_3 are close to 1 (AR ≅ 1) [1]. Based on this concept, we consider d_1 as longitudinal eigen value as it is the largest eigen value, and d_2 and d_3 are two other transverse eigen values where ($d_1 \geq d_2, d_3$). We also assume the values of these eigen values are determined considering Volume constraint [5], which retains the geometric mean of the eigen values. While we determine *AR*, we apply similar process of Head model A to classify GM and WM as

$$tissue\_type = \begin{cases} 1, & AR < threshold \\ 10, & AR \geq threshold \end{cases} \quad (4)$$

Classifying the brain tissues, we assign isotropic conductivity to GM and anisotropic conductivity to WM elements and are ready to perform forward computation.

C. Head model C

It is well known that human head is composed of scalp, skull, cerebrospinal fluid (CSF), gray matter (GM), WM, and other sub cellular domains or compartments [6]. Therefore, we construct five-layered spherical head model to represent above mentioned compartments [8][12][13]. Among these compartments, skull and WM show the complicated tissue anisotropy property because of their physiological structure. The anisotropy ratio of WM is considered with the value of 1:10 (transverse: longitudinal) [5][8]. Based on the literature [5][12], the WM conductivity tensor (σ) is assumed and the tensor eigen values are determined considering Volume constraint, as shown in Table 1. Assigning conductivities to each piecewise element, forward computation is performed by means of FEM to measure the electric potentials at scalp.

Table 1 Simulated values for the WM tensor eigen values

σ_{trans}: σ_{long}	σ_{trans}	σ_{long}
1:1 (isotropic)	0.14	0.14
1:2	0.11	0.222
1:4	0.088	0.353
1:8	0.07	0.56
1:10	0.0649	0.65

The electric potentials at scalp are measured by using forward computation with a known current source. The electric field $\mathbf{E}$ is obtained as the negative gradient of scalar potential, φ, that is $\mathbf{E} = -\nabla\varphi$. According to Ohm's law, the current density $\mathbf{J}$ and $\mathbf{E}$ are related as $\mathbf{J} = \sigma\mathbf{E}$, where σ is the conductivity tensor of the medium. In the event that a source density $\mathbf{I}_v$ is present, then $\nabla\mathbf{J} = \mathbf{I}_v$. Finally, the relationship between φ and $\mathbf{I}_v$ can be given as [4][5]

$$\nabla\mathbf{J} = \mathbf{I}_v = -(\nabla \bullet \sigma(\nabla\varphi)) \quad (5)$$

Equation(5) is solved using Dirichlet and Neumann boundary conditions noting that current can passes only one head layer to another but there is no current getting out of the scalp, respectively [4].

$$\varphi = \varphi_0 \quad \text{on inner surface} \tag{6}$$

$$\sigma(\nabla\varphi) \bullet \mathbf{n} = 0 \quad \text{on outer surface} \tag{7}$$

where **n** is unit normal. Some literature found where the analytic solution [8] is used to solve the electric potential problem for isotropic and homogeneous purposes. However, due to complicated head structure, anisotropy and inhomogeneity, some numerical techniques must be employed for solving the forward problem [1][4]. One of the most common numerical techniques is FEM which computes an estimation of the potential field over each element, taking into account the material properties of each individual element. Therefore, it is possible to specify different conductivities over different regions, or even for each element.

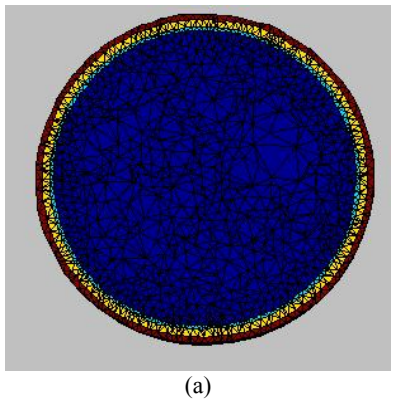
We use two error criteria that are commonly used in different literature [4][5][12] to analyse the measured EEGs. We use relative difference measure (RDM) and magnification (MAG) errors, to compare the forward solutions under different conductivity approximations. The minimum RDM value is 0 and MAG value is 1.

III. Implementation and Experimentation

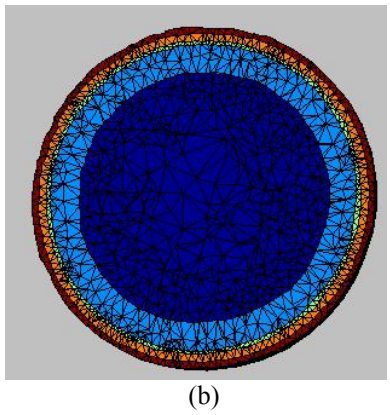
At first, we implement a four-layered spherical head model [4] with different radii for different tissue layers (shown in columns 2 and 3, in Table 2) using Matlab [14] for Head models A and B. Head modelling using FEM requires mesh generation. Mesh represents the piecewise geometry of the head where electric properties are easily included. The first step of mesh generation is to create the surfaces with a given distances d_r from the centre of the spherical head. The second step is the generation of the vertices of the tetrahedral mesh. The third step is the computation of Delaunay triangulation [14] for all generated vertices. The Delaunay triangulation confirms that no vertex resides inside the circumsphere of any generated tetrahedron. In the final step, each generated tetrahedron is labelled as to which each compartment it belongs to. Then the conductivities are assigned to each finite element. We create surfaces with 14 mm thinning for different tissue layers. We mesh the sphere into 315K elements (shown in column 4 in Table 2) from 54K nodes using Tetgen® package provided by Brainstorm [15]. We assign homogeneous isotropic conductivities (shown in last column in Table 2) for different tissue layers according to the literature [4][5] and perform forward computation for the reference model. Later on, we classify the brain layer into GM and WM tissues using Head model A and B techniques implemented in Matlab. Then we assign anisotropic conductivities to WM while isotropic conductivities to GM and other compartments for measurement models. To construct five-layered head model (Head model C), we segment the sphere into five compartments using the same surface thinning of four-layered model. The details of five-layered head model construction are described in our previous papers [12][13]. Brain tissue layers of four-layer model consists 118366 piecewise elements. Analogy, GM consist 66665 elements and 50489 elements are in WM for five-layered model. Approximately, 99% elements of the brain compartment are contained within GM and WM. Moreover, there is a surface between GM and WM. For this model, we segment into approximately the same number of piecewise elements. Table 2 shows the parameters for both head models and Fig. 1 shows the meshed head models. Analogous to previous models, we assign isotropic conductivities to each tissue layer for reference model and we assign WM anisotropic conductivity with the ratio of 1 to 10 (transverse to longitudinal) keeping other compartment isotropic for measurement models. In our research, we consider longitudinal conductivity only.

Table 2 Head model parameters

Head model		Radii	Elements	Isotropic conductivity
Four-layered	scalp	8.8	51792	0.33
	skull	8.5	66427	0.0042
	CSF	8.1	78852	1.00
	brain	7.9	118366	0.33
Five-layered	scalp	8.8	52519	0.33
	skull	8.5	67403	0.0042
	CSF	8.1	78846	1.00
	GM	7.9	66665	0.33
	WM	6.5	50489	0.14



(a)



(b)

Fig. 1 Meshed head models: (a) four-layered head model and (b) five-layered head model.

After constructing the head geometry and assigning the conductivities to each layer, we perform the forward computation. As the brain activity can originate from GM only, we assume the dipole located in axial, coronal and sagittal planes. A dipole can be decomposed into three orthogonal dipoles along the main axis, therefore, we consider three orthogonal orientations. These orientations are X orientation (along left-right), Y orientation (along back-front) and Z orientation (along bottom-top). We choose the orientations for the dipole indicated by an azimuth angle $\theta \in [-\pi, \pi]$ and an elevation angel $\varphi \in [-\pi/2, \pi/2]$. We place a fixed dipole at 2mm below the cortex surface inside the GM at the right hemisphere with the value of $\pi/4$ and $\pi/5$ for θ and φ, respectively. We choose the unit magnitude of the dipole. Then we solve equations (5) to (7) using FEM into a set of linear equations. These linear equations are solved by preconditioned conjugate gradient method using Cholesky factorization preconditioning [14] with a drop tolerance of $1e^{-4}$. In this study, 64 electrodes provided by Brainstorm are used to measure the EEG on scalp. Finally, we measure the electric potentials on scalp using the electrodes, where the electrode potentials $\mathbf{V} \in \Re^{m\times 1}$, m is the number of electrodes. We implement these models using an Inter® dual core 2.0 Ghz processor. A single computation for the FEM used in this research takes more than three hours CPU time.

IV. Result Analysis

In this study, we carry out forward computations to measure EEGs for Head models A, B and C independently. For Head model A and B, firstly, we calculate the electric potential differences (EEG) on scalp using the isotropic conductivities for each tissue layer and use for reference model. Then, we logically divide the brain tissue layer into GM and WM based on different thresholds using either Head model A or B. We use 0.2, 0.4, 0.5, 0.6 and 0.9 values for different thresholds. We assign anisotropic conductivities into WM but remaining other tissue layers are isotropic. Finally, we perform forward computation to measure EEGs and we consider these as measurement models.

Fig. 2 shows the RDM and MAG errors generated by Head model A. In this experiment, when we use 0.2 values for threshold, we find very few GM elements (approximately 10% of brain elements). As a result, it generally produces more errors. We also find that the RDM and MAG values are high for 0.2 threshold value. We measure RDM values are between 2.2% and 18%, 17% and 38%, and 3% and 178% for X, Y and Z orientations, respectively. Therefore, on an average, it produces 72% RDM differences. Similarly, it produces 0.58 average MAG errors. When the threshold value is 0.9, RDM and MAG errors exhibit their ideal values because the brain tissue classification based on equation (2) generates 99% GM elements, i.e. as reference model. For other threshold values, the number of elements varies in the resulting variations shown in Fig. 6.

Fig. 3 shows the RDM and MAG errors generated by Head model B. RDM values for X orientation are in the range of 10% to 75%, for Y orientation lie between 8% to 31% and for Z orientation are from 3% to 45% except for the threshold value 9. Therefore, Head model B produces average 43% RDM error. Similarly, it also produces 1.7 average MAG error. For the maximum threshold, it generates the maximum GM elements; as a consequence, it produces the same electrode potentials like the reference model. In this case, we consider 2, 4, 5, 6, and 9 values for AR thresholds defined in equation (4).

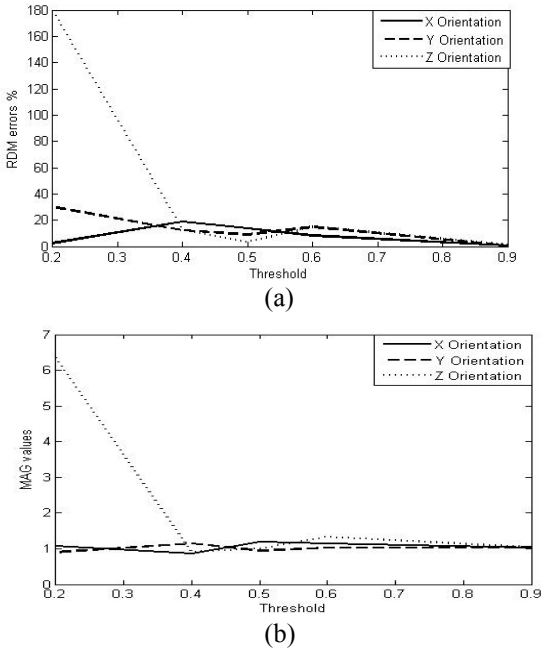


Fig. 2 RDM and MAG errors produced by reference model and anisotropic model generated by WM anisotropic conductivity using Head model A. (a) presents RDM errors against thresholds and (b) shows MAG values vs thresholds.

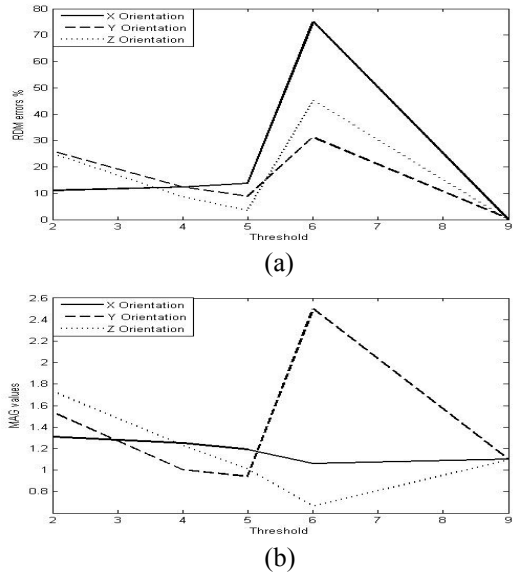
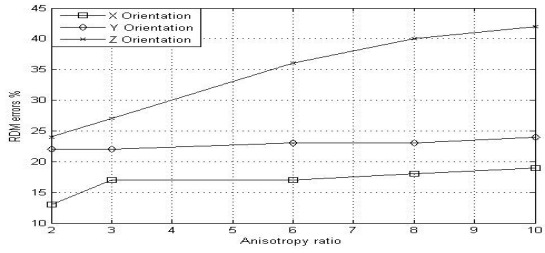
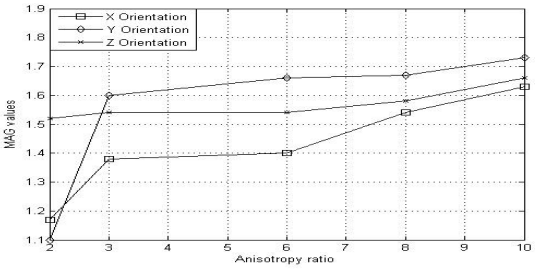


Fig. 3 RDM and MAG errors produced by reference model and anisotropic model generated by WM anisotropic conductivity using Head model B. (a) presents RDM errors against thresholds and (b) shows MAG values vs thresholds.

Fig. 4 shows the same RDM and MAG errors (shown in Figs. 2 and 3) produced by Head model C with different anisotropy ratios. With the increasing of anisotropy ratio, the conductivity values are also increased, shown in table 1. The more conductivity value increases, the more electrode potential differs from the reference. As the number of WM elements are fixed, it produces smooth upward curves for increasing conductivities with increasing anisotropy ratio. The RDM variation in X orientation is 6%; 4% is for Y orientation and 18% is for Z orientation. Therefore, it produces only 9% average RDM error. It also produces 0.36 average MAG. Analyzing the results produced by Head model A and B shown in Figs. 2 and 3, we observe that Head model C produces less errors than others.



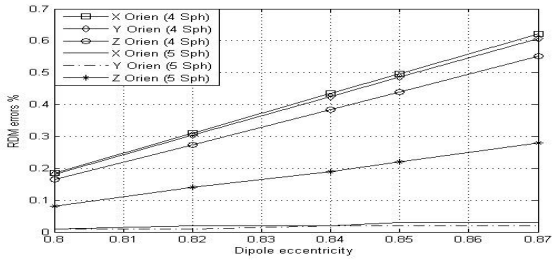
(a)



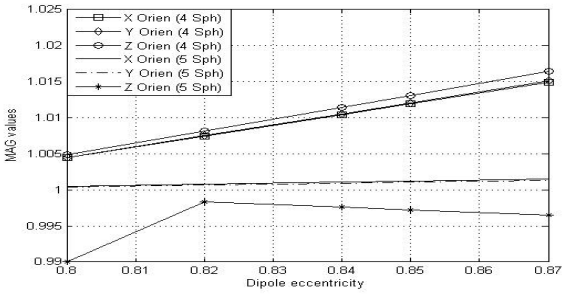
(b)

Fig. 4 RDM and MAG errors for different WM anisotropy ratio on X, Y and Z orientations in Head model C using longitudinal conductivity. (a) Anisotropy ratio vs RDM errors and (b) Anisotropy ratio vs MAG values.

Fig. 5 shows RDM and MAG errors for dipole eccentricity where the more eccentric the dipole, the greater the differences. The dipoles are located at the same places for both four-and five-layered head models. As the depth of GM is limited due to five-layered head model, we place the dipole from 2mm outer the WM surface to 2 mm inner the cortex. As a result, dipole eccentricity starts from 0.8 and finishes to 0.87. For the reference EEG of this computation, we assume the dipole is located at 2mm outer the WM surface. Comparing the errors, we realise that four-layered head model produces more errors than five-layered head models. For instance, the maximum RDM and MAG differences between these two head models for X orientation are 0.6% and 0.0148, respectively, where the eccentricity is 0.87 (the point closer to cortex).



(a)



(b)

Fig. 5 Dipole eccentricity for X, Y and Z orientations for four- and five- layered spherical head model. (a) Dipole eccentricity vs RDM errors and (b) Dipole eccentricity vs MAG values.

V. Discussion

In this paper, we have investigated the effects of WM on EEG scalp potentials to analyze the influence of WM to construct an accurate head model for forward as well as inverse computations. Most of the spherical head models are either three- or four-layered. However, there is limited spherical head model consisting five layers. In reality, it is difficult to segment GM and WM from brain. Moreover, the cortex (the outer part of the brain) and thalamus (the core part of the brain) area are belonged to GM. So, there arises a question is it necessary to separate GM and WM from the brain tissue layer?

The classification of brain tissue has an important application in studying the structure and function of the brain [9]. Diffusion coefficient of water molecule in brain tissue decreases quickly after stroke and other brain injuries by thirty to forty percentages [16]. Diffusion tensor imaging (DTI) is widely used in the study and diagnosis of neurological diseases involving in the WM such as stroke, tumors, multiple sclerosis, dyslexia, and schizophrenia [16]. However, many neurological and neurodegenerative diseases, such as Alzheimer's and Creutzfeldt-Jakob diseases are generally considered involving the GM [17]. Thus, it becomes necessary to separate GM and WM to study and investigate neuronal diseases. On the other hand, most head models in different literature assume an isotropic conductivity for brain tissues. However, GM has isotropic and WM has anisotropic conductivities. As spherical head models are easy to construct, many researchers use it to analyze scalp potentials or source reconstruction. Using spherical head models, we have shown that the head model consisting WM produces less average errors than the head model excluding WM. Therefore, we consider that it is important to classify the brain and we suggest that excluding WM

compartment will be a cause of inaccurate head model construction, which will greatly affect on the EEG source reconstruction.

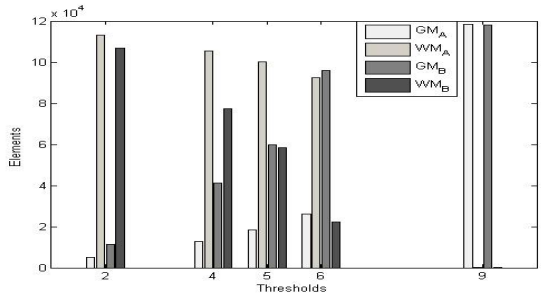


Fig. 6 Number of classified elements for GM and WM of Head model A and Head model B based on different thresholds. In the figure, subscript A represents Head model A and B represents Head model B. Thresholds of Head model A is multiplied by 10.

GM and WM are classified according to either Head model A or B by means of different thresholds, as shown in Fig. 6. Errors are changing with the changing of thresholds (Fig. 2 and Fig. 3). From Fig. 6, when threshold=5, it produces 59663 and 58703 elements for GM and WM, respectively, for Head model B. Coincidentally, we get the close numbers of GM and WM elements to those elements of Head model C. The RDM and MAG errors shown in Fig. 3 are close to those errors produced by Head model C (shown in Fig. 4) for the maximum anisotropy ratio. While for threshold=9, the number of GM elements is 118242, that is most of the brain elements. Therefore, it produces the minimum errors. As no brain signal can be generated from WM, we are bound to locate the dipole inside the GM. In the five-layered head model, the width of GM is 14mm. In reality, the GM is located at the cortex, thalamus and cerebellum region. Therefore, dipoles can be placed at any position within these regions. We shall investigate more similar studies on realistic head model using MRI in the near future.

VI. Conclusion

This study investigates the importance of WM on human head modeling. In this paper, we implement four- and five-layered head models excluding and including WM compartment. The preliminary results obtained in this research using statistical error quantifications are testified to investigate the effects of excluding and including WM on EEG. Performing different experiments, we find that the head model excluding WM produces 72% and 0.58 average RDM and MAG errors, respectively, whereas the WM inclusion head model produces 9% and 0.36 average RDM and MAG errors, respectively, with their corresponding reference model. Consequently, the EEG model that excludes WM layer causes more dipole eccentricity errors than the EEG model that includes WM layer. This study suggests that an accurate head model requires WM layer to measure error pruned EEG.

Acknowledgement

This project is supported by Australian Research Council Discovery Program DP0665216.

References

[1] Hallez, H., *et al.*, "Dipole estimation errorsdue to differences in modeling anisotropic conductivities in realistic head models for EEG source analysis," Journal of Phys. Med. and Bio., vol. 53, pp. 1877-1894, 2008.

[2] J. O. Ollikainen, *et al.* "Effects of local skull inhomogeneities on EEG source estimation," Med. Eng. and Phy., vol. 21, pp. 143-154, 1999.

[3] J. Haueisen, "The influence of forward model conductivities on EEG/MEG source reconstruction," IEEE Proceedings of NFSI & ICFBI, pp. 18-19, 2007.

[4] P. Wen, "Human Head Modelling and Computation for the EEG Forward Problem," PhD dissertation, The Flinders University of Western Australia, Australia, 2000.

[5] C. H. Wolters, "Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain," PhD dissertation, University of Leipzig, France, 2003.

[6] Z. Zhang and D. L. Jewett, " Insidious error in dipole localization parameters at a single time-point due to model misspecification of number of shells," Electroence.. Clin. Neurophy.l. vol. 88, pp. 1-11, 1993.

[7] K. Wendel, *et al.*, "The Influence of CSF on EEG Sensitivity Distributions of Multilayered Head Models," IEEE Trans. on Bio. Engr., vol. 55, no. 4, pp. 1454-1456, 2008.

[8] H. Hallez, *et al.*, "Dipole Localization Errors due to not Incorporating Compartments with Anisotropic Conductivities: Simulation Study in a Spherical Head Model," IJEM, vol 7, no. 1, pp. 134-137, 2005.

[9] H. Li, *et al.*, "Brain Tissue Segmentation based on DWI/DTI Data," 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.

[10] S. Kim *et al.*, "Influence of Conductivity Tensors in the Finite Element Model of the Head on the Forward Solution of EEG," IEEE conference on Nuclear Science Symposium Conference Record, 2001.

[11] L. Li, *et al.,* "A Study of White Matter Anisotropic conductivity on EEG Forward Solutions," IEEE Proceedings of NFSI & ICFBI, pp.130-132, 2007.

[12] R. Bashar, *et al.*, "Influence of white matter inhomogeneous anisotropy on EEG forward computing," Australasian Physical & Engineering Sciences in Medicine, vol. 31, no. 2, pp. 122-129, 2008.

[13] R. Bashar, *et al.*, "Tissue Conductivity Anisotropy Inhomogeneity Study in EEG Head Modelling," Int. Conf. on BioComputing (BioComp'08), USA, 2008.

[14] Mathwork software, The Matlab.

[15] S. Baillet, *et al.*, "Electromagnetic Brain Imaging using Brainstorm," IEEE Int. Symposium on Bio. Engr.: Macro to Nano, pp. 652-655, 2004.

[16] P. N. Sen and P. J. Basser, "A Model for Diffusion in White Matter in Brain," Biophysical Journal, vol. 89, pp. 2927-2938, 2005.

[17] T. Liu, *et al.*, "76-space Analysis of Grey Matter Diffusivity: Methods and Application," MICCAI, Palm Springs, California, USA, 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Acquisition and Analysis of Electrogastrogram for Digestive System Disorders Using a Novel Approach

G. Gopu[1], R. Neelaveni[2], and K. Porkumaran[3]
1. Research Scholar, Dept. of EEE. PSG College of Technology, Coimbatore- 04,
2. Assistant Professor, Dept. of EEE. PSG College of Technology, Coimbatore- 04,
3. Professor & Head, Dept. of BME, Sri Ramakrishna Engineering College, Coimbatore-22.
E-mail: gopugovindasamy@gmail.com

Abstract - The digestive system is the one of the important system in the human body, which plays major role directly or indirectly for the function of human body. Most of the people around the world have the digestive system disorders due to improper digestion of food due to inefficient performance of stomach activity. The main objective of this paper is to propose a novel method of finding the digestive system disorders using Electrogastrogram [EGG], which is a non-invasive, cheap and painless method by detecting the electrical signal from the stomach cutaneously and also it act as a preliminary investigation without a need for Endoscopy which is painful investigation. The recording setup explained in this proposed system includes LabVIEW software and hardware which is used to record the EGG for more than hundred patients, nearly 75% of the patients suffered from results in digestive system disorders such as Dyspepsia, Stomach ulcer, nausea, cyclic vomiting syndrome, etc. For the above said digestive system disorders dissimilarity is found in its frequency and amplitude compared with its normal individual parameter (3cpm) at a fair amount of accuracy.

I. Introduction

An Electrogastrogram (EGG) is a non-invasive test used to measure gastric myoelectrical activity [2, 18]. The normal gastric myoelectrical activity consists of a slow wave and spike potential. The abnormality arises due to recurrent nausea, vomiting, Dyspepsia, Stomach ulcer, Cyclic vomiting syndrome, *etc* which signals that the stomach is not emptying food normally. If the electrogastrogram is abnormal, it confirms that the problem probably is with the stomach's muscles or the nerves that control the muscles. This paper deals with the novel approach of recording of the electrical signals that travel through the muscles of the stomach and control the muscle's contraction. The EGG can be considered as an experimental procedure since its exact role in the diagnosis of digestive disorders of the stomach has not been defined yet.

II. Electrogastrogram

An electrogastrogram is similar to an electrocardiogram of the heart. It is a recording of the electrical signals that travel through the muscles of the stomach and control the muscle's contraction. It is used when there is a suspicion that the muscles of the stomach or the nerves controlling the muscles are not working normally. It is done by placing the electrode cutaneously over the stomach and the electrical signals coming from the stomach's muscles are sensed by the electrode and recorded on a computer for analysis by lying patient quietly. In normal individuals the electrogastrogram is a regular electrical rhythm generated by the muscles of the stomach and the power (voltage) of the electrical current increases after the meal. In patients with abnormalities of the muscles or nerves of the stomach, the rhythm often is irregular or there is no post-meal increase in electric power. It will not have any side effects and it is painless study.

III. Proposed EGG Recording Setup

The Ag/Agcl electrodes [5, 12] are used as a sensor to record the electrical activity of the stomach's muscle cutaneously.The electrical signals generated are usually of very low amplitude ranging from 0.01 to 0.5 mV is given to signal conditioning unit (SCU) because proper conditioning of signals is necessary to produce analog signal without noise before giving it to the ADC. The SCU consist of instrumentation amplifier and low pass

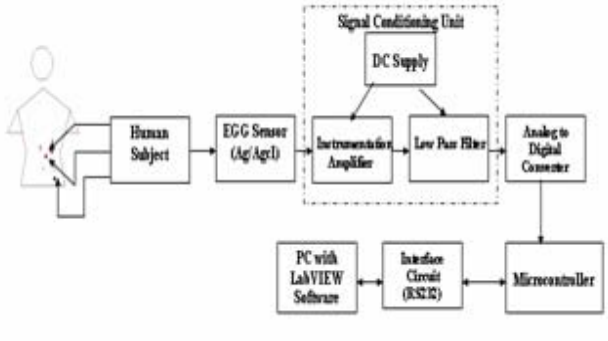


Fig.1 General Block diagram for recording EGG

filter. The instrumentation amplifier has a gain of 1000 to 10,000. The amplified signals are input to a second order low pass Butterworth filter to remove the noises and ripples. The analog output of the filtered signal is converted to digital through Analog to Digital Converter (ADC).These outputs are then given to a microcontroller circuit which transfers the digital outputs to the PC via RS232 to view the recorded signal in a readable form. The

LabVIEW software which uses the graphical data flow programming technique for the purpose of recording EGG for the investigation of digestive system disorders as shown in Fig.1.

IV. The Anatomy of the Stomach

The main function of the stomach is to process and transport food [2]. After feeding, the contractile activity of the stomach helps to mix, grind and eventually evacuate small portions of chyme into the small bowel, while the rest of the chyme is mixed and ground. Anatomically, the stomach can be divided into three major regions: fundus (the most proximal), corpus and antrum. Histologically, the fundus and corpus are hardly separable. In the antral area, the density of the smooth muscle cells increases. The area in the corpus around the greater curvature, where the split of the longitudinal layers takes place, is considered to be anatomically correlated with the origin of gastric electrical activity. The stomach wall, like the wall of most other parts of the digestive canal, consists of three layers: the mucosal (the innermost), the muscularis and the serosal (the outermost). The mucosal layer itself can be divided into three layers: the mucosa (the epithelial lining of the gastric cavity), the muscularis mucosae (low density smooth muscle cells) and the submucosal layer (consisting of connective tissue interlaced with plexi of the enteric nervous system). The second gastric layer, the muscularis, can also be divided into three layers: the longitudinal (the most superficial), the circular and the oblique. The longitudinal layer of the muscularis can be separated into two different categories: a longitudinal layer that is common with the esophagus and ends in the corpus, and a longitudinal layer that originates in the corpus and spreads into the duodenum as shown in Fig 2.

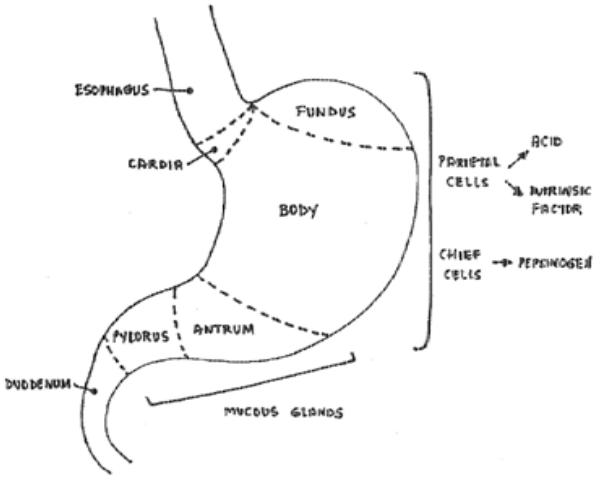


Fig.2 The Anatomy Of The Stomach

V. Electrodes Positioning

The electrical signals are generally produced in the mid-corpus of the stomach where the electrical activity takes place. The positioning of the Ag/ Agcl electrodes for tapping of these signals is as follows as shown Fig 3: Two electrodes A and B are placed in the fundus and the mid corpus of the stomach. The third electrode C is placed as ground at the end of the stomach region for patient safety [4].

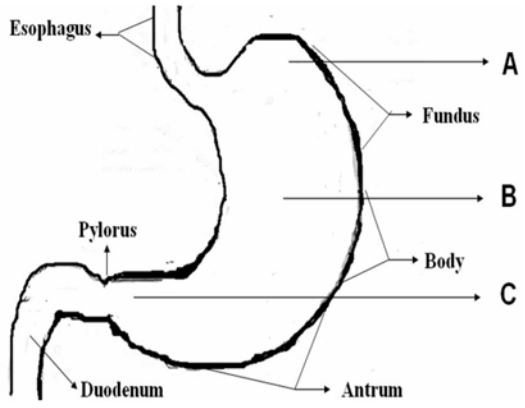


Fig.3 Electrode Positioning for Recording EGG

VI. Materials and Methods

Patients consecutively attending a hospital outpatient's gastroenterology clinic were studied. All patients were considered by their general practitioner to have significant symptoms to merit referral to a specialist clinic for further evaluation and investigation. The patients included 50 with Dyspepsia, 40 with Stomach ulcer, and 25 with Nausea (Table 1). Hematological and biochemical profiles were normal in all patients.

Table .1: Sex and Age Distribution Of Patient Groups

Disorders	Mean Age (years)	Male	Female
Dyspepsia(D) (n=70)	42	28	22
Stomach Ulcer(SU) (n=40)	31	19	31
Nausea(N) (n=25)	45	25	20

A gastrointestinal symptom profile was recorded on all patients immediately before the EGG [3, 7]. This profile detailed the presence or absence of the following dyspeptic symptoms in the previous 2 weeks: Dyspepsia (D), Stomach ulcer (SU), Nausea (N) and Cyclic vomiting syndrome. The EGG was performed after a 6 hours fasting. All medication with the potential affect to gastric function was discontinued for more than 48 hours before the recording. Patients were studied in a semireclaining position and requested to avoid any major movements. The skin was lightly abraded with gauze before placement of adhesive gel EGG electrodes. Two bipolar skin electrodes were placed on the abdomen, one on the fundus and the other on the mid-corpus. A reference electrode was placed on the right side of the abdomen. The electrodes were connected to a signal conditioning unit (SCU). The EGG recording included a 1 hour fasting study, after which the patient ate a sandwich (575 kcal,

50% carbohydrate, 25% protein, 25% fat) and drank 200ml of water. This was immediately followed by another 1 hour recording. Visual inspection of the waveform detected any obvious major movement artifacts. These were defined as abnormally large positive or negative peaks in the tracing and were detected from the analysis. The EGG data was observed by the LabVIEW software running on a personal computer. The EGG frequency ,amplitude are recorded and its respective values are stored for further analysis of the same to investigate the digestive system disorders as shown in Fig.4.

A sampling frequency of 4Hz was used. The EGG analysis is based on the frequency and amplitude of the signals. For the periods before and after the test meal, the frequency and amplitude values are observed. Normal electrical activity was defined as a frequency between 2-4 cycles/minute.

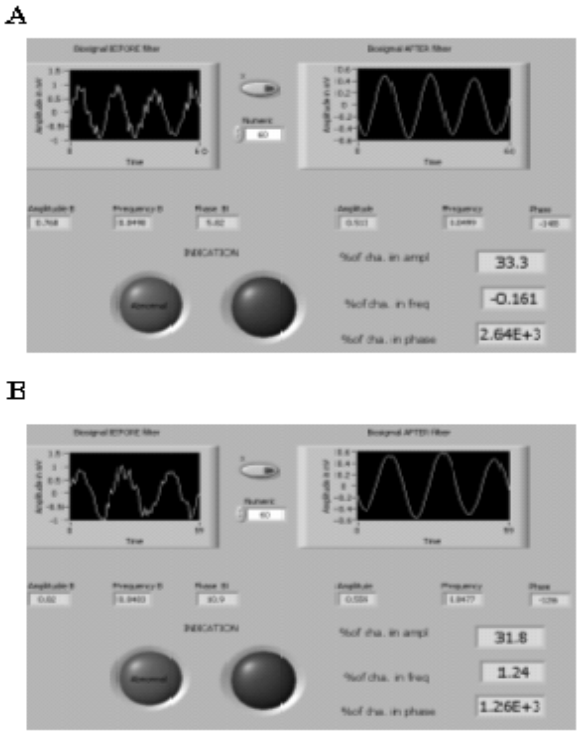


Fig.4 A and B are recorded EGG from Dyspepsia Nausea patients respectively

[9, 17, 20]. Activity of 0-2 cycles / minute was termed bradygastria, and 4-9 cycles / minute as tachygastria. The percentage of normal electrical activity, bradygastria, and tachygastria were calculated both before and after the test meal. The amplitude of the dominant frequency was measured both before and after the test meal. The electrical frequency is stable and does not change significantly after a standard test meal.

VII. Algorithms and Flowchart

The algorithm for this proposed recording setup includes the following steps for recording EGG signal as given below

- Initialization of Analog to Digital Converter
- Enable Serial Port to transmit data.
- Initialization of Timer.
- ADC values are read for amplitude.
- Timer values are read for cycles.
- Conversion of values in to decimal.
- Data are sent to PC through serial port.
- LabVIEW software is used to observe and Record.
- Continue from step 4.

The pictorial representation of the above algorithm is shown in Fig.5

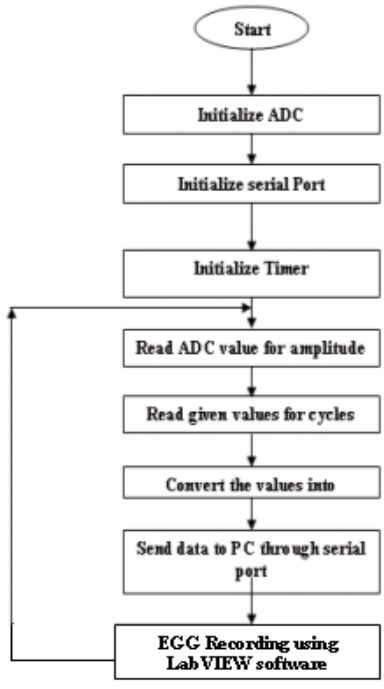


Fig.5 Flowchart of EGG Acquisition System

VIII. Results and Discussions

The effectiveness of the proposed methodology is illustrated and the Table. 2 shows the range of frequencies and amplitudes for various abnormalities at preprandial and postprandial condition as observed from the recording setup in consultation of physician.

Table.2: Frequency and Amplitude Ranges of Various Disorders of Stomach

Conditions	Frequency (cpm)	Amplitude (V)
Preprandial		
Dyspepsia	4-5	0.65-0.95
Nausea	3.5-5.8	0.4-0.6
Vomiting	5.5-6.3	0.54-0.0.72
Stomach Ulcer	6-9	0.6 -0.78
Postprandial		
Dyspepsia	1-2.8	0.5-0.7
Nausea	2-3	0.3-0.8
Vomiting	2-3.5	0.42-0.67
Stomach Ulcer	1.3-2.5	0.3-0.5.5

The Variation in the frequency and amplitude values are observed between preprandial and postprandial condition as Shown in Fig.6

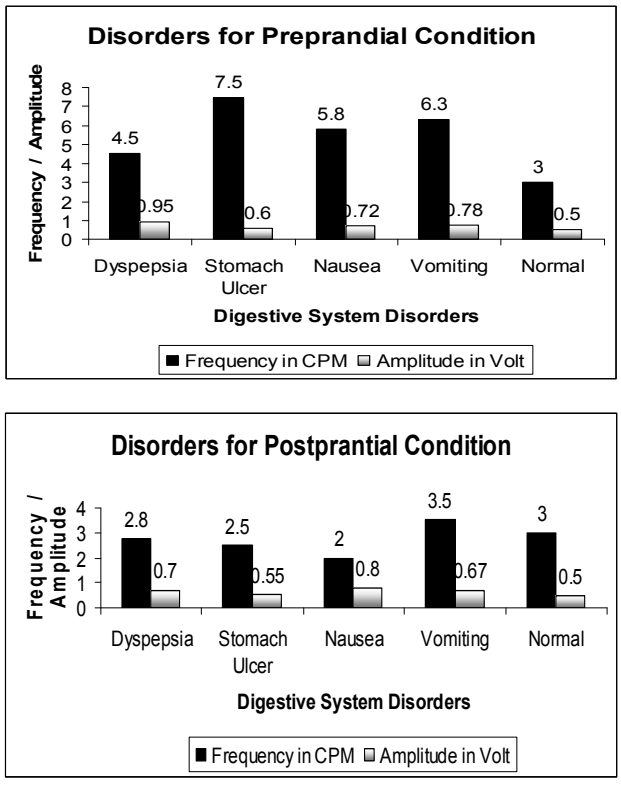


Fig.6 Bar Chart represents the variation in Frequency and Amplitude for the digestive system disorder for preprandial and postprandial condition

The Fig.7 shows the various waveforms obtained for various diseases.

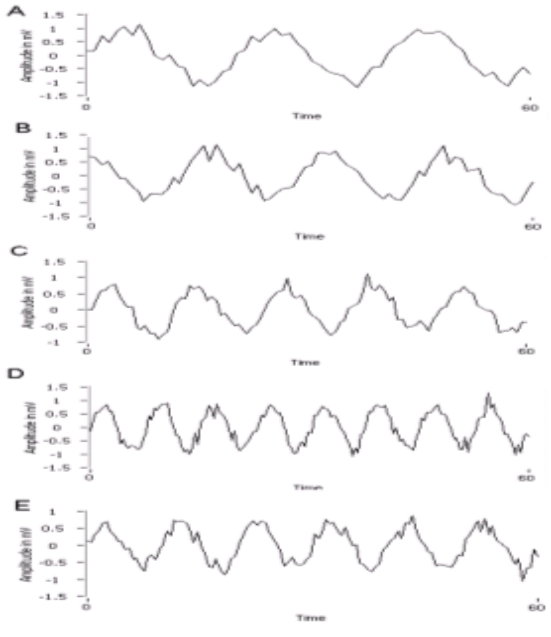


Fig.7 Waveforms for different disease patterns. Normal patient Waveform with 3 cycles per minute as frequency (recording A). A dyspepsia patient wave pattern (recording B). A Nausea patient wave pattern (recording C). A Vomiting patient wave pattern (recording D). An Ulcer patient wave pattern (recording E)

In this study, we have used the EGG to record myoelectrical activity for the patients suffering from Dyspepsia, Stomach ulcer, and Nausea. The abnormal EGG has a high specificity for the detection of abnormal myoelectrical activity [15, 16]. The observation that the test meal can worsen or correct the EGG reflects the interaction between fasting, feeding, and electrical activity. The effect of fasting or feeding indicates that both an abnormality and change in the EGG recording. A postprandial power reduction has been proposed as an EGG abnormality, and correlated with gastroparesis [11]. However, in this study, eight controls demonstrated a postprandial power reduction, indicating that power reduction is an unreliable sign. The significance of power reduction after meals has also been questioned by a recent study comparing internal gastric electrical activity with the EGG [6]. There are important practical implications stemming from this study. In the absence of readily available tests in functional dyspepsia, there has been an assumption that patients with functional dyspepsia entered into clinical trials represent a homogeneous group. This study indicates that it is possible to identify a subgroup of patients with EGG evidence of abnormal gastric myoelectrical activity that is likely to have antral hypomotility. The EGG could help distinguish stomach ulcer from dyspepsia in clinical trials. The treatment of dyspepsia should be aimed at correcting physiological abnormalities [8]. A normal or abnormal EGG will not only help distinguish patient heterogeneity in clinical studies but might also provide a useful objective marker of treatment effect [10,13,14,19]. Further studies of statistical analysis, power spectrum and by using LabVIEW software in dyspepsia patients, with and without an abnormal EGG will help to determine whether these other abnormalities define more coherent or overlapping subtypes of disorder.

IX. Conclusion

A simplified novel approach is proposed in this paper for recording of gastro electrical activity. The EGG setup is used to record the activity of patients suffering from digestive disorders like Dyspepsia (D), Stomach ulcer (SU), Nausea (N) and Cyclic vomiting syndrome. About more than 100 patients were tested using this setup and the results are summarized based on the observation. In future, NI based LabVIEW with Data Acquisition Card (DAC) can be used to perform the recording of EGG to the fair amount of accuracy for further analysis using wavelet transform, statistical analysis ,etc which may improve the accuracy of diagnosis.

ACKNOWLEDGEMENT

The authors acknowledge their indebtedness to the following medical experts Dr L Venkatakrishnan, Dr,J.Krishnaveni, from PSG Hospitals, Coimbatore and Dr.M.G.Shekar, from Stanley Medical College and Hospital, Chennai for their support and for permitting us to use the facilities at the hospitals for live testing of the recording setup and sharing valuable patient database with us.

References

[1] A.B Luckhard , H.T Phillips ,A.J Carlson , "Contributions to the physiology of the stomach." Am. J. Physol., vol.50, pp. 57-62, 1919.

[2] W.C.Alvarez. "The Electrogastrogram and what it shows". vol .78, pp.1116–1118, *JAMA* 1922.

[3] W.C Watson, S.N Sullivan, M Corke, et al. "Incidence of oesophageal symptoms in patients with irritable bowel syndrome", vol 17, pp. 827A, *Gut* 1976.

[4] A.J.P.M Smout, E.J Van Der Schee, J.L Grashuis, "What is measured in electrogastrography." Digestive Diseases and Sciences , A253, 1980.

[5] K L Koch, R M Stern, W R Stewart, et al. "Gastric emptying and gastric myoelectrical activity in patients with diabetic gastroparesis: Effect of long term domperidone treatment." Am. J Gastroenterology, vol. 84, pp. 1069-75, 1989.

[6] J. Chen, R.W. McCallum," Electrogastrography: measurement, analysis and prospective applications". Med Biol Eng Comput, vol.29, pp.339–350, 1991.

[7] J Chen, RW McCallum. "Gastric slow wave abnormalities in patients with gastroparesis." Am J Gastroenterology, vol .87,pp. 477-82, 1992.

[8] KL Koch, M Medina, S Bingman, et al. "Gastric dysrhythmias and visceral sensations in patients with functional dyspepsia", Am J Gastroenterology, A469, 1992.

[9]. RD Rothstein, A Alavi, JC Reynolds, "Electrogastrography in patients with gastroparesis and effect of long term cisapride." Digestive Diseases and Sciences, vol. 38, pp. 1518-24, 1993.

[10] Fumitaka Asa no, Yoshio Yamada, "Method of, and apparatus for, measuring Electrogastrogram and intestinal Electrogastrogram." Digestive Diseases and Sciences, A489, 1995.

[11] Pfaffenban B,Adamek RJ,kuhn K,et al., "Electrogastrography in healthy subjects. Evalation of normal values influences of age and gender." Digestive Diseases and Sciences ,vol. 40, pp. 1445-50, 1995.

[12] NJ Talley, "Review article: Functional dyspepsia- should treatment be targeted on distributed physiology?" Aliment Pharmacol Ther, vol 9, pp. 107-15, 1995.

[13] G Riezzo, S Cucchiara, M Chiloiro, et al. "Gastric emptying and myoelectrical activity in children with non-ulcer dyspepsia. Effect of cisapride." Digestive Diseases and Sciences; vol .40, pp. 1428-34, 1995.

[14] HP Parkman, AD Harris, MA Miller, et al, "Influence of age, gender and menstrual cycle on the normal Electrogastrogram". Am J Gastroenterology, vol .91, pp. 127-33, 1996.

[15] JDZ Chen, Z Lin, J Pan, et al, "Abnormal gastric myoelctrical activity and delayed gastric emptying in patients with symptoms suggestive of gastroparesis." Digestive Diseases and Sciences , vol .41,pp. 1538-45, 1996.

[16] XM Lin, D Levanon, MH Mellow, et al. "Prevalence of impaired gastric myoelectrical activity in patients with functional dyspepsia." Am J Gastroenterology, A199, 1997.

[17] JF Fielding. "The irritable bowel syndrome. Clinic Gastroenterology, vol .6, pp.:607-22. 1997.

[18] M P Mintchev, A Stickel, K L Bowes. "Comparative assessment of power dynamics of gastric electrical activity" , Digestive Diseases and Sciences, vol .42, pp. 1154-7, 1997.

[19] K Besherdas, ACB Leachy, I Mason, et al. "The effect of cisapride on the electrogastrogram and dyspepsia score in nonulcer dyspepsia." Gut, 41(suppl 3):A154, 1997.

[20] Simon Zhao, Giancario Succi, Martin P. Mintchev, "Tele-Electrogastrography." Digestive Diseases and Sciences, A214, 1999.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Bioinformatics Web Data and Service Integration - An Experiemnt with Gene Regulatory Networks

Emdad Ahmed
PhD Candidate, Integration Informatics Laboratory
Department of Computer Science, Wayne State University
Detroit, Michigan, USA
emdad@wayne.edu

Abstract

Web data integration has been an active research topic for the last few years. Bioinformatics is a huge domain wealth of data. By reverse engineering web data, in this paper we report a Java based implementation of Biologically relevant application of web data integration consisting of few sites such as Flybase, YMF and STAMP. We have developed a tool that can automate the finding of transcription factor binding sites for arbitrary Gene Regulatory Networks. We employed vertical (i.e *link*) and horizontal (i.e *combine*) integration for the whole process. As a running example and proof of concept of our approach, we report here a real example motivated by the biological domain of *Drosophila Melanogaster* model organism database. For a given set of genes, we could find a set of sequences (800˜1500 bp) from Flybase site, then we feed those into YMF site to get a motif of 6˜8 bp. The result motifs were fed into STAMP site and by combining result from Fly curated by Bergman and by TRANSFAC, we were able to get a sample correct result for a *Bicoid* gene, namely *bin*. The lesson learned from the exercise can be applied in generic dynamic web form based data integration application like e-commerce, comparison shopping etc.
Keywords: Web Data Integration, mediators, metadata, Ontologies, Web Services, wrapper, gene regulatory network, Ontology, HTML forms, Hidden Web, web data reverse engineering.

1 INTRODUCTION

Web data integration has been an active research topic for the last few years. Bioinformatics is a huge domain wealth of data. We have seen a paradigm shift of web application from static web site to dynamic web services. Data Intensive Web Application (DIWA) has been prolifirated in our daily life i.e in e-commerce, comparison shopping etc. The legacy web based systems are web form based and autonomously developed. How to automatically integrate one site to another has been a research challenge. Here in this paper we report a Java based implementation of web data integration consisting of few sites such as Flybase (www.flybase.org, A Database of Drosophila Genes and Genomes), YMF (http://wingless.cs.washington.edu/YMF/YMFWeb/YMFInput.pl (YMF 3.0: Finds short motifs in DNA sequences), STAMP (http://www.benoslab.pitt.edu/stamp (Alignment, Similarity and Database Matching for DNA Motifs). We have developed a tool that can automate the finding of transcription factor binding sites for arbitrary Gene Regulatory Networks. We employed vertical (i.e *link*) and horizontal (i.e *combine*) integration for the whole process. As a proof of concept of our approach, we report here a real example motivated by the biological domain of Drosophila Melanogaster (*fruit fly*) model organism database. For a given set of genes, we could find a set of sequences (800˜1500 bp) from Flybase site, then we feed those into YMF site to get a motif of 6˜8 bp. The result motifs were fed into STAMP site and by combining result from *Fly curated by Bergman* and by *TRANSFAC*, we were able to get a correct sample result for a *Bicoid* gene, namely *bin*. We materialize the web response in relation database MySQL as it outperform other DB system for Data Intensive Web Appliaction [2]. To the end, our system can solve web data intergration problem in an ad hoc manner as follows:

```
Define function RegulatoryRegion
Extract 800bp sequence
Using wrapper FlyBaseToYMF in ontology GeneMapping
From URL Flybase
Submit FBgnNum  string;
```

```
Define function motif
Extract 6..8bp sequence
Using wrapper FlyBaseToYMF in ontology GeneMapping
From URL YMF
Submit 800bp string;
```

2 RELATED WORK

The present research work is our ongoing work where we have identified and proposed ontology based and declarative workflow query language for ad hoc web data integration on the fly [3, 4]. In our previous work [6, 11, 10, 12, 8, 7], we have implemented a number of prototype systems such as WebFusion, PickUp and OntoBuilder that are much more scalable approach toward web data integration.

3 DATA INTEGRATION

3.1 A Framework for Data Integration

In this section, we describe a formal framework for *data integration systems* (DISs). We will use a unique semantic domain $\triangle$ composed of a fixed, infinite set of symbols representing real world objects. Thus, all databases have $\triangle$ as domain. Formally, a DIS $\mathcal{D}$ is a triple $\langle \mathcal{G}, \mathcal{S}, \mathcal{M}_{\mathcal{G},\mathcal{S}} \rangle$, where $\mathcal{G}$ is the global schema, $\mathcal{S}$ is the set of source schemas, and $\mathcal{M}_{\mathcal{G},\mathcal{S}}$ is the mapping between $\mathcal{G}$ and the source schemas in $\mathcal{S}$. We denote with $\mathcal{A}_\mathcal{G}$ the alphabet of terms of the *global schema*, and we assume that the global schema $\mathcal{G}$ of a DIS expressed in the relational model with two kinds of constraints:

- *Key constraints*: given a relation r in the schema, a key constraint over r is expressed in the form $key(r) = \mathbf{X}$, where $\mathbf{X}$ is a set of attributes of r. Such a constraint is satistfied in a database $\mathcal{DB}$ if for each $t_1, t_2 \in r^{\mathcal{DB}}$ with $t_1 \neq t_2$ we have $t_1[\mathbf{X}] \neq t_2[\mathbf{X}]$, where $t[\mathbf{X}]$ is the projection of the tuple t over $\mathbf{X}$.
- *Foreign key constraints*: a foreign key constraint is a statement of the form $r_1[\mathbf{X}] \subseteq r_2[\mathbf{Y}]$, where r_1, r_2 are relations, $\mathbf{X}$ is a sequence of distinct attributes of r_1, and $\mathbf{Y}$ is a sequence formed by the distinct attributes forming the key of r_2. Such a constraint is satisfied in a database $\mathcal{DB}$ if for each tuple $t_1 \in r_1^{\mathcal{DB}}$ there exists a tuple $t_2 \in r_2^{\mathcal{DB}}$ such that $t_1[\mathbf{X}] = t_2[\mathbf{Y}]$

We assume that a DIS $\mathcal{D}$ is constituted by n sources, and we denote with $\mathcal{S}$ the set of the n corresponding *source schemas* $\mathcal{S}_1, \ldots, \mathcal{S}_n$. We denote with A_{S_i} the alphabet of terms of the source schema $\mathcal{S}_i$ and with $\mathcal{A}_\mathcal{S}$ the union of all the A_{S_i}'s. We assume that the various A_{S_i}'s are mutually disjoint, and each one is disjoint from the alphabet $\mathcal{A}_\mathcal{G}$. Each source schema is expressed in the relational model.

EXAMPLE 1 An example of data integration system is $\mathcal{D}^1 = \langle \mathcal{G}^1, \mathcal{S}^1, \mathcal{M}_{\mathcal{G},\mathcal{S}}{}^1 \rangle$, where $\mathcal{G}^1$ is constituted by the relation symbols `student(`*`Scode, Sname, Scity`*`)`, `university(`*`Ucode, Uname`*`)`, `enrolled(`*`Scode, Ucode`*`)`, and the constraints

$$key(student) = \{Scode\}$$

$$key(university) = \{Ucode\}$$

$$key(enrolled) = \{Scode, Ucode\}$$

$$enrolled[Scode] \subseteq student[Scode]$$

$$enrolled(Ucode) \subseteq university[Ucode]$$

$\mathcal{S}^1$ is constituted by sources s_1 of arity 4, and sources s_2, s_3 of arity 2. Finally, the mapping $\mathcal{M}_{\mathcal{G},\mathcal{S}}{}^1$ is defined as

$$\{\langle student, st(X,Y,Z) \leftarrow s_1(X,Y,Z,W) \rangle$$

$$\langle university, un(X,Y) \leftarrow s_2(X,Y) \rangle$$

$$\langle enrolled, en(X,W) \leftarrow s_3(X,W) \rangle\}$$

Consider the following biological query *"Find the DNA sequences associated with a 3D protein structure"*. In order to answer the query, one need to consult 3D protein structures that are stored in PDB and also to consult DNA sequences that are stored in GENBANK. Let us consider another biological query *"Find the mutation of a 3D protein structure known to cause a disease"*. To answer the query, one need to look up 3D protein structures that are stored in PDB, DNA sequences that are stored in GENBANK, also need to check mutations in sequences that are stored in OMIM. One more example of biological query is as follows: "Find all genes in the human genome that are expressed in the liver and have a TTGGACAGGATCCGA (allowing for 1 or 2 mismatches) followed by GCCGCG within 40 symbols in a 4000 symbol stretch upstream of the gene". In order to answer the query, the following steps are required: perform BLAST or a Smith-Waterman search; combine all the results of step 1; Consult a gene DB; DB search to check if the gene is expressed in the liver.

The above examples show that there is need for Bioinformatics Web Data and Service Integration, which is the focus of this paper.

Table 1: Interoperability challenges

Synthesis
Semantics
Structure
Syntax
System Integration

3.2 Data Integration Challenges

In this section we provide a high-level overview of the data-integration and system-interoperability challenges that often await a scientist who wants to employ IT infrastructure for scientific knowledge discovery. Data heterogeneity has traditionally been divided into *syntax*, *structure*, and *semantic* differences. Scientific data analysis and information integration scenarios often involve additional levels, which include low-level *system* integration issues as well as higher-level *synthesis* issues. We briefly discuss the various levels of heterogeneities and inter-operability challenges below. By system aspects we mean differences in low-level issues relating to, e.g, network protocols (e.g http, ftp, GridFTP, SOAP), platforms (OS) *Sometimes different version of API has significant effect on the performance outcome. For example, we could not make it work auto form filling using JDK 1.5, whereas it works fine for JDK 1.6.* By synthesis we mean the problem of putting together databases, including semantic extensions, queries and transformations, and other computational services into a scientific workflow. The problem of synthesis of such workflow encompasses all previous challenges. For example, if a scientist wants to put together two processing steps A and B into the following simple analysis pipeline:

$$\xrightarrow{x} A \xrightarrow{d} B \xrightarrow{y}$$

many questions arise: In what format does A expect its input x? Does the output d of A (e.g 800 bp sequence from Flybase) directly "fit" the format of B (e.g in YMF), or is a data transformation necessary (e.g from flybase site to construct 800 bp sequence, we need to fill in another form to add 2K sequence) to add? In addition to these *syntactic* and *structural* heterogeneities, *system* and *semantic* issues exist as well. For example, what mechanism should be used to invoke processes and how should data be shipped from A to B? (*need to address auto form filling*) Is it meaningful and valid to connect A and B in this way (*valid web site composition*)? The main challenges for synthesis include process composition and the modeling, design, and execution of reuseable process components and scientific workflows.

3.3 Data Integration Approaches

The data integration systems are characterized by an architecture based on a global schema and a set of sources. The sources contain the real data, while the global schema provides a reconciled, integrated, and virtual view of the underlying sources. Modeling the relation between the sources and the global schema is a crucial aspect. Figure 1 shows a data integration system over several sources that deals with Gene Regulatory Network. Given a user query that is formulated in the query interface (also called the *mediated schema*), the system uses a set of *semantic mappings* to translate the query into queries over source schemas, then executes the queries and combines the data returned from the sources to produce the desired answers to the user. A mediated schema is a set of *virtual* relations, which are designed for a particular data integration application. The relations in the mediated schema are not actually stored anywhere. As a consequence, the data integration system must first *reformulate* a user query into a query that refers directly to the schemas in the sources. In order to perform the reformulation step, the data integration system requires a set of *source descriptions*. A description of an information source specifies the contents of the source (e.g contains gene sequence), the attributes that can be found in the source (e.g Drosophila, species), constraints on the contents of the source (e.g contains only Drosophila genome), completeness and reliablity of the source (*e.g we can assert that DBLP site has the com-*

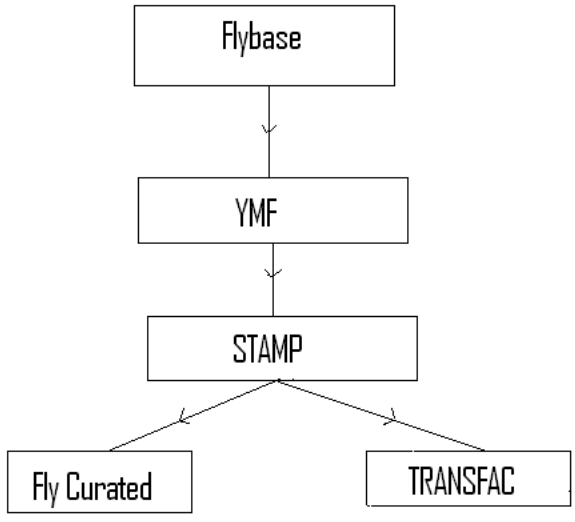


Figure 1: Data Integration for Gene Regulatory Network

plete set of papers published in most major database conferences), and finally, the query processing capabilities of the source (e.g can perform selections, search by submiting a gene name, FBgn number or CG number)

The workflow of Fig. 1 illustrates the need to support both data source integration (*e.g Fly curated and TRANSFAC*) as well as service reconciliation to enable the output of Flybase to form the input to YMF. Our approach deals with both issues in a single framework. It uses our toolkit to reconcile a sequence of semantically compatible services that need to form a pipeline: there is no need for a single "global schema".

3.3.1 Vertical Integration

We refer interested reader to Wikipedia [9] to make understand vertical data integartion. *We can think of the web data integration with Flybase to that of YMF as vertical data integration as Flybase takes some gene name say sin3, produces a sequence 800˜ 1500 bp, then those sequences can be fed into YMF to find motif of 6˜ 8 bp.*

3.3.2 Horizontal Integration

We refer interested reader to Wikipedia [9] to make understand horizontal data integartion. *We can think of several ouptut in STAMP site for a given motif to find a number of Motif Similarity Matches (e.g one chosen by Fly curated by Bergman, another chosen by TRANSFAC as horizontal data integartion).* Two main approaches for Mediator-based Integration are: Global-as-view (GAV), also known as Query-centric and Local-as-view (LAV), also known as Source-centric. In GAV approach, mediator relations are written in terms of source relations, i.e incrementally define target schema construct in terms of source schema constructs, whereas in LAV approach, source relations are defined in terms of the mediator schema, i.e define source schema constructs in terms of target schema construct.

3.3.3 Navigational Integration

There are a number of examples for link-driven integration such as: Sequence Retrieval System (SRS) and Entrez. *Integrating data from Flybase to YMF also falls in this category.* Pure navigational integration eliminates relational modeling of the data and instead applies a model where sources are defined as sets of pages with their interconnections and specific entry-points, as well as additional information such as content, path constraints, and optional or mandatory input parameters. However, the concept of navigational integration has in fact not yet established itself as a true alternative to the other, more common integration approaches. The biologist must know in advance which tool to use for each step in this process [5]. (*For our experiment with Gene Regulatory Networks, it consist of the following sites: Flybase, YMF and STAMP*)

4 WEB SITE COMPOSITION

4.1 Deep Web Query Model

Structured Web databases can be queried via query forms or through Web service interfaces. We uniformly refer to both access methods as "query interfaces". Through query interfaces, data consumers (e.g end users) are able to express their information needs by imposing selection conditions on certain attributes of interest. Our system views a Web database as a single relational table DB with a set of queriable attributes $A_q = \{attr_{q1}, attr_{q2}, \ldots, attr_{qn}\}$ (interface schema) and a set of result attributes $A_r = \{attr_{r1}, attr_{r2}, \ldots, attr_{rm}\}$ (result schema). Each $attr_{qi} \in A_q$ represents the queriable attribute through the query interface, while the result attribute $attr_{rj} \in A_r$ corresponds to the attributes displayed in the result pages. Each query operation can be modelled using SQL syntax as:
SELECT $\{attr_{r1}, attr_{r2}, \ldots, attr_{rm}\}$
FROM DB
WHERE $attr_{q1} = val_{q1}, attr_{q1} = val_{q1}, \ldots, attr_{qn} = val_{qn}$
where val_{qi} is the corresponding attribute value filled into the query form. We model the dynamic web site as $S \subseteq Q \times R$, where Q is the query interface schema and can be represented as $Q \subseteq F \times P$ and R is the result schema, can be represented as $R \subseteq L \times V$. The semantics of the above definition is as follows: we have a set of form label F and a set of corresponding parameter P. In response to the web form, we will have tabular data, which have a set of values V. The label L of the values V may or may not be in the web page. One of the major fundamental research problem is how to find L, given F, P and V.

4.2 Web Form Filling

Information transmitted to the server in a CGI request is fundamentally just a list of (name, value) pairs with appropriate URL encoding string. Thus we can characterize a form with n controls as a tuple $F = \langle U, (N_1, V_1), (N_2, V_2), ..., (N_n, V_n)\rangle$, where U is the URL to which the encoded CGI request is sent, and the (N_i, V_i) are (name, value) pairs to be sent. There are two ways to submit a form for CGI processing. HTTP POST method submit the form in the body of the request. HTTP GET method submit the

form as part of the URL. A question mark (?) separates the base URL and action path from the encoded names and values. This query is then sent directly to the Web site of interest. It has the same effect as that of a user clicking the Go button without selecting or typing anything on the Web form. For our experiment with *Flybase* and *YMF* site form filling, the following Java code snippet captures those functionality:

```
url3String = new StringBuffer();
url3String.append
("http://wingless.cs.washington.edu/YMF/YMFWeb/YMFInput.pl?
motifsize=8&spacersize=0&jokersize=2&organismname=dmelepd&sequences=");
for(i=0; i< InthowManyGene; i++){
    url1String = new StringBuffer();
    url2String = new StringBuffer();
    url1String.append("http://flybase.bio.indiana.edu/cgi-bin/uniq.html?
    species=Dmel&cs=yes&db=fbgn&caller=genejump&context=");
    url1String.append(inputStr[i]);
    url1 = url1String.toString();
    FlyBaseToACEGenericV3 fetcher1 = new  FlyBaseToACEGenericV3(url1);
    if ( HEADER.equalsIgnoreCase(option) ) {
        pageHeader1 = fetcher1.getPageHeader();
    }
    else if ( CONTENT.equalsIgnoreCase(option) ) {
      pageContent1 = fetcher1.getPageContent();
      idIndex = pageContent1.indexOf("FlyBase ID</th><td>");
      if (idIndex != - 1){
          fbgnID[i] = pageContent1.substring(idIndex+19, idIndex+30);
          System.out.println("Extracted FBgn ID from overview page"+
          fbgnID[i]);
      }
      else
       System.out.println("Could not extract FBgn ID from overview page");
     rangeIndex = pageContent1.indexOf("Sequence location</th><td>");
     if (rangeIndex != -1){
       ourRange = pageContent1.substring(rangeIndex+26, rangeIndex+56);
       checkForwardReverse =
       pageContent1.substring(rangeIndex, rangeIndex+62);
     }
     else
        System.out.println("Could not compute the sequence range");
       computedRange = deleteComma(ourRange);
     finalRange = computedRange.substring(0, computedRange.indexOf(" "));
     url2String.append
     ("http://www.flybase.org/cgi-bin/getseq.html?id=");
     url2String.append(fbgnID[i]);
     url2String.append("&dump=DecoratedFasta&source=dmel&range=");
     url2String.append(finalRange);
     url2String.append("&addrange=");
     if (checkForwardReverse.indexOf("+") != -1){
         url2String.append("-"); //the gene is forward add Up Stream
         url2String.append(addRegion);
     }
       else {  //the gene is reverse add Up and Down Stream
          url2String.append(addRegion);
       }
       url2  = url2String.toString();
       FlyBaseToACEGenericV3 fetcher2 = new  FlyBaseToACEGenericV3(url2);
       System.out.println(">"+inputStr[i]);
       url3String.append(">");
       url3String.append(inputStr[i]);
       url3String.append("%0D%0A");

       if (checkForwardReverse.indexOf("+") != -1){
           pageContent2 = fetcher2.getPageContent();
           output1000bp = processUp(pageContent2);
       }
       else
       {
           pageContent2 = fetcher2.getPageContent();
           output1000bp = processDown(pageContent2);
       }                       //accumulate individual gene's region
          url3String.append(output1000bp);
}
else
  log("Unknown option.");
} // end for
url3 = url3String.toString();
FlyBaseToACEGenericV3 fetcher3 = new  FlyBaseToACEGenericV3(url3);
pageContent3 = fetcher3.getPageContent();
words = pageContent3.split ("\n");
for (int j=0; j < words.length; j++){
 motifIndex = words[j].indexOf("Add the top motif");
 if (motifIndex != - 1)
  System.out.println(words[j].
  substring(motifIndex, words[j].length()));
}
```

The traditional process of applying the workflow technology in a distributed environment involves three, mostly sequential phases: *first* the discovery of matching resources (*e.g Flybase and YMF*), *second* the preparation of the workflow description (*e.g auto form filling in Flybase, extract relevant gene sequence by wrapper, then submit the extracted 800 bp sequence in YMF etc*) and *third* the execution and gathering of the results (e.g extract 6-8 bp motif from YMF, submit those in STAMP and gather motif related info from STAMP such as by Fly curated and by TRANSFAC). The composite service is similar to a scientific workflow.

5 EXPERIMENTAL RESULTS

We report here a table pertaining to our experiment, on the fly data integration from Flybase and YMF site. Flyabse site has celeta gene number, CG, flybase gene number, FBgn as well as gene name. From the table, it is noted that some of the gene don't have name gene name (NULL). Each gene has sequence direction (forward or reverse) as well as its position in the chromosome arm. These two parameters are crucial for the subsequent processing. The table also shows top motif of size 6 obtained from YMF site. The last row conforms to our finding of the bicoid gene, *bin*. It should be noted that a gene might have a number of synonyms. It has been reported in the literature that on average a gene has 2.3 synonym name. We went further in geneontology site to find the synonymous name for gene. Taking care of the synonym advocates the necessity for ontology based web data integration. [1]

6 CONCLUSION AND FUTURE WORKS

Integration of web data demands significant advances in middleware. The Integartion of Web Data Sources is a large research field. This paper has demonstrated the application of a uniform approach to address the problems of heterogeneous data integration and service reconciliation. Using our approach, we are able to define workflows that utilise integrated resources and we are also able to semi-automatically reconcile workflow services. In this paper a framework

Table 2: Subset of the Genes Experimented with

CG Number	FB Number	Gene Name	F/R	Range	Top Motif (size 6)
CG12055	FBgn0001091	Gapdh1	R	2R:3679403..3680885	ATCSAT
CG6871	FBgn0000261	Catalase	F	3L:18815706..18821294	TAGCCS
CG17903	FBgn0000409	Cytochrome c proximal	F	2L:16719874..16722659	GGTCAC
CG7070	FBgn0003178	PyK (Pyruvate kinase)	F	3R:18193234..18198464	AAGAAG
CG4581	FBgn0025352	Thiolase	R	2R:19754124..19756139	ACTGTA
CG7176	FBgn0001248	Idh (Isocitrate dehydrogenase)	R	3L:8349503..8354628	CCATCC
CG10120	FBgn0002719	Men (Malic enzyme)	R	3R:8538819..8548267	CAACTC
CG3476	FBgn0031881	NULL	F	2L:7027594..7028959	AGAAGA
CG2107	FBgn0035383	NULL	F	3L:2981023..2983551	TCCGCA
CG12653	FBgn0000233	btd (buttonhead)	F	X:9588221..9591606	AGAGAG
CG15321	FBgn0030150	NULL	R	X:9576998..9577705	CAGCAG
CG15319	FBgn0015624	nej (nejire)	R	X:9563240..9576909	GGAGSA
CG8316	FBgn0030852	NULL	R	X:17523818..17526053	GCTTAC

for querying and integrating data from heterogeneous biological data sources was outline. The framework will use the mediator-based approach with intention to solve the problems exist in the current integration approaches. Much scope for future work remains. Some interesting outstanding problems include:

- Treat traditional legacy dynamic web site as web service, as Web form takes some user input and produces result, may be stuctured/ unstructured
- Devise algorithm for Web site composition, whether two sites are composible or not
- Develop and implement a system which takes SQL like user query, map the query to *Interface schema* and return output from *Result schema*

We intend to address these and other questions in the future.

Acknowledgement

The work is partially supported by Wayne State University, Computer Science Department conference, travel fund.

References

[1] Emdad Ahmed. Use of ontologies in software engineering. In *17th International Conference on Software Engineering and Data Engineering (SEDE)*, pages 145–150. ISCA, Los Angeles, USA, June 30 - July 2 2008.

[2] Emdad Ahmed, K. M. Ibrahim Asif, and Miftahur Rahman. Performce analysis of data intensive web application (diwa) - a case study. In *7th International Conference on Computer and Information Technology*, pages 226–231. ICCIT, December 2004.

[3] Emdad Ahmed and Hasan M. Jamil. A survey on bioinformatics data and service integration using ontology and declarative workflow query language. In *PhD Qualifying Survey Report*, pages 1–60. Department of Computer Science, Wayne State University, March 2007.

[4] Emdad Ahmed, Hasan M. Jamil, and Lori Pile. Resource capability discovery and description management system for bioinformatics data and service integration - an experiment with gene regulatory networks. In *PhD Prospectus Technical Report*, pages 1–70. Department of Computer Science, Wayne State University, November 2008.

[5] David Buttler, matthew Coleman, Terence Critchlow, Renato Fileto, Wei Han, Ling Liu, Calton Pu, Daniel Rocco, and Li Xiong. Querying multiple bioinformatics information sources: Can semantic web research help. In *lookup*, pages 1–10. ACM, June 2002.

[6] Liangyou Chen and Hasan M. Jamil. *AD HOC INTEGRATION AND QUERYING OF HETEROGENEOUS ONLINE DISTRIBUTED DATABASES*. PhD thesis, Mississipi State University, 2004.

[7] Bilal El-Hajj-Diab and Hasan M. Jamil. Autoflow: A web-based declarative scientific workflow language. In *2nd International Workshop on Scientific Workflow*, pages 1–8. IEEE, July 8 2008.

[8] Avigdor Gal, Giovanni Modica, and Hasan M. Jamil. Ontobulider: Fully automatic extraction and consolidation of ontologies from web sources. In *20th International Conference on Data Engineering*, pages 1–10. IEEE, November 2004.

[9] http://wikipedia.org. Wikipedia: The free encyclopedia.

[10] Chen L and Hasan M. Jamil. Supporting remote user defined functions in heterogeneous biological databases. In *International Symposium on Bio-Informatics and Biomedical Engineering*, pages 144–152. IEEE, November 2001.

[11] Chen L and Hasan M. Jamil. On using remote user defined functions as wrappers for biological database interoperability. *International Journal of Coopeartive Information Systems*, 12(2):161–195, June 2003.

[12] Chen L, Hasan M. Jamil, and Wang N. Automatic wrapper generation for semi structured biological data based on table structured identification. In *DEXA International Workshop on Biological Data Management*, pages 1–10. IEEE, September 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A New Digital Signal Processor for Doppler Radar Cardiopulmonary Monitoring System

Mohammad Shaifur Rahman, Byung-Jun Jang, and Ki-Doo Kim

School of Electronic Engineering, Kookmin University
861-1 Jeongneung-dong, Seongbuk-gu, Seoul 136-702, Korea
E-mail: msrahman@kookmin.ac.kr

Abstract - Remote sensing and monitoring of cardio-pulmonary activities based on direct conversion Doppler radar shows promise in medical and security applications. For accurate sensing, demodulation of the quadrature outputs of a direct-conversion Doppler radar is a great challenge. A digital signal processor based on Kalman filtering and principal component combining of quadrature channels is suggested. Rate detection ability and success ratio is evaluated and compared with other techniques. This radar sensor system achieves good detection accuracy in increased noise power level.

I. Introduction

Microwave Doppler radar with quadrature direct-conversion is a promising method for noncontact detection and monitoring of human cardiopulmonary activities. This monitoring system can be efficient for regular health care, emergency, military, security as well as in the case of neonates, infants or burn victims where contact sensors are not suitable. Microwave Doppler radar has been used for physiologic sensing since the early 1970s [1]. This system included bulky, heavy and expensive waveguide components which were not practicable everywhere. However, integration of Doppler radar transceivers on a single chip is now achievable due to the recent advancements of micro-fabrication and wireless technology. Further, robust digital signal processors have opened up enormous possibilities for processing and extracting the information from noisy data. Several research groups are working on remote sensing and monitoring of cardiopulmonary activities using direct conversion Doppler radar [2-5]. Demodulation of the noisy quadrature outputs of a direct-conversion Doppler radar is a great challenge for accurate monitoring. Park et al. [4] suggested a demodulation technique which is found to be effective in solving the problem of demodulation sensitivity to target position. They combined the quadrature outputs using arctangent demodulation with DC offset compensation. The arctangent technique takes advantage of the 90° phase difference between I and Q channels and combines the signals by taking the arctangent of the ratio of Q signal to I signal. Successful arctangent demodulation of quadrature channels depends on correction of channel imbalances and removal of unwanted DC offsets resulting from receiver imperfections and clutter reflections while preserving the required DC information. Channel imbalance can be corrected using Gram-Schmidt procedure. However, accurate prior knowledge of the amplitude and phase errors is required. Further, removal of unwanted DC offsets from the quadrature signal is also difficult.

Our present study suggests a new digital signal processor for Doppler radar monitoring of vital signs. Kalman filter is used for the estimation of quadrature I/Q signals containing the information related to heart and respiration rate. Kalman filter [6] is an estimator for linear quadratic problem of a linear dynamic system perturbed by white noise. The estimation is done by using the measurements linearly related to the state but corrupted by white noise. After the estimation, principal component combining (PCC) is applied to combine the quadrature channels. PCC is a technique that reduces data dimensionality by performing a covariance analysis between factors, suppressing redundant information and maximizing the variance in the data.

II. Quadrature Doppler Radar Transceiver

According to Doppler theory, if a radio signal is reflected from a target with a time-varying position but with zero net velocity, it will be phase modulated. In that case, the modulation is proportional to the time-varying target position. If the change in target position is small compared to the wavelength of the radio signal, the phase change will be small and the PM signal can be directly demodulated by mixing it with the original signal. Human chest has a periodic motion for heart beat and respiration with a net zero velocity, and therefore, a Doppler radar with the chest as a target will receive a signal similar to the transmitted signal with its phase modulated by the time-varying chest position. Figure 1 shows the block diagram of a Doppler Radar cardiopulmonary monitoring system. Typically this transceiver transmits a radio wave and receives a phase modulated signal reflected from the target. The LO and RF output signals are generated from the same source. A 90° power splitter is used to divide the LO signal. These two LO signals are mixed with the reflected RF signal to provide two orthonormal baseband

This study was supported by Mid and Long Term Strategic Technology Development Program, Ministry of Knowledge Economy, Republic of Korea (10030045)

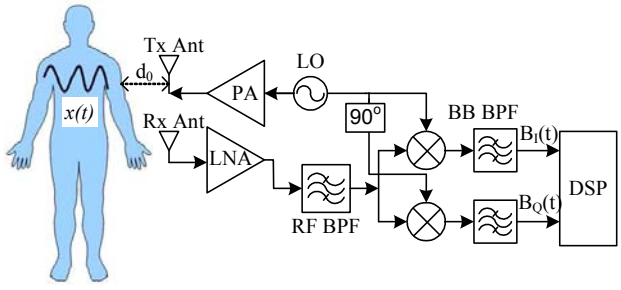


Fig. 1 Doppler radar cardiopulmonary monitoring system block diagram. The LO and RF output signals are provided from the same source. The LO signal is split by a two-way 90° power splitter to obtain in- and quadrature phase signals. These two signals are mixed with the reflected RF signal and bandpass filtered to get I and Q baseband signals.

output signals. Use of quadrature receiver eliminates the problem of null points in single channel receivers [7]. For 2.4 GHz radar operating frequency, the null points occur at every 3 cm which is difficult to avoid by adjusting the target position.

Let $f_H(t)$ and $f_R(t)$ be the heart beat and respiration frequencies of the target, respectively. After mixing of the reflected and LO signals and lowpass filtering, the baseband I and Q signals can be expressed as follows, respectively [3]:

$$B_I(t) = A\cos(\theta + \pi/4 + 4\pi A_R \sin 2\pi f_R(t)/\lambda + 4\pi A_H \sin 2\pi f_H(t)/\lambda + \Delta\varphi(t)) \quad (1)$$

$$B_Q(t) = A\cos(\theta - \pi/4 + 4\pi A_R \sin 2\pi f_R(t)/\lambda + 4\pi A_H \sin 2\pi f_H(t)/\lambda + \Delta\varphi(t)) \quad (2)$$

Here, θ is the constant phase shift related to the nominal distance between the antenna and the target, the phase change at the target surface, and the phase delay between the mixer and the antenna. $\Delta\varphi(t)$ is the residual phase noise. When $\theta + \pi/4$ is an integer multiple of π, I channel signal will be in null point [7]. At the same time, the Q channel signal will be in optimum point. On the other hand, the condition will be reversed if $\theta - \pi/4$ is odd multiple of $\pi/2$. At a frequency of 2.4 GHz, I or Q channel null points occur at every 3 cm. By using the quadrature receiver, it can be assured that at least one output won't be in null point. If θ becomes integer multiple of π, both I and Q channel will neither be in null nor in optimum points. However, still the heart rate can be detected provided that target displacement due to heart beat is smaller than the wavelength of the radio signal.

In order to demodulate the heart and respiration signals from the baseband signals regardless the target position, I and Q signals can be combined using arctangent demodulation. However, the divided-by-zero problem occurs in arctangent calculation when the denominator value is zero. In addition, quadrature channel imbalance and DC offset issues add to system complexity. PCC can be used to overcome these problems. However, it shows poor success ratio at the starting time. We suggest Kalman filtering technique combined with PCC which shows good success ratio from the starting and also gives better performance than the other two techniques in terms of SNR.

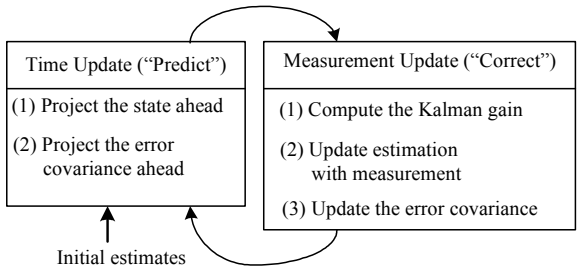


Fig. 2 Discrete Kalman filtering cycle. The time update projects the current state estimate ahead in time. The measurement update adjusts the projected estimate by an actual measurement at that time.

III. Kalman Filtering

The Kalman filter estimates a process by using a form of feedback control. The filter estimates the process state at some time and then obtains feedback in the form of noisy measurements. The equations for the Kalman filter fall into two groups: time update equations and measurement update equations. The time update equations are responsible for projecting forward the current state and error covariance estimates to obtain the *a priori* estimates for the next time step. The measurement update equations are responsible for the feedback i.e. for incorporating a new measurement into the *a priori* estimate to obtain an improved *a posteriori* estimate. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. The steps for Kalman filter estimation are show in Fig. 2. The details of the Kalman filter can be found in [8]

IV. Signal Processing Methodology

Figure 3 shows the block diagram of the proposed digital signal processor. The analog bandpass filtered signals are digitized at 50 Hz sampling rate. The quadrature I/Q signals containing noise are state estimated by using Kalman filter. The Kalman estimated I/Q signals are combined using the PCC. The signal is then filtered to separate the heart and respiration components. 5th order bandpass digital elliptic filter with 1 dB of peak-to-peak ripple and a minimum stopband attenuation of 60 dB is used to extract the heart component and 5th order lowpass digital elliptic filter with 0.5 dB of peak-to-peak ripple and

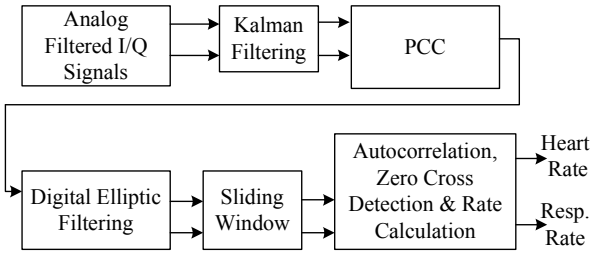


Fig. 3 Digital signal processor block diagram. Processing blocks are simulated using MATLAB.

a minimum stopband attenuation of 20 dB is used to extract the respiration component. The cutoff frequency selected for heart and respiration components are 1 to 3 Hz and 0.3 Hz, respectively. Calculations are done on 10-second data using a sliding window. The sliding window is shifted over the samples at one sample increment. For the determination of the heart and respiration rate, autocorrelation function is used. One of the properties of autocorrelation function is that if the input signal contains a periodic component, the autocorrelation function will yield a periodic component of the same frequency. The period of the signals is determined by detecting the zero crossings since the DC offset is removed from the signal. Success ratio is calculated for the simulated signal. Success ratio is defined as the percentage of time the calculated rate is within 2% of the reference rate.

V. Results and Discussion

Raw I and Q signals are generated according to Equations (1) and (2). Amplitude ratio of the heart and respiration signals are set to be 1:100, since the amplitude of the respiration signal is typically about 100 times greater than that of the heart beat signal. For that reason the respiration rate can be detected without signal processing. Hence, our analysis is mainly concerned with heart rate monitoring. For the simulation, heart and respiration rates are taken to be 82 beats and 21 breaths per minute, respectively. White Gaussian noise is added with noise power level of -75 dBW. The signal processing is done using MATLAB software. Some representative signals at different steps of the digital signal processor are displayed in Fig. 4. In this figure, the raw I and Q signals, Kalman estimation of I and Q signals, heart and respiration signals extracted from the PCC signal, and the reference heart signal are shown. Here, θ is set to a value so that I channel is closer to the optimum point and the Q channel is closer to the null point. The simulation is performed for 20 seconds. For the rate calculation, a 10-second sliding window with 500 samples is used. The sliding window is shifted over the samples at one sample increment. Heart rate for the proposed Kalman filtering technique is calculated by detecting the zero crossings of the autocorrelated heart signal and the result is compared with two other techniques. One is arctangent combining of I and Q signals and the other is PCC of quadrature channels. The result is shown in Fig. 5. Heart rate measurement history evaluated by the proposed technique shows a good conformity with the reference rate of 82 beats per minute. The other two techniques show some errors at the beginning. It is found in Fig. 5 that the beat rate is initially erroneous for about 9 seconds in PCC and for about 2 seconds for arctangent demodulation technique. On the other hand, proposed Kalman filtering technique gives almost accurate result from the very beginning.

Average success ratio for last 10 seconds of the 20-second simulation is calculated for the three different techniques by varying the value of θ in I and Q signals from 0 to 2π radians. The result is shown in Fig. 6. Solid, dotted and dashed lines show the average success ratios for Kalman filtering with PCC, only PCC and arctangent combining techniques, respectively. In case of arctangent

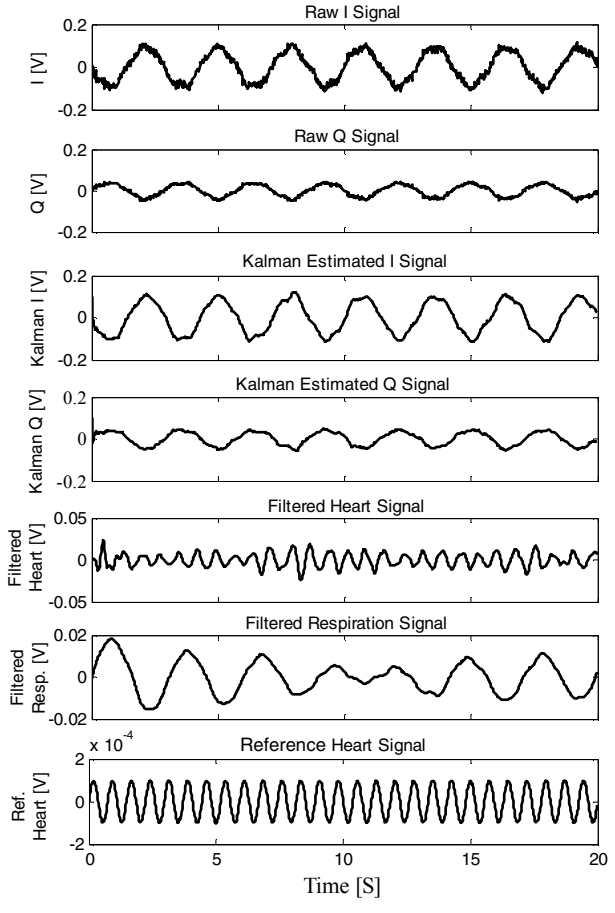


Fig. 4 Signals at different levels of the rate measurement. Doppler radar raw I and Q signals, Kalman estimated (filtered) I and Q signals, heart and respiration components extracted by digital elliptic filtering from the PCC signal and the reference heart signal are shown. Here, I channel is near to optimum point while the Q channel is near to null point.

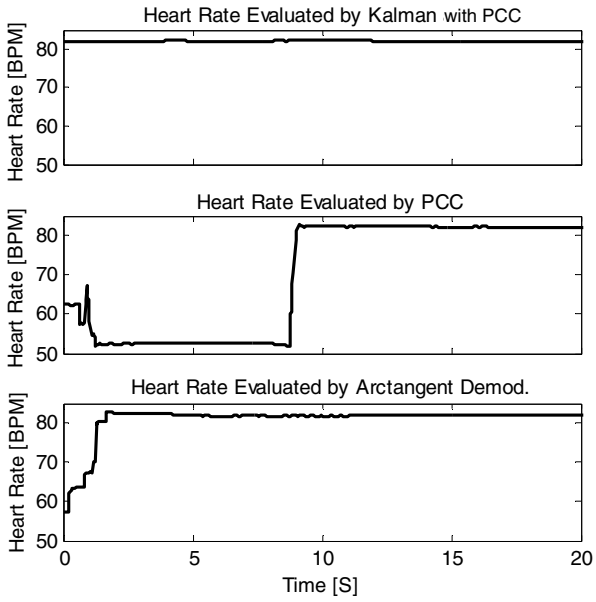


Fig. 5 Heart rate measurement history for 20 second. Heart rate evaluated by Kalman filtering with PCC technique shows a good agreement with the reference rate of 82 beats per minute from the beginning. The other two evaluations are done by PCC and arctangent demodulation, which show some errors at the beginning.

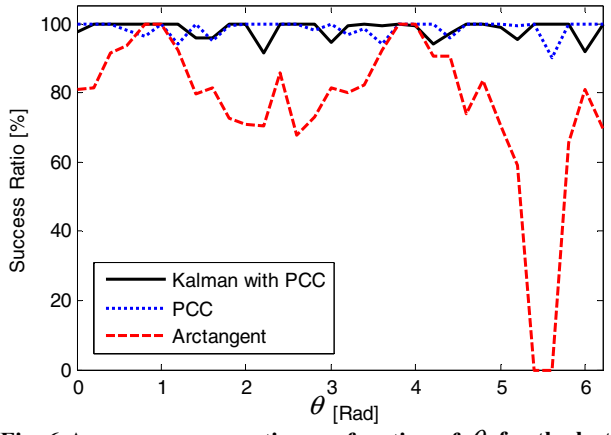


Fig. 6 Average success ratio as a function of θ for the last 10 seconds of the 20-second simulation. θ depends largely on the distance between the antenna and target. Solid line shows result for the Kalman filtering with PCC where average success ratio is above 90% for all cases. PCC also shows good success rate even though its performance is not satisfactory at the initial seconds. Arctangent has the limitations of divided-by-zero.

combining technique, average success ratio falls in some situations due to divided-by-zero problem since the I and Q channels are combined by taking the arctangent of the ratio of Q signal to I signal. Other two techniques show θ independent performance. However, proposed Kalman filtering with PCC technique gives an additional benefit of better detection ability from the beginning.

Performance of the proposed technique is also evaluated increasing the noise power level and the result is compared with the two other techniques. The result is shown in Fig. 7. It is seen in Fig. 7 that the success ratio of Kalman with PCC technique is above 95% at -70dBW noise power but for the other two techniques it is below 85% for the same noise power. Hence, performance of the proposed technique would be better than the other two techniques even if the noise power level is higher.

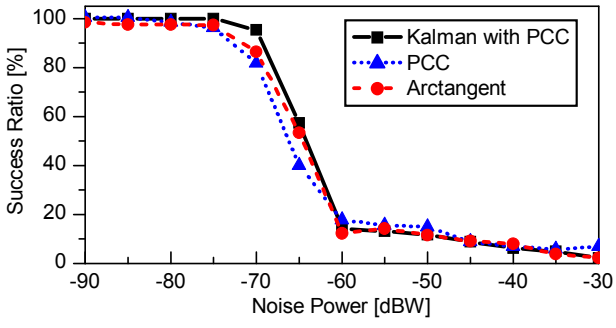


Fig. 7 Average success ratio as a function of noise power in dBW for the last 10 seconds of the 20-second simulation. Solid, dotted and dashed lines show Kalman filtering with PCC, PCC, and arctangent demodulation success ratios, respectively, as noise power level is varied. This figure demonstrates some better performance of combined Kalman and PCC technique in case of increasing the noise power.

VI. Conclusion

In the present study, a digital signal processor has been described which extracts the respiration and heart rate information from I and Q quadrature signals of the direct conversion Doppler radar physiologic monitoring system. Kalman filtering is used to estimate the I and Q quadrature signals from the noisy baseband signals. The estimated signals are then combined using PCC. Autocorrelation operation is performed to produce the periodic signal and heart rate is evaluated by zero crossing detection. Results obtained from the simulation show that the proposed technique can be successfully applied for remote monitoring of vital signs. This technique also shows better performance in noisy environment. Hence, Kalman filtering technique can be a good candidate for digital signal processing of Doppler radar remote monitoring system.

References

[1] J. C. Lin, "Microwave sensing of physiological movement and volume change: A review," *Bioelectromagnetics*, vol. 13, no. 6, pp. 557-565, 1992.

[2] B. Lohman, O. Boric-Lubecke, V. M. Lubecke, P. W. Ong, and M. M. Sondhi "A digital signal processor for doppler radar sensing of vital signs," *IEEE Eng. in Medicine and Biology*, pp. 161-164, September/October 2002.

[3] A. D. Droitcour, O. Boric-Lubecke, V. M. Lubecke, J. Lin, and G. T. A. Kovacs, "Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for noncontact cardiopulmonary monitoring," *IEEE Trans. Microwave Theory Tech*, vol. 52, no. 3, pp. 838-847, March 2004.

[4] B. K. Park, O. Boric-Lubecke, and V. M. Lubecke, "Arctangent demodulation with DC offset compensation in quadrature Doppler radar receiver systems," *IEEE Trans. Microwave Theory Tech*, vol. 55, no. 5, pp. 1073-1079, May 2007.

[5] B. K. Park, V. M. Lubeck, O. Boric-Lubecke, and A. Host-Madsen, "Center tracking quadrature demodulation for a Doppler radar motion detector," *in IEEE MTT-S International Microwave Symp.*, pp. 1323-1326, June 2007.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, vol. 82 (series D), pp. 35-45, 1960.

[7] J. Seals, S. R. Crowgey, and S. M. Sharpe, "Electromagnetic vital signs monitor," Georgia Tech. Res. Inst., Atlanta, GA, *Final Rep. Project A-3529-060*, 1986.

[8] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," http://www.cs.unc.edu/~tracker/media/pdf/SIGGRAP-H2001_CoursePack_08.pdf

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Design of a Cost-effective EMG Driven Bionic Leg

T. Latif [1], C. M. Ellahi [1], T. A. Choudhury [1] and K. S. Rabbani [2]

1 Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT), Gazipur
2 Department of Physics, University of Dhaka, Dhaka

Room 302, Academic Building, Department of EEE, Islamic University of Technology (IUT), Board Bazar, Gazipur - 1704

tahmidlatif@yahoo.com, mahboob.ellahi@gmail.com, tanveerchou_25@hotmail.com, srabbani@agni.com

Abstract – Conventional low cost leg prosthesis, available in Third World countries, is essentially a fixed passive structure which makes walking possible with some difficulty, and climbing stairs with such is extremely difficult. The present work was taken to design prosthesis for above-knee amputees. The low cost bionic leg prosthesis, which has an active (battery powered) limited rotational movement of the knee joint, is controlled by voluntary EMG (electromyogram) signals from two opposing muscles from the thigh; one to rotate the leg backward (flexion) and the other forward (extension). The design involved developing the necessary EMG and processing circuitry, interfacing the output to the microcontroller, developing the driving circuitry for bidirectional rotation of the motor, and programming the microcontroller.

During the course of the present work it was possible to control the rotation of a motor using a simulated EMG signal. The completed designed prosthesis will allow a user not only to walk with a better gait, but also to climb stairs with ease.

Keywords EMG-based control, Prosthesis, Bionic leg,

1 Introduction

The human leg is a complex and flawless part of the body. Consisting of a number of complex functions and providing multiple degrees of freedom, the human leg is almost impossible to imitate by any means of prostheses. Commercially available, locally made simple leg prosthesis has zero degree of freedom or in some cases just one. The latter provides a passive knee joint, which cannot be utilised to its best due to *lack of voluntary involvement*. These do *not* prove to be *user-friendly* and the weight turns out to be a major problem. While walking, the user has to pull the leg along and climb stairs in a similar way; control requires a lot of energy that easily *fatigues the user*.

In the Third World countries and in this sub-continent, people solely focus on the use of passive prosthesis. Large-scale developments in prosthesis have been made in India by Jaipur Foot [1][2][3] with rubber-based prosthetic leg starting from 1968-69 [1]. These prostheses are inexpensive and quick to manufacture, but being passive, occasionally lack the option of wilful control of the prosthesis. Developments in the west have produced very sophisticated 'bionic' legs with many advanced features. These bionic legs are not the typical mechanical structures, but are prostheses voluntarily controlled by the user by means of electrical circuitry. The bionic leg developed by Victhom [4] (commercialised by Ossur [5] under the name POWER KNEE), with an on-board AI module, is a fantastic example of the latest innovations in prostheses. Ossur is involved in development of both below-knee and above-knee bionic prosthesis. Work has been done by researchers at Arizona State University's Polytechnic campus and the Military Amputee Research Program at Walter Reed Army Medical Center to develop powered prosthesis nicknamed SPARKy (Spring Ankle with Regenerative Kinetics) [6][7]. The device is a below-knee prosthesis based on light-weight energy storing springs. However, these types will remain out of bounds for most of the Third World because of 'cost'. The goal of this work is to provide above-knee amputees with improved leg prostheses at 'low cost'.

The prosthetic leg is designed with a motor operated active knee joint (hence the term bionic leg), being operated by voluntary control of the user through electrical signals (EMG) from thigh muscles. In short, the bionic prosthesis will:

- provide voluntary movement of motorised knee joint
- improve gait (manner of moving)
- allow stair climbing with ease
- eliminate undesired movement at knee

The designed bionic leg focuses on cost-effectiveness and quality as well as comfort and ease of use.

2 Use of Electromyogram

Whenever a physically fit person is walking or an amputee is walking, using a passive prosthesis or crutch, their muscles generate signals due to movement of his/her thighs that corresponds to movement during walking.

These signals are EMG (Figure 2.1), which when observed using suitable means would give pictorial representation of electrical activity of the contracting muscle. [8]

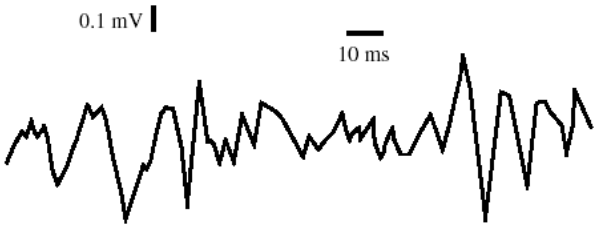


Figure 2.1 Typical EMG Signal [9]

Voluntary EMG results from voluntary contraction of muscles, under wilful action of the brain. Voluntary EMG from thigh muscles is used for rotation of motor at knee. While identifying a site for EMG signal extraction, it is important to give careful consideration to the skin condition and acceptability of EMG signal strength of that site [10].

EMG signal, acquired from thigh muscles (using differential electrodes), is processed to control a dc servo motor operating the knee joint, using a microcontroller. The microcontroller, used to sense the presence of this output, is programmed to decide whether to activate the motor or not, and if yes, in which direction to rotate it. EMG signals from the amputee's thigh will rotate the motor operated knee-joint.

To make things simpler and to use a non-invasive technique, the signal is sensed using surface electrodes, placed over the desired muscle. Use of needle electrodes will make the use of the prosthesis, with needles pierced into the skin, a painful experience, which is not a practical proposition.

3 The Designed System

3.1 Signal Acquisition and Amplification

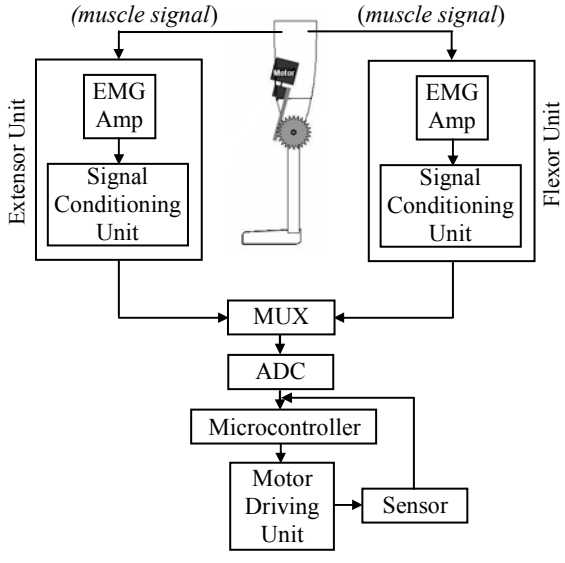


Figure 3.1 Block Diagram of Designed System

As shown in the block diagram of the designed system (Figure 3.1), signals from thigh muscles, extracted using skin surface electrodes, are amplified by EMG Amplifier units followed by Signal Conditioning units. There are two sets of EMG Amplifier and Signal Conditioning units. One set processes muscle signal for extension at the knee and the other set for flexion. Since the amplitude of the muscle signals are very small (ranges from less than 1 μV to 10 mV [11]), their amplification is very much essential for effective usage in our system. An instrumentation amplifier is used for this purpose, keeping in mind its good de-noising property. With the chosen resistance values, the amplifier has

Theoretical value of differential gain = 100.909

Obtained value of differential gain = 94.118

The common-mode rejection ratio (CMRR) for the amplifier,

$$\text{CMRR} = \frac{|A_d|}{|A_{cm}|} \approx 90 \text{ dB} \qquad \ldots\text{(i)}$$

where A_d = Differential gain

A_{cm} = Common-mode gain

The higher the CMRR value, the better it is. Such a CMRR value would indicate that the amplifier amplifies the wanted signal and discriminates against the common-mode signal, that is, the noise. For best results, the operational amplifiers used in the amplifier have to be closely matched and so has to be taken from the same IC package [12]. The instrumentation amplifier is constructed using three operational amplifiers (op-amps) of TL074 [13] quad operational amplifier IC package. TL074 ICs are used in constructing the instrumentation amplifier as well as other circuits where operational amplifiers are needed.

The amplified signal was then fed to the Signal Conditioning Unit (Figure 3.2), which comprises a band pass filter, an active rectifier and a peak detector.

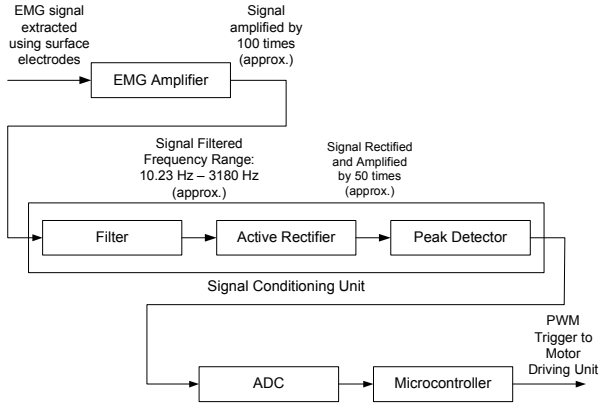


Figure 3.2 Block Diagram of the Designed System, partially redrawn, showing the Signal Conditioning Unit

The effective frequency range of EMG signals (extracted using surface electrodes) is between 10 Hz and 5 kHz [14]. The wideband band pass filter in the design is made by cascading a 10.23 Hz active 40-dB/decade high pass Butterworth filter and a 3.18 kHz passive low pass filter. Using a second-order high pass filter gives a much required sharper lower cut-off. A passive low pass filter is

used since surface electrode EMG signal at high frequencies contains less useful information than that from needle electrodes [11]. The selection of the chosen cut-off frequencies also ensure that high frequency components, which might obstruct smooth motor rotation, are eliminated. Experimentally, the lower and higher cut-off frequencies are found to be somewhat less than 10 Hz and around 4 kHz respectively.

The output signal of the instrumentation amplifier being bidirectional, will not allow smooth movement of the motor due to signal polarity changes. So, the signal needed to be rectified. A simple diode rectifier could not be used because of the forward voltage drop. So an active rectifier, using op-amps, was used in the design to overcome the diode voltage drop. Not only did the active rectifier rectify the signal, but it was designed in such a manner that it amplified the signal for use in the later stages.

The variable dc output, obtained through rectification, being pulsating in nature would have made it unsuitable for driving the motor of the prosthesis. The peaks of the output were detected using a simple active peak detector. Generally, it is assumed that,

$$RC >> T \qquad ...(ii)$$

where T is the period of the input

It must not be forgotten here that the input to the peak detector, the rectified EMG signal, is an aperiodic signal consisting of different frequency ranges. So, a generalised assumption would fail here. The RC time constant should be large in order to give a smooth control voltage for driving the motors. On the other hand, it should be small enough to follow changes in EMG amplitudes, particularly for reductions, when the user desires. Therefore, the RC time constant needs some optimisation.

Word of caution: If the peak detector circuit fails to give output at all the peaks of the signal, then improper signal may be send to subsequent stages. Currently the peak detector circuit is under analysis with different RC values.

3.2 Microcontroller Operation

The use of microcontroller in the design makes it and the control process sophisticated. Researchers have been implementing microcontrollers in prosthetic knees to enhance the basic mechanical design [15][16].

ADC0808 [17], used in the design, is a fast switching 8 bit ADC which can multiplex between eight analogue signal inputs. It has a conversion time of 100 µs only. A fast conversion time is very prominent for our purpose since we are dealing with smooth knee movement. Only two input channels of the ADC are used; envelop signal outputs from the peak detectors of both the signal conditioning units (Flexor and Extensor) were connected to two of these channels. The 8-bit microcontroller PIC16F84A [18] has a stored programme with the help of which it selects one channel of the ADC at a time and triggers the ADC to produce the 8 bit digital output. The digital output is sent to the microcontroller via an 8 x 1 Multiplexer 74153 [19] (Figure 3.3). The stored programme is written in a manner such that the microcontroller keeps on checking and comparing the magnitudes of the digitised signals, from the two muscles, in real time. The motor rotates in the direction defined by the amplitude of signal larger among the two.

An important role as played by the microcontroller is modulating (PWM - Pulse Width Modulation) the current through the motor in accordance to the signal amplitude. PWM allows variation of the average current through the motor without changing the original peak value of the current, thus controlling speed of the motor. A change in value of 1 in the digital domain represents a change of magnitude of $(5\text{ V} - 0\text{ V})/2^8 = 19.53$ mV of the analogue signal. The microcontroller generates PWM signal by considering a time period of 255 units, which was divided into two frames, namely on-time and off-time. On-time equals the digital value of the signal amplitude. For higher the signal amplitude, more was the on-time in the time slot and hence lesser the off-time. Together, many such time slots make up a PWM signal to control the average current flowing through the motor.

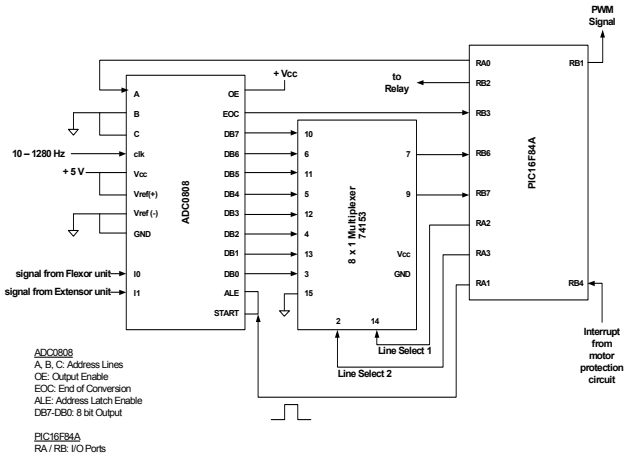


Figure 3.3 Data Acquisition: Data Transfer between ADC0808 and Microcontroller PIC16F84A

This PWM signal from the microcontroller acts as a gate signal to a transistor connected in series with the motor.

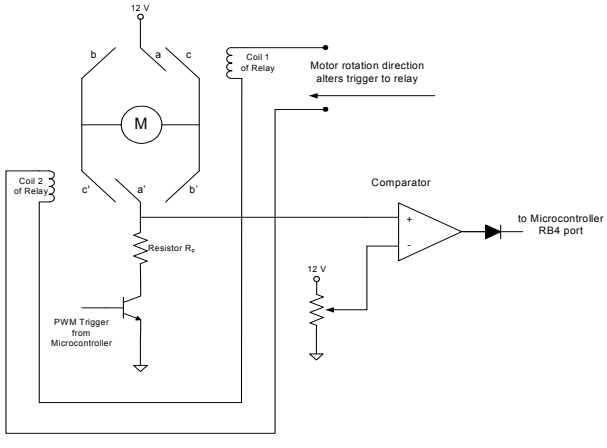


Figure 3.4 Motor Driving and Protection Unit

The motor driving unit (Figure 3.4) utilises a 2-contact 5V relay for driving the motor. Flow of current in a

certain direction (say, current from Flexor unit) links contacts *a* and *a'* to *c* and *c'* respectively; the motor rotates in a certain direction. Flow of current in the opposite direction (say, current from Extensor unit) results in connection of contacts *a* and *a'* to *b* and *b'* respectively; the motor rotates in another direction. The PWM signal to the transistor causes variation of the average current through the motor, which on the other hand varies the speed of the motor proportionately.

A kneecap-like metal lock prevents bending at the knee beyond full extension. If the lower leg is fully extended and muscles are still working, the continuing EMG signal will force the knee to bend further against the mechanical lock. In such a situation, the motor is in blocked rotor condition, during which a very large current flowing through the rotor winding of the motor may damage the motor. The motor protection unit is provided as a remedy against blocked rotor condition. It is a simple op-amp comparator circuit (Figure 3.4), which compares the voltage across the resistor Rp of the motor driving circuit with a reference voltage experimentally preset by the voltage divider. When the current through the motor approaches near short circuit current value, as the potential difference across the resistor Rp just exceeds the reference voltage, the output of the comparator goes high interrupting the microcontroller. The interrupt procedure suspends motor operation, resets PWM signal to zero and after a delay of 1 second the motor rotates in a newly determined direction of rotation.

Figure 3.5 shows the flowchart for the microcontroller operation. It is important for the amplitude of the input to the ADC to be of the order of 5 V for proper functioning [17]. For this purpose, EMG signals are amplified by the EMG Amplifier and later stages to appropriate voltage level.

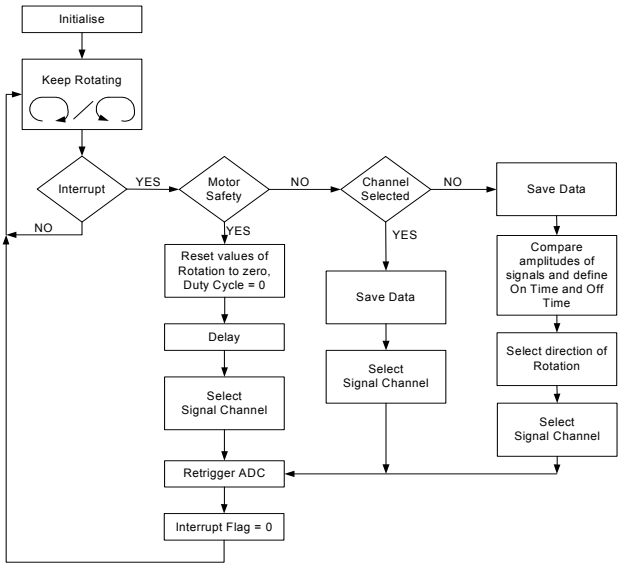


Figure 3.5 Programme Flowchart for the Microcontroller Operation

3.3 Safety and Isolation

The whole system, which includes the circuits and associated equipments in the design, are all battery powered. Since the system is not connected to the mains, there is no necessity to consider the safety of the amputee from severe electrical shock.

4 Mechanical Design of the Leg

The role of mechanical system not only is to provide with the desired motion on rotation of the motor, but also to support the whole body of the amputee.

The fitting frame of a commercially available passive prosthetic leg is used. The lower leg has to be light enough since this part has to be in constant motion when an amputee would walk. Corrugated iron is used to make a lower leg weighing around 0.825 kilograms. The centre of gravity of the lower leg is shifted closer to the knee joint so that the maximum toque required to lift it will be less. In this case, the required torque is approximately equal to 1.17 Nm.

For driving the prosthesis, a servo motor is chosen over a stepper motor because of the former having higher speed, lighter weight and providing with smooth rotation at uniform angular speed. Moreover, the driving circuits for such motors are cost-effective. A 12 V, 3 W motor, with a rated rpm of 3000, is used in the design.

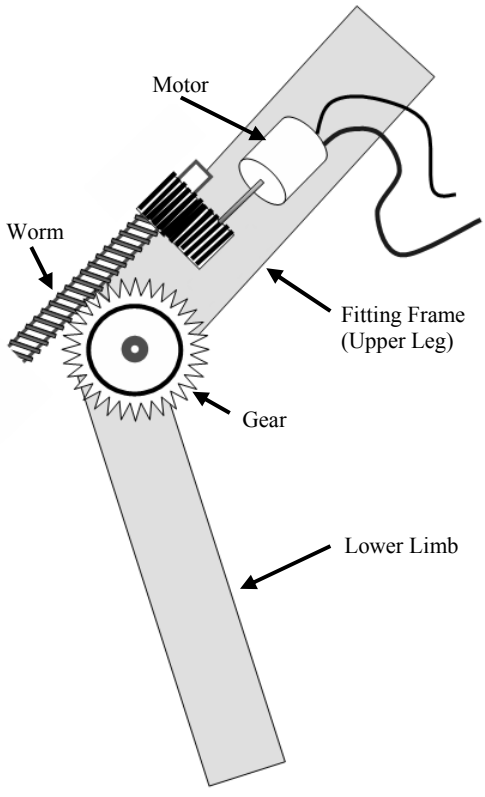


Figure 4.1 Implemented Design of the Mechanical System

The design utilises a worm-gear as part of the driving system. The worm-gear and motor, etc. need to be fitted onto the fitting frame of the prosthesis. All of the above are implemented in the design shown in Figure 4.1. As the motor rotates, it rotates the worm (screw rod) coupled to its shaft. The motion of the worm in turn rotates the gear. For extension of the lower leg, the gear would rotate in a clockwise direction while for flexion in an anticlockwise direction. The lower part of the leg, welded

to the gear, moves along with the gear. The square threading of the worm-gear prevents flexion or extension of the lower leg by any external pressure. The necessary torque required by the motor to move the lower leg is improved by carefully selecting gear ratio between the gears attached to the motor shaft and the worm. Torque and speed are optimised in the process. In this paper, the calculations carried out in the mechanical design are not shown.

5 Conclusion

Some developments have already been made in the field of bionic legs with no or little concern for Third World countries. The paper discussed a simple, but meaningful design of an above-knee bionic leg with a focus on cost-effectiveness.

At this point, the primary concern is on perfecting the design: both electrical and mechanical. This paper dealt mainly with the electrical part leaving out the details of the mechanical design. Research is being done on the mechanical design in order to see the face of perfection. The largest barrier that lies ahead is the availability of suitable material here in our region. The lower leg and the worm-gear, currently being used, are made of iron and this makes the prosthesis heavy and very slow in performance. If the leg could have been made using a light-weight, but strong material and the worm-gear with material lighter than iron, the weight of the whole prosthesis would have significantly reduced and the performance might have improved.

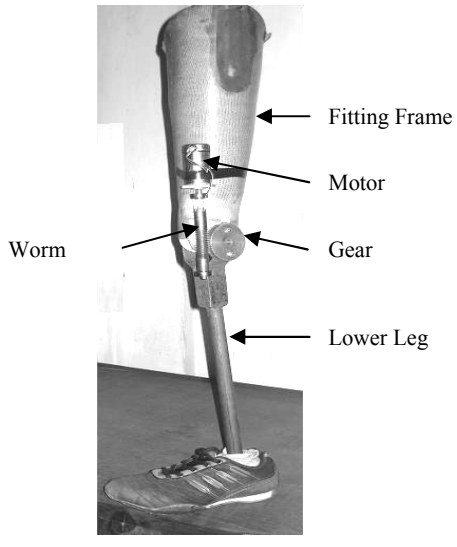


Fig 5 The Worm-gear, Motor and Lower Leg attached to the Fitting Frame

Currently, the system works with simulated signals, but needs to be tested with an amputee, who, at first, has to be trained on using the prosthesis.

References

[1] Jaipur Foot: http://www.jaipurfoot.org/

[2] T. K. Kulshreshtha, "The jaipur Below KneeProsthesis HDPE – Fabrication manual", Bhagwan Mahaveer Viklang Sahayata Samiti

[3] M. K. Mathur, "Jaipur Artificial limbs"

[4] Victhom Human Bionics. Bionic Leg – The Power Knee: http://victhom.com/en/realization-bionic-leg-the-power-knee-1.htm

[5] Ossur. Power Knee: http://ossur.com/pages/2749

[6] Scientists Create Prosthesis of the Future. Science Daily. May 2, 2007: http://www.sciencedaily.com/releases/2007/05/070501151726.htm

[7] Human machine Integration Laboratory, Arizona State University: http://robotics.eas.asu.edu/research.htm

[8] M. B. I. Reaz, M. S. Hussain, F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications", Biological Procedures Online 2006; 8(1):11-35.

[9] B. H. Brown, R. H. Smallwood, D. C. Barber, P. V. Lawford and D. R. Hose, "Medical Physics and Biomedical Engineering", Medical Science Series, IOP Publishing, Figure 16.24 An EMG recorded from surface electrodes.

[10] M. Zecca, S. Micera, M. C. Carrozza and P. Dario, "Control of Multifunctional Prosthetic Hands by Processing the Electromyographic Signal", Critical Reviews in Biomedical Engineering, 30(4-6):459-485 (2002).

[11] B. H. Brown, R. H. Smallwood, D. C. Barber, P. V. Lawford and D. R. Hose, "Medical Physics and Biomedical Engineering", Medical Science Series, IOP Publishing, 16.4.2 EMG Equipment.

[12] K. S. Rabbani, Educational Material: Neuro-Physiological Study and Diagnosis using Evoked Responses, International Conference on Medical Physics Website. http://www.icmpdubai.com/sp/K%20S%20%20Rabbani%20tutorial.pdf

[13] TL074, Texas Instruments, Datasheet: http://focus.ti.com/lit/ds/symlink/tl074.pdf

[14] B. H. Brown, R. H. Smallwood, D. C. Barber, P. V. Lawford and D. R. Hose, "Medical Physics and Biomedical Engineering", Medical Science Series, IOP Publishing, 16.4.1 Signal sizes and electrodes.

[15] State of Washington, Department of Labor and Industries, Office of the Medical Directory, Technology Assessment, "Microprocessor-Controlled Prosthetic Knees", August 16, 2002.

[16] C. W. Martin, "Otto Bock C-leg: A review of its effectiveness", WCB Evidence Based Group, Compensation and Rehabilitation Services Division, November 27, 2003.

[17] ADC0808, National Semiconductor, Datasheet: http://cache.national.com/ds/DC/ADC0808.pdf

[18] PIC16F84A, Microchip Technology Inc, Datasheet: http://ww1.microchip.com/downloads/en/DeviceDoc/35007b.pdf

[19] Multiplexer 74153, National Semiconductor, Datasheet: http://www.national.com/ds/54/54153.pdf#page=1

Appendix

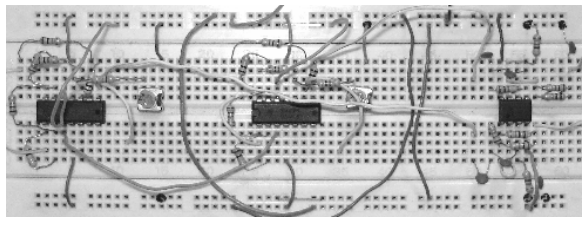


Figure I Breadboard Implementation of EMG Amplifiers and Band Pass Filters

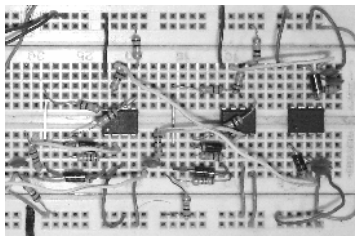


Figure II Breadboard Implementation of Active Rectifiers and Peak Detectors

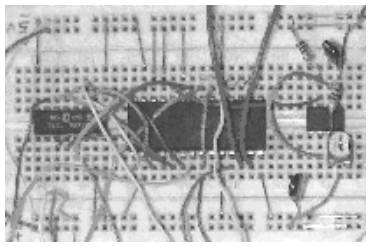


Figure III Breadboard Implementation of circuitry for ADC and Multiplexer

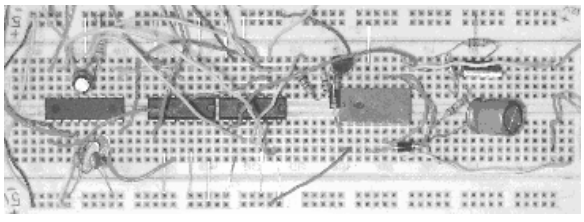


Figure IV Breadboard Implementation of circuitry for Microcontroller and Relay

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

EM Radiation from Wi-LAN Base Station and Its' Effects in Human Body

Hikma Shabani, Md. Rafiqul Islam, AHM Zahirul Alam and Hany Essam Abd El-Raouf

Department of Electrical and Computer Engineering, Kulliyyah of Engineering
International Islamic University Malaysia, P.O.BOX.10, Kuala Lumpur, 50728, Malaysia
E-mail: rafiq@iiu.edu.my

Abstract - EM Fields have adverse effects on Human health. Tissue damage could occur due to the body inability to cope with the excessive heat that could be generated during the exposure to Radio Frequency. The parameters of concern with respect to human health are the rate at which a person absorbs electromagnetic energy, called the specific absorption rate (SAR). Electric field intensities from wireless LAN were measured at IIUM Campus. Using a developed flat phantom model, electric field and SAR variations for human organs are simulated and presented in this paper. The peak SAR values are compared with recommended values.

I. Introduction

The electromagnetic field is gaining importance with rapid proliferation of large number of electrical and electronic devices which radiate unwanted EM emissions. Computer networks and the Internet are accessed using wireless techniques (Wi-LAN) at frequencies in the microwave range from 2 to 4 GHz [1].

In order to support the growing number of users, more base stations are installed in almost everywhere. Therefore, the wide spread use of wireless devices that require higher and higher amounts of bandwidth due to growing of necessary transfer rates has increased a public concern about possible health hazards resulting from exposure to electromagnetic (EM) wave.

The Finite Difference Time Domain (FDTD) method [2] is used for investigating the interaction between the human body model and EM waves. The FDTD is based on a particular Finite Difference scheme (Yee's algorithm) that is applied to Maxwell's curl equations in the time domain. It is an explicit marching-in-time procedure that simulates the propagation and interaction of electromagnetic waves in a region of space [3]. The antennas for this case are not used in close proximity to the head unlike the commercial cellular phone, but attached to the trunk. In order to incorporate the inhomogeneous human body model into the FDTD scheme, the dielectric properties of the tissues are required. These tissue parameters or tissue dielectric properties are available at Federal Communications Commission (FCC) Web Site [4]. Basically, the FDTD approach involves solving Maxwell's curl equations using a finite difference approximation to both space and time derivatives [5]. To facilitate the absorption condition on the outer surface of the boundary space, the total electric and magnetic fields $(\vec{E},\vec{H})$ are partitioned into incident $(\vec{E}^i,\vec{H}^i)$ and scattered $(\vec{E}^s,\vec{H}^s)$ fields.

A tri-axis isotropic probe with portable spectrum analyzer FSH3 and RFEX software is used to measure the incident electric field and incident power densities for all existing Wi-LAN signals at student hostels and office premises near base stations in International Islamic University Malaysia, Gombak campus. The evaluated values are compared with the safe guidelines given by the recognized body such as the International Commission on Non-Ionizing Radiation Protection (ICNIRP) [6], American National Standards Institute/Institute of Electrical and Electronic Engineers (ANSI/IEEE) [7], and Malaysia Communication and Multimedia Commission (MCMC). Generally, the ICNIRP, ANSI/IEEE, and National Council on Radiation Protection and Measurements (NCRP) Standards are the most widely accepted all over the world [8]. In order to estimate the risk caused by the use of wireless devices on different parts of the body, the following parameters have been estimated:

1. a maximum SAR or local peak SAR,
2. a whole body average SAR

To avoid adverse health effects, several reputed organizations such as the Institute of Electrical and Electronic Engineers (IEEE), Federal Communications Commission (FCC), the National Council on Radiation Protection and Measurements (NCRP) and the International Committee on Non- Ionizing Radiation Protection (ICNIRP) and the National Radiation Protection Board (NRPB) have adopted exposure guidelines for the general public as well as for RF workers in the course of their regular duties. According to ICNIRP Guidelines, the safe level of SAR for the general public is 2W/Kg and the average whole body SAR is 0.08W/kg for a person exposed to Wireless Base Station. In appendix 1, we have the SAR limits as recommended by ANSI/IEEE and ICNIRP. Furthermore, appendix 2 shows the limit of exposure level in terms of the incident

electric field strength, E (V/m), and the power density, S (mw/cm^2) as recommended by ICNIRP and adopted by some countries. In this paper, electric field intensities have been measured in three locations at IIUM campus. The measured electric fields have been used as incident field on flat phantom model [9]. The variations of electric fields and SAR have been calculated for twenty different organs in human body. The paper also focused on comparing peak SAR with recommended values.

II. Radiation measurement

The portable EMF measurement system, TS-EMF from Rohde and Schwarz is used to measure the electric field and power density from mobile base station. The equipment consists of 3 main components, which are the tri-axis probe, portable spectrum analyzer FSH3 and the RFEX software, which are installed in a laptop were used to measure electric field intensities and power densities for all existing signals ranging from 80 MHz to 2.5 GHz. The equipment has wide frequency range which from 300kHz to 3 GHz covering all common radio services as Mobile radio (GSM, CDMA and UMTS), DECT, Bluetooth™, WLAN (802.11b), Sound broadcasting and TV broadcasting [10].

The portable EMF Measurement System TS-EMF is designed for short-term and long-term measurements of the electromagnetic field (EMF). The tri-axis probe has an isotropic characteristic, so the measurement is done independent from direction or polarization of the emitter. In contrast to directional antennas, it is no longer necessary to move the antenna for covering all directions and polarization. The specification of the probe is shown is table1.

Table 1: Specification of Tri-axis Probe

Frequency range	80 MHz to 2.5 GHz
VSWR	≤ 2.0 (f > 800 MHz)
Measurement range	about 1mV/m up to 100V/m
Isotropic deviation	± 1.0 dB (900 MHz), ± 1.7 dB (1800 MHz)
Temperature range	-10° C to 50° C
Humidity	85%
Current consumption	500 mA max.

For the accurate results, the equipment (fig.1) was used for the average level of 6 min.

Fig. 1: TS-EMF Equipments

The measurement was done in three locations namely: IIUM's Library, Celcom Building, and Rector's office. The measured peak electric fields and power density are shown in table2.

Table 2: Highest measured peak values

	Freq. (MHz)	Electric Field (V/m)	Power Density (mW/cm^2)
Central Library	2467.0	37.7920	0.37884
Celcom Building	2437.0	37.2263	0.36759
Rector Office	2437.0	37.3164	0.36937

Hence, observing the recorded data in table 2, the highest measured electric field and power density are located at central library. The intensity of E-field is 37.7920V/m, which represent 61% of the limit while the power density is 0.37884mW/cm^2, which is 37.9% of the limit. These limits are based on the International Commission on Non-ionizing Radiation Protection (ICNIRP) standards for general public as presented in appendix2.

III. Biological Effects of EM Radiation

Since the introduction of mobile phones in the mid-1980s, it has been know that the exposure to electromagnetic (EM) waves can be harmful due to the ability of Radio Frequency energy to heat biological tissue rapidly [11].

It is reported that short term exposure to very high levels of RF radiation have caused cataracts in rabbits. Temporary sterility, caused by such effects as changes in sperm count and sperm mobility, is reported to be possible after exposure of the testis to high-level RF radiation. Mice and rabbits have been employed for most of the experimental investigations on biological effects of RF exposure [12]. Furthermore, an epidemiological study of cancer has discovered that the main cancers associated with EMF exposure are leukemia, nervous system tumors, lymphoma and breast cancer among children in residential settings and adults in occupational settings [13].

Other effects of EMF exposure are the depression as well as fatigue, irritation, and headaches. Epidemiological studies have found higher ratio of depression-like symptoms and higher rates of suicide among people living near transmission lines. On the other hand, adverse pregnancy outcome, including miscarriages, still birth, congenital deformities, and illness at birth have been associated with maternal occupational exposure to electromagnetic fields. Moreover, paternal occupational exposure to electromagnetic fields has also been linked to reduce fertility, lower male to female sex ratio in offspring, congenital malformations and teratogenic effects expressed in the form of childhood cancer. *In vivo* studies with rates showed that exposure to high electric fields reduced plasma testosterone concentrations and reduced sperm viability [14].

IV. Human Body Absorption and SAR Evaluation

An extremely challenging problem is the prediction of the penetration of electromagnetic fields (EMF) into a human body. The structure of the body is quite complicated, and the constitutive parameters vary with position [15].

A. EM Fields Distribution inside Human Body

During the last years, efforts have been applied to precisely predict electromagnetic behavior of outdoor microwave channels due to the explosion of the commercial use of wireless devices. The Finite Difference Time Domain Method (FDTD) is the most often used method for evaluation of EMF in human tissue. In this method, Maxwell's equations in time domain differential form are solved when an incident uniform sinusoid plane wave propagates through the body as shown in fig2.

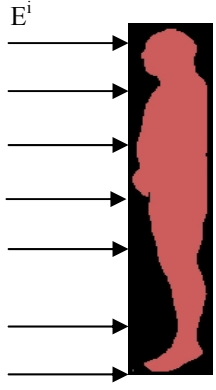


Fig. 2: Uniform incident plane wave on the body

Though, the time domain differential Maxwell's equations governing electromagnetic propagation in an isotropic medium are:

$$\nabla \times \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t} \tag{1}$$

$$\nabla \times \vec{H} = \sigma \vec{E} + \varepsilon \frac{\partial \vec{E}}{\partial t} \tag{2}$$

In which $\vec{E}$ is the electric field intensity vector in (V/m), $\vec{H}$ is the magnetic field intensity vector in (A/m), ε and μ are, respectively, electric permittivity (farads/m) and magnetic permeability (henrys/m), and σ is the electric conductivity of the tissue in S/m. The body under investigation is divided into a large number of small cubic cells of size $\Delta x, \Delta y, \Delta z$, which is the space incremental and the investigation is carried out at Δt which is time increment. Yee positions the components of E and H at alternate half time steps interval about a unit cell [16]. Using the MKS system of units, and assuming that the dielectric parameters ε, μ and σ are independent of time, the equations (1) and (2) are equivalent to the following system of scalar equations in the rectangular coordinate system (x, y, z):

$$\frac{\partial H_x}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y}\right) \tag{3.a}$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z}\right) \tag{3.b}$$

$$\frac{\partial H_z}{\partial t} = \frac{1}{\mu}\left(\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x}\right) \tag{3.c}$$

$$\frac{\partial E_x}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \sigma E_x\right) \tag{3.d}$$

$$\frac{\partial E_y}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \sigma E_y\right) \tag{3.e}$$

$$\frac{\partial E_z}{\partial t} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \sigma E_z\right) \tag{3.f}$$

In which E_x, E_y, E_z and H_x, H_y, H_z are the $\vec{E}$ and $\vec{H}$ components, respectively. The propagation of an incident electric field ($\vec{E}^i$) through the human body is governed by equation (3a) to (3f) which are solved by converting the six partial differential equations into six explicit finite difference approximation in which E_x, E_y, E_z, H_x, H_y,and H_z are obtained. To yield accurate results, the grid spacing δ in the finite difference simulation must be less than the wavelength, usually less than $\lambda/10$. When the step size δ is the same in all directions ($\delta = \Delta x = \Delta y = \Delta z$), the condition for stability is [17]:

$$\Delta t \le \frac{\delta}{\sqrt{n}\, c_0} \tag{4}$$

With n the number of spatial dimensions in the problem and $c_0 = 3 \times 10^8 \, m/\sec$ is the velocity of light in free space.

Using the leap frog integration scheme to the explicit finite difference approximation, the flat phantom model of human body is derived as follow [18].

$$E^{s,n} = \left(\frac{\varepsilon}{\varepsilon + \sigma \Delta t}\right) E^{s,n-1} - \left(\frac{\sigma \Delta t}{\varepsilon + \sigma \Delta t}\right) E^{i,n} - \left(\frac{(\varepsilon - \varepsilon_0)\Delta t}{\varepsilon + \sigma \Delta t}\right) \dot{E}^{i,n} + \left(\nabla \times H^{s,n-\frac{1}{2}}\right)\left(\frac{\Delta t}{\varepsilon + \sigma \Delta t}\right) \tag{5}$$

The total fields were obtained by adding the incident and scattered fields as follow [19]:

$$E^T(i+\tfrac{1}{2},j,k) = E_0(i+\tfrac{1}{2},j,k)\sin(2\pi f n\Delta t) + E^{s,n}(i+\tfrac{1}{2},j,k) \quad (6)$$

During each time step, the value of $E^{s,n}(i+\tfrac{1}{2},j,k)$ is calculated in the normal manner of the algorithm implemented in Matlab Codes by using the explicit finite difference approximation and stored in memory. Thus, for $n=1$, the new value of the sinusoid wave is calculated and added to the stored value of $E^{s,n}(i+\tfrac{1}{2},j,k)$. This new modified value of $E^{s,n}(i+\tfrac{1}{2},j,k)$ is stored in memory. The Eq. (6) is then incremented (n is increased by 1), and the process is repeated until all time steps have been completed. Figure3, 5, 7 and 9 show the variations of electric fields inside four organs of human body.

B. SAR Evaluation

The specific absorption rate (SAR) is used to specify the amount of radio frequency energy absorbed in the body [20]. In the frequency range of approximately 100 KHz to 10GHz, the SAR is the important dosimetric quantity. SAR is defined as the average rate of energy absorption in tissue [21], i.e.

$$SAR = \frac{d}{dt}\left(\frac{dW}{dm}\right) = \frac{d}{dt}\left(\frac{dW}{\rho dV}\right) \quad (7)$$

However, as the EM field is being stepped throughout the scatterer, a peak value detector is constantly monitoring at each point for new maximum amplitude. When steady state is reached (typically after the incident wave has been generated for approximately two to three periods of oscillation), the stored values of maximum amplitude are retained, and the local specific absorption rate (SAR) is calculated in the following manner:

$$SAR = \frac{\sigma |E|^2}{\rho} (W / kg) \quad (8)$$

Where σ the electric conductivity of the tissue in is S/m, ρ is the mass density in Kg/m^3, and E is the root-mean-square (rms) magnitude of the electric field strength in V/m inside the material. Thus, SAR is a measure of the electric field, and indirectly the magnetic field and current density at the point under study, and also a measure of the local heating rate [22].

$$\frac{dT}{dt} = \frac{SAR}{c} \ ^0C/s \quad (9)$$

Where c is the specific heat capacity of the tissue expressed in $J/kg\ ^0C$. The Eq. (8) was plotted into developed Matlab Codes. Figure4, 6, 8 and 10 show the distribution of SAR inside four organs of human body.

V. Results and Analysis

In order to study the field penetration into human body and calculate the SAR, the E-field at a single frequency is estimated as a uniform sinusoid plane wave of that particular frequency. The Conductivities and mass density of tissues as presented in table3 are also needed [23].

Table3. Densities and electrical properties of the tissues.

Tissue	Masse density $\rho(kg\,m^{-3})$	2.4 GHz ε_r	2.4 GHz $\sigma(S\,m^{-1})$
skin	1125	38.06	1.44
Tendon, pancreas	1151	43.21	1.64
Fat	943	5.29	0.10
Cortical bone	1850	11.41	0.38
Cancellous bone	1080	18.61	0.79
Blood	1057	58.35	2.50
Muscle	1059	54.49	1.84
Grey Matter (brain)	1035.5	48.99	1.77
White matter	1027.4	36.23	1.19
Cerebro_spinal_fluid	1000	66.32	3.41
Vitr. humor (eye)	1000	68.24	2.44
Bladder	1132	18.03	0.67
Cartilage	1171	38.88	1.72
Gall Bladder bile	928	68.42	2.76
Thyroid/tongue	1059	57.27	1.93
Stomach/Esophagus	1126	62.24	2.17
lung	563	48.45	1.65
kidney	1147	52.86	2.39
Testis-prostate	1151	57.63	2.13
Small intestine	1153	54.53	3.13

The developed Flat Phantom Model based on the FDTD methods was used to study the variation of electric fields inside human body. Hence, the electric fields and SAR in different organs of human body are simulated by using the FDTD method implemented in Matlab codes.

A. Blood

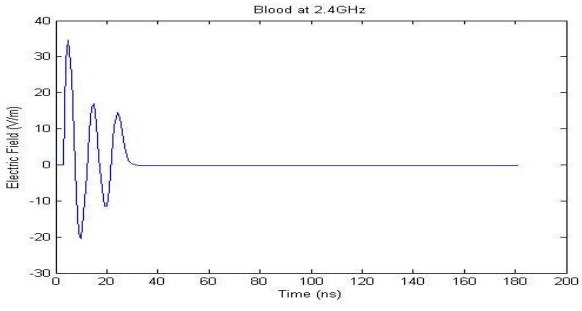


Fig. 3: Electric Field Variation

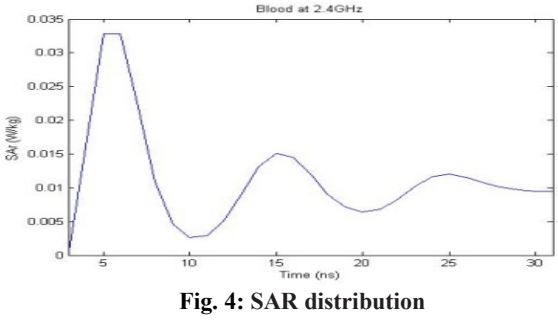


Fig. 4: SAR distribution

B. Brain

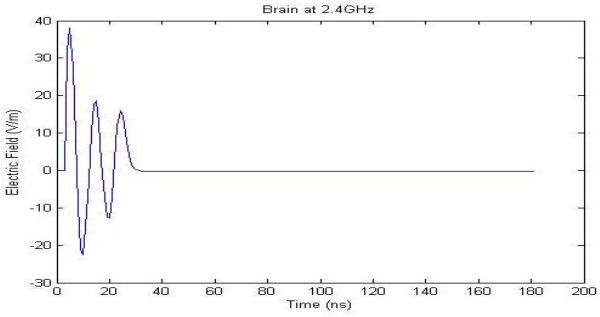


Fig. 5: Electric Field Variation

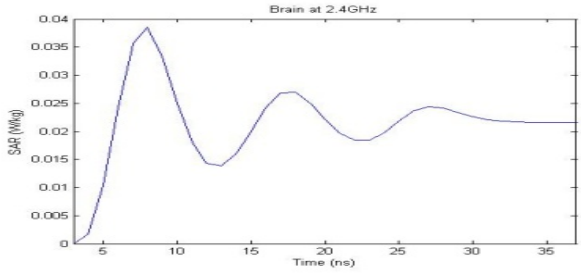


Fig. 6: SAR distribution

C. Skin

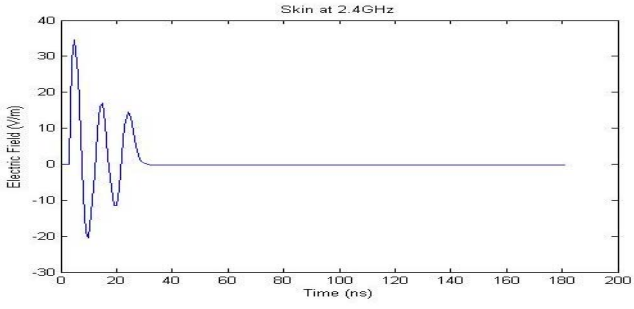


Fig. 7: Electric Field Variation

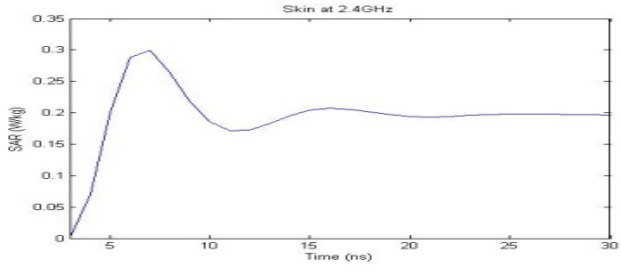


Fig. 8: SAR distribution

D. Eye

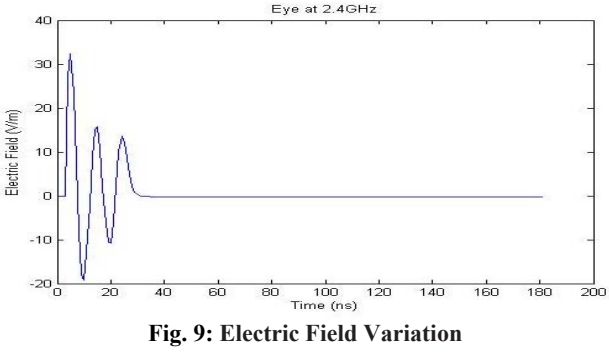


Fig. 9: Electric Field Variation

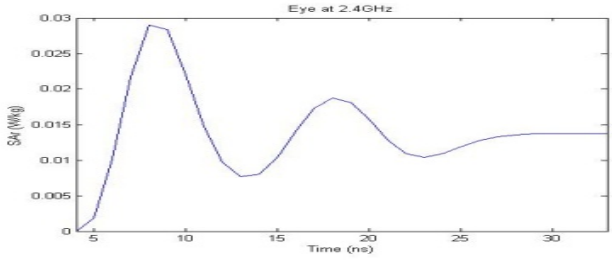


Fig. 10: SAR distribution

From the graphs, it is observed that, the Skin is the most vulnerably with the highest SAR value of 0.2914W/kg. Second is the Brain with 0.0393W/kg of SAR value. The Eye is at third place with SAR value of 0.0281W/kg. At last position, the Blood with SAR value of 0.0245W/kg. They are respectively 14.6%, 1.97%, 1.41% and 1.23% of the limit. However, due to the dielectric properties of tissues, the highest variation of electric field does not mean the highest SAR distribution inside the tissue. Hence, the blood has the highest electric field variation of 37.5V/m and SAR value of 0.0245W/kg (1.23% of the limits) while the electric field in the Brain is 35V/m only with SAR value of 0.0393W/kg (1.97% of the limits). Table4 presents SAR values of twenty organs of human body exposed in the worst place in term of electric fields at IIUM Campus.

Table 4: Highest simulated SAR Values over limits

Organs	W/Kg	%
Skin	0.2914	14.6
Tendon	0.2339	11.7
Fat	0.0212	1.06
Cortical bone	0.0261	1.31
Cancellous bone	0.0203	1.02
Blood	0.0245	1.23
Muscle	0.1105	5.53
Grey Matter (brain)	0.0393	1.97
White matter	0.0126	0.63
Cerebro_spinal_fluid	0.0273	1.37
Vitreous humor(eye)	0.0281	1.41
Bladder	0.0037	0.19
Cartilage	0.0065	0.33
Gall Bladder bile	0.2339	11.7
Thyroid/tongue	0.0212	1.06
Stomach/Esophagus	0.0261	1.31
lung	0.0203	1.02
kidney	0.0245	1.23
Testis-prostate	0.1105	5.53
Small intestine	0.0393	1.97

VI. Conclusion

The radiation from W-LAN has been measured at IIUM Campus. The measured electric field has been used as incident electric field to evaluate the distribution of electric field and SAR variation for twenty organs inside the human body. It is observed that the highest SAR is 0.2914W/kg in the Skin and the lowest SAR is 0.0037W/kg in Bladder. They are 14.6% and 0.19% of the recommended safety limits. However, long term radiation measurement and careful investigation considering all sub-bands are needed to confirm the final conclusions.

Appendixes

Appendix1: SAR limits recommended by ANSI/IEEE and ICNIRP

Organizations	Exposure Characteristics	Frequency Range	Whole-body average SAR (W/kg)	Localized SAR (Head) in W/kg	Localized SAR (Limbs) in W/kg
ANSI/IEEE	Occupational	100KHz-6GHz	0.4	8	20
	General Public	100KHz-6GHz	0.08	1.6	4
ICNIRP	Occupational	100KHz-10GHz	0.4	10	20
	General Public	100KHz-10GHz	0.08	2	4

Appendix2: Radio frequency and Microwaves radiation exposure limits for member of the public as recommended by ICNIRP and adopted by some countries

Country	Radio frequency and Microwaves			
	Frequency	Electric field (v/m)	Magnetic field (a/m)	Power density (mw/cm^2)
USA /ANSI	100MHz-1GHz	61.4 (f/100)	0.163 (f/100)	f/100
IEEE	1GHz-300GHz	194.16	0.515	10
MALAYSIA	300MHz-1.5GHz	$1.616\ f^{0.5}$	$0.00433\ f^{0.5}$	f/1500
(MCMC)	1.5GHz-300GHz	62	0.16	1

Acknowledgement

The authors wish to acknowledge the support from the International Islamic University Malaysia-Research Management Centre by funding this research through IIUM/504/RES/G/14/3/05/FRGS 0106-07 grant and to express their gratitude to Telekom Malaysia R & D unit for their portable EMF, measurement system TS-EMF, portable spectrum analyzer FSH3 and the RFEX software which were used for measurements.

References

[1] M. Rafiqul Islam and Hikma Shabani, "EM Radiation and Evaluation of Specific Absorption Rate (SAR) in Human body Exposed to Wireless-Base Station Fields at IIUM Campus," ICCCE 2008, Kuala Lumpur, Malaysia, May 2008.

[2] A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method.* Artech House, Inc, 1995.

[3] David B. Davidson, *Computational Electromagnetics for RF and Microwave Engineering,* Cambridge CB2 2RU, UK, 2005.

[4] Federal Communications Commission (FCC), "Tissue dielectrics", http://www.fcc.gov/fcc-bin/dielec.sh

[5] Ronald J. Spiegel, M.B.A. Fatmi, Stanislaw S. Stuchly and M.A.Stuchli, "Comparison of Finite-Difference Time-Domain SAR Calculations with Measurements in a Heterogeneous Model of Man," Transactions on Biomedical Engineering, Vol. 36, no. 8, (1989).

[6] ICNIRP Guidelines, "Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields up to 300GHz," Health physics, Vol. 74, no. 4, pp. 508-509, April 1998.

[7] ANSI C95.1, "American National Standard safety levels with respect to human exposure to radio frequency electromagnetic fields, 300kHz to 100GHz," September, 1982.

[8] Md. Rafiqul Islam, "Radiation Measurement from Mobile Base Station at a University Campus in Malaysia". American Journal of Applied Sciences 2, 2006.

[9] IEEE C95.1, "IEEE standard for safety levels with respect to human exposure to radio frequency electromagnetic fields, 3kHz to 300GHz," September, 1991.

[10] Rohde and Schwarz, "Portable EMF Measurement System TS-EMF and portable spectrum analyzer FSH3 and the RFEX software: User manual", 2003

[11] J. C. Lin, "Biological Bases of Current Guidelines for Human Exposure to Radio-Frequency Radiation," IEEE Antennas and Propagation Magazine, Vol.45, no.3, September 2003,

[12] J. C. Lin, "Microwave cataracts and personal Communication radiation," IEEE Microwave, vol. 4, pp. 26–32, September 2003.

[13] M. A. A. Karunarathna and J.Dayawana, "Human Exposure to RF radiation in Sri Lanka," Sri Lankan Journal of Physics, Vol. 6, pp.19-32 (2005).

[14] California EMF Program, "An evaluation of the possible risks from electric and magnetic fields (EMFs) from power lines, internal wiring, electrical occupations and appliances," Draft 3, California Department of Health Services, Oakland, April 2001.

[15] K. S. Kunz and R. J. Luebbers, "The Finite Difference Time Domain Method for Electromagnetics," CRC Press, 1993.

[16] K. S. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Medi," IEEE Transactions on Antenna and Propagation, Vol. AP-14, No.3, 1996.

[17] W. Sui, "Time Domain Computer Analysis of Nonlinear Hybrid Systems," CRC Press, 2002.

[18] M. Rafiqul Islam and Hikma Shabani, "Development of Analytical Flat Phantom Model for EM Radiation to Evaluation Specific Absorption Rate (SAR) in Human body Exposed to Wireless-Base Station Fields at IIUM Campus," ICCCE 2008, Kuala Lumpur, Malaysia, May 2008.

[19] Taflove A, "Computational Electrodynamics: The Finite-Difference Time-Domain Method," Artech House, Inc, 1995

[20] Ericsson AB, "Radio waves and health: Mobile communications," 2006.

[21] NCRP, "National Council on Radiation Protection & Measurement," www.ncrponline.org/Publications, 2001.

[22] J. M. Osepchuk and R. C. Petersen, "Safety Standards for Exposure to RF Electromagnetic Fields," IEEE, 2001.

[23]. J. Keshvari., "Emerging Wireless Mobile Technologies, Measurement & Computational RF Compliance Assessment", EMF Research and Standards, Nokia, 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Maximization of System Lifetime in Body Sensor Networks

[1] Md. Nazrul Islam Mondal [2] Kazi Mohiuddin Ahmed
[1] Department of Computer Science & Engineering, Faculty of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Bangladesh, Email: nimbd@yahoo.com
[2] Department of Telecommunications, School of Engineering & Technology, Asian Institute of Technology, Bangkok, Thailand, Email: kahmed@ait.ac.th

Abstract- In this paper, the target is to reduce the energy consumption in the body sensor network as well as maximize the system lifetime of sensor nodes when they will make communication among body sensors and personal communication unit. The best compression technique like LPC is selected for energy saving based on some calculations. Formulation of a linear programming problem where is to maximize the system lifetime which is equivalent to the time until the first node runs out of battery. Maximum system lifetimes are calculated by MATLAB optimization technique using and without using efficient compression algorithm like LPC in various environments. The results show that maximum system lifetimes calculated in different scenarios using efficient compression technique like LPC is better than without using compression technique.

I. Introduction

Consider a wireless body sensor network of static nodes as depicted in Fig.1, where each node operates on limited battery energy. Assume that each node generate constant bit rate and send or receive information to or from neighbors with and without compression technique. Also assume that the patient's body sensor (**one sensor node for every patient to detect heart beat signal**) is fixed and its distribution is shown in the Fig.1. In this paper, we select the best compression algorithm like LPC and calculate the transmitting and receiving energy according to the new energy dissipation model shown in Fig.4. Then we formulate the linear programming problem where objective is to maximize the system lifetime. Finally compare the results with the help of optimization technique and show that maximum system lifetimes calculated in various scenarios using efficient compression technique like LPC is better than without using compression technique. Most of the previous works deal with low energy consumption for WSN but they did not maximize the system lifetime with the help of optimization technique using energy efficient LPC algorithm. In [1], the authors described the application of WSN just to monitor the battle field. In [2], the authors described the WBAN architecture for broader telemedical system using energy efficient local communication protocols such as ZigBee, Wi-Fi. The authors in [3] explained the energy aware network layer protocols like LEACH, CICADA, and S-MAC for maximize the battery lifetime.

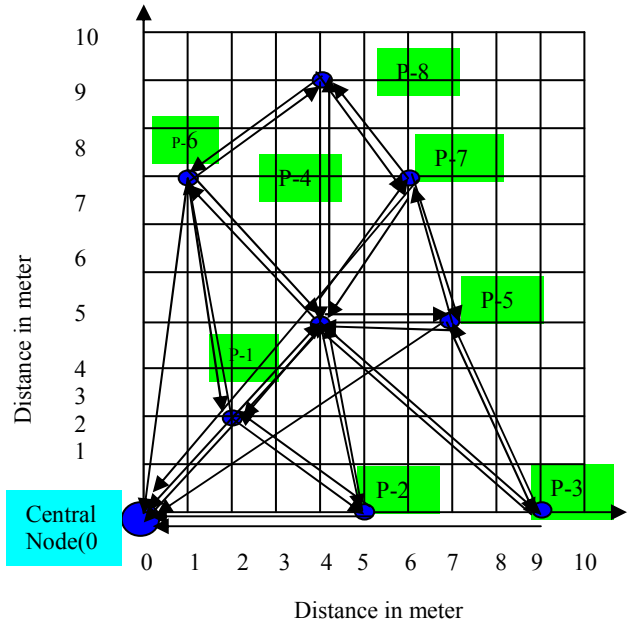


Fig 1 Application scenario for maximum system lifetime

The work in [4] considered multi destination case instead of single destination case for maximizing the system lifetime. In [5], instead of minimizing the total consumed energy, they tried to maximize the network lifetime till first node dies.

II. Maximum System Lifetime

Maximum system lifetime which is equivalent to the time until the first node runs out of battery for a particular network.

A. Formulation of linear problem

The formulation of maximum system lifetime routing problem is described by the author for the case where information generation rates are fixed [4], [6]. Since T_{sys} in equation (1) is not a linear function of flow variables, x_{ij}^{k} , to formulate as a linear programming problem, the author defined T_{sys}^{-1} as an inverse system lifetime shown in equation (6) [6]. Linear programming problem is formulated considering some parameters described below:

▪ **Given information for the network**

The parameters that are considered to formulate the problem for the network as follow [6]**:**

- υ **:**set of nodes (including node 0 which is central node)
- $\upsilon^{'}$ **:** set of nodes (excluding node 0)
- υ_i **:** set of nodes that can be reached by node *i*
- t^k **:** traffic rate from sensor node *k* to central node
- E_i **:** initial battery energy of sensor node *i*
- α_{ij}^{t} **:** energy consumed by sensor node *i* to transmit a packet (total bits) to node *j*
- α_i^{r} **:** energy consumed by sensor node *i* to receive a packet (total bits)

So, the *Lifetime* of node i is written by

$$T_i = \frac{E_i}{\sum_{k\in\upsilon^{'}}\sum_{j\in\upsilon_i}\alpha_{ij}^{t}x_{ij}^{k} + \sum_{k\in\upsilon^{'}}\sum_{j\in\upsilon_i}\alpha_i^{r}x_{ji}^{k}} \quad (1)$$

▪ **Decision variables, system lifetime and their relations**

- x_{ij}^{k} **:** traffic flow in bps from node i to node j for data of sensor node k.

The *Lifetime* of node i, denoted by T_i can be computed as in (1) follows [5]**,** [6]**.**
But, our target is to maximize the system *lifetime*, called MLR (Maximum Lifetime Routing), denoted by $\boldsymbol{T_{sys}}$**.** The system lifetime under the flow x_{ij}^{k} can be defined as the length of time until the first node dies among all nodes in the network, which is the same as the minimum lifetime over all nodes, shown in equation (2).

$$T_{sys} = \min_{i\in\upsilon^{'}} T_i \quad (2)$$

So, maximizing $T_{sys} = \min_{i\in\upsilon^{'}} T_i$ can be rewritten by

$$\text{minimizing } T_{sys}^{-1} = \max_{i\in\upsilon^{'}} T_i^{-1} \quad (3)$$

▪ **Constraints for the problem**

To maximize the system lifetime, the author needs to find the flow under flow conservation condition and satisfaction of traffic demands as follows [6]**:**

Flow conservation and satisfaction of traffic demands**:**

$$\forall i\in\upsilon, k\in\upsilon^{'}, \sum_{j\in\upsilon_i}x_{ji}^{k} - \sum_{j\in\upsilon_i}x_{ij}^{k} = \begin{cases} -t^k, i=k \\ t^k, i=0 \\ 0, otherwise \end{cases} \quad (4)$$

Nonnegativity of flows:

$$\forall k\in\upsilon^{'}, i\in\upsilon, j\in\upsilon_i, x_{ij}^{k} \geq 0 \quad (5)$$

▪ **Overall optimization problem**

The overall problem is to minimize the cost function mentioned in equation (2) subject to the above set of constraints. From equation (1), it is obvious that the cost function is not yet a linear function of decision variables but it should be linear as considered as linear programming problem. So, the overall linear optimization problem is as follows [5], [6]:

$$\text{minimize } T_{sys}^{-1} \quad (6)$$

subject to

$$\sum_{k\in\upsilon^{'}}\sum_{j\in\upsilon_i}\alpha_{ij}^{t}x_{ij}^{k} + \sum_{k\in\upsilon^{'}}\sum_{j\in\upsilon_i}\alpha_i^{r}x_{ji}^{k} - T_{sys}^{-1}E_i \leq 0 \quad (7)$$

$$\forall i\in\upsilon, k\in\upsilon^{'}, \sum_{j\in\upsilon_i}x_{ji}^{k} - \sum_{j\in\upsilon_i}x_{ij}^{k} = \begin{cases} -t^k, i=k \\ t^k, i=0 \\ 0, otherwise \end{cases} \quad (8)$$

$$\forall k\in\upsilon^{'}, i\in\upsilon, j\in\upsilon_i, x_{ij}^{k} \geq 0 \quad (9)$$

III. Selection of Efficient Algorithm

The energy consumption of the transceiver depends on the both system and circuit level design. If the CMOS circuits for designing the sensor nodes increase, total energy consumption must be increased. It is given that 20k and 10k CMOS gate circuits are needed for compression and decompression respectively [7], [8]. Every 1k gate needs 24 µW powers for compression and decompression. So, computation power depends on the hardware of the sensor nodes. Another one is the transmission power and it is related to data rate minimization, distance and frequency. This data rate minimization can be achieved by local processing at the sensor node like data compression algorithms. We compare the different data compression algorithms by the following equation (10), [7], [8] and select **LPC** as the best algorithm for our purpose based on some calculations.

$$E_{total} = P_{TX}\times(T_{ON-TX}+T_{START-TX}) + P_{OUT}\times T_{ON-TX} + d\times P_{RX}\times(T_{ON-RX}+T_{START-RX}) + P_{OFF}\times T_{OFF} \quad (10)$$

Where,

$P_{TX/RX}$ = compression and decompression power consumption of the transceiver

$T_{ON-TX/RX}$ = transmitter/receiver on-time

$T_{START-TX/RX}$ = start up time of the transceiver

P_{OFF} = power consumption of the transceiver in the standby/off state

T_{OFF} = transceiver off-time =*1*-[$T_{ON-TX} + T_{START-TX} + d\times(T_{ON-RX}+T_{START-RX})$]

d = ratio of received packets to transmitted packets

- **Energy consumption in different methods varying transceiver ON time and Data Rate**

We calculate the energy consumptions in different methods varying ON time of the transceiver with help of equation (10) and finally compare to select the best algorithm like LPC that are shown in the following Fig.2 and Fig.3.

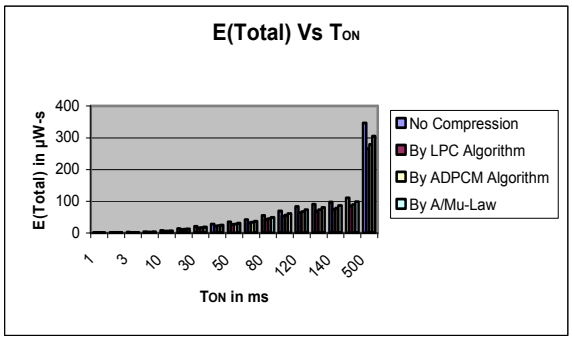


Fig.2 Energy consumption in different methods varying transceiver ON time

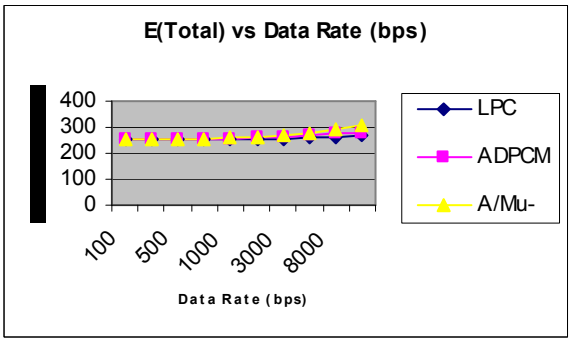


Fig.3 Energy consumption in different methods varying transceiver data rate

IV. New Energy Dissipation Model

We used our new energy dissipation model for calculating the transmitting and receiving energy consumption of the transceiver which is shown in the following Fig.4 [9].

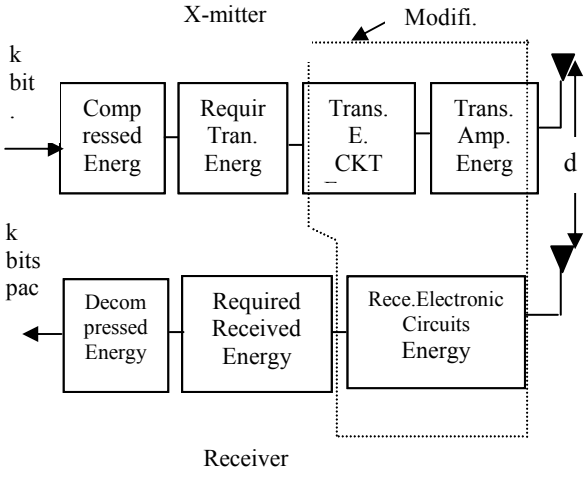


Fig.4 New energy dissipation model

Case I: Without compression algorithm

We calculate transmit and receive energy consumption according to proposed radio energy dissipation model for our work as follows [9]:

Transmit energy consumption,

$$\alpha_{ij}^{t} = E_{total} + P_{out} \times T_{on} + P_{off} T_{off} + E_{amp} \times d_{ij}^{n} \quad (11)$$

Where,

E_{total} = total electronic circuit energy consumption (i,e ADC, filtering etc)

$P_{out} \times T_{on}$ = required transmission energy consumption

$P_{off} T_{off}$ = transmitter off energy consumption

E_{amp} = amplifier energy consumption

d_{ij}^{n} = distance between node i and j where n is the path loss exponent

Received energy consumption,

$$\alpha_{i}^{r} = E_{total} + P_{rec} \times T_{on} + P_{off} T_{off} \quad (12)$$

Where,

E_{total} = total electronic circuit energy consumption (i,e ADC, filtering etc)

$P_{rec} \times T_{on}$ = required received energy consumption

$P_{off} T_{off}$ = receiver off energy consumption

▪ Case II: Using compression algorithm

We also calculate transmit and receive energy consumption according to proposed radio energy dissipation model for our work as follows [9]:

Transmit energy consumption,

$$\alpha_{ij}^{t} = E_{total} + P_{TX} \times \left(T_{ON-TX} + T_{START-TX}\right) + P_{out} \times T_{on} + P_{off} T_{off} + E_{amp} \times d_{ij}^{n} \quad (13)$$

Where,

E_{total} = total electronic circuit energy consumption (i,e ADC, filtering etc)

$P_{TX} \times \left(T_{ON-TX} + T_{START-TX}\right)$ =compression circuit energy with start-up time of compressor

$P_{out} \times T_{on}$ =required transmission energy consumption

$P_{off} T_{off}$ = transmitter off energy consumption

E_{amp} = amplifier energy consumption

d_{ij}^{n} = distance between node i and j where n is the path loss exponent

Received energy consumption,

$$\alpha_{i}^{r} = E_{total} + P_{RX} \times \left(T_{ON-RX} + T_{START-RX}\right) + P_{rec} \times T_{on} + P_{off} T_{off} \quad (14)$$

Where,

E_{total} = total electronic circuit energy (i,e ADC, filtering etc)

$P_{RX} \times \left(T_{ON-RX} + T_{START-RX}\right)$ = decompression circuit energy consumption with start-up time of decompressor

$P_{rec} \times T_{on}$ = required received energy consumption

$P_{off} T_{off}$ = receiver off energy consumption

V. Performance Comparison

We measure the performances of our system by the following ways:

- **Case I: For free space model (When n = 2)**

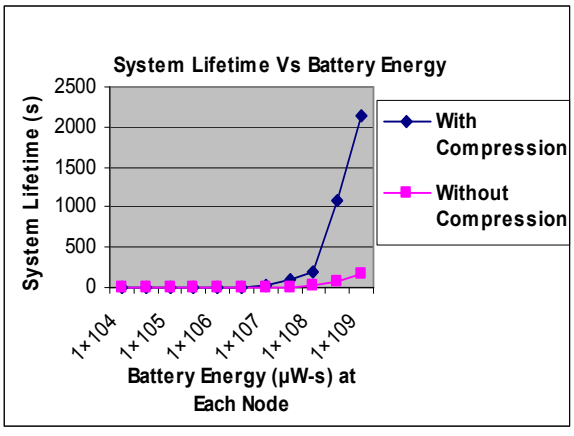


Fig.5: Comparison of the system lifetimes in 8-node body sensor networks

- **Case II: For multipath model (When n = 4)**

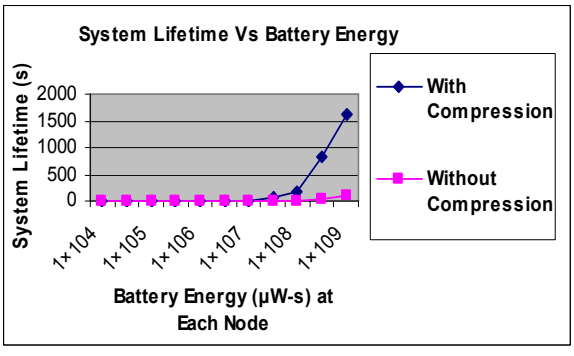


Fig.6 Comparison of the system lifetimes in 8-node body sensor networks

VI. Conclusion

In wireless body sensor networks, the nodes operate on limited battery energy. Our major target to support uninterrupted operation by enhancing the battery lifetime of the body sensor nodes for diagnosis the heart beat signal. Therefore, the main objective of this paper is to study and analyze the different algorithms for selecting the best one for low energy consumption in the body sensor nodes and then apply this efficient technique to maximize the system lifetime for a particular application scenario. In this research work, a new energy dissipation model is proposed and applied this best technique for this improved model to calculate the energy consumption parameters like transmitting and receiving energy consumption with respect to a particular application environment. MATLAB optimization technique is used to find out the system lifetimes with respect to our application scenario where we assumed our problem as a linear programming problem. Finally, we compared our results (system lifetimes) using and without using compression technique in two different environments like free space model where n = 2 and multipath model where n = 4. Results show that the maximum system lifetime using compression technique is definitely better in both cases.

References

[1] Pearson Matthew, Beckford Jamal & Akin John, "Wireless sensor Network Proposal, ECE290", 2007. URL: www.engr.uconn.edu/ece/SeniorDesign/projects/ecesd83/Proposal.pdf.

[2] Otto Chris, Milenkovic Aleksandar, Sanders Corey & Jovanov Emil, "System Architecture of a Wireless Body Area Network for Ubiquitous Health Monitoring". Journal of Mobile Multimedia, Vol. 1, No 4, 307-326, 2006.

[3] Estrin D., Heidemann J, & Ye W, "Energy Efficient MAC Protocol for Wireless Sensor Networks", In Proceedings of the IEEE Infocom, USC/Information Sciences Institute. New York, NY, USA: IEEE June, pp. 1567-1576, 2002

[4] Chang Jae-Hwan & Tassiulas L., "Energy Conserving Routing in Wireless Ad-hoc Networks,". In Proceedings of IEEE INFOCOM, Tel Aviv, Israel, Mar, pp.22-31, 2000.

[5] Chang Jae-Hwan & Tassiulas L., "Maximum Lifetime Routing in Wireless Sensor Networks". IEEE/ACM Transactions on Networking, Vol.12, No.4, August, 2004.

[6] Saengudomlert P., "Lecture Note on Optimization for Communications and Networks". AT77.9011, P-32-36, Dept. of Telecommunications, School of Engineering & Technology, Asian Institute of Technology, August,2007. URL:http://203.159.18.14/tcweb/AT77_9011_Optimization.

[7] Toumaz Technology Limited, "Consideration for the Design of an Ultra-Low Power Wireless Transceiver for Sensor Network Applications". UbiMon First Deliverable, 85 Milton Park Abingdon, UK, March 2005.

[8] Toumaz Technology Limited, "Trade-Off in Communication versus Computation for Body Sensor Network Nodes". UbiMon Second Deliverable, 85 Milton Park Abingdon, UK, April, 2005.

[9] Heinzelman Wendi B., Chandrakasan Anantha P. & Balakrishnan Hari, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks". IEEE Transactions on Wireless Communications, Vol. 1, No. 4, October, 2002.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Power spectral analysis for identifying the onset and termination of obstructive sleep apnoea events in ECG recordings

Ahsan H. Khandoker, *Member, IEEE,* Chandan K. Karmakar, *Student member, IEEE,* and Marimuthu Palaniswami, *Senior Member, IEEE*

Department of Electrical & Electronic Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia. Email: ahsank@unimelb.edu.au.

Abstract— The high prevalence of obstructive sleep apnoea (OSA) requires a simplified, unattended screening device that would be useful for diagnosis at the early stage. This study presents a method for screening individual OSA event based on sleep ECG signal. The overnight ECG recordings were divided into 5-second epochs containing normal (N) breathing and onset (O), maximum (M) & termination (T) of OSA events. Power spectral analysis of ECG epochs was employed to extract features. The area under receiver operating characteristics curve was estimated to determine the discrimination capability of each feature (or power in each frequency bin). The maximum ROC areas for N/O, N/M and N/T were found to be 0.78, 0.81,0.71 in the ranges of powers of 57-65 Hz, 52-72 Hz, 52-66 Hz bands respectively. An heuristic rule was applied to recognize the individual OSA events from spectral features of N,O,M,T epochs. Results show good agreement with the original annotations in an overnight sleep study. These results, therefore, could have considerable potential in ECG based screening and can aid sleep specialist in the assessment of patients with suspected sleep apnoea syndrome.

I. INTRODUCTION

OBSTRUCTIVE sleep apnoea (OSA) is commonly characterized by recurrent airflow obstruction due to total or partial obstruction of the upper airway. This collapse of the upper airway occurs when tongue or soft palate touches the posterior pharyngeal wall. In previous studies, it is shown that the prevalence of OSA is 4% in adult men and 2% in adult women [1] and considered as highly under-diagnosed due to unavailability of effective treatment [2-3]. The most common symptom for OSA is
daytime sleepiness, which in turn decreases patient's effectiveness in regular activities. It has a strong link with cardiovascular diseases, poor cognitive performance, and increased risk of motor vehicle and workplace accidents due to reduction in sleep quality [4].

OSA is commonly known as sleep apnoea whereas hypopnoea is defined as partial collapse of narrowed pharynx during OSA. The reduction of thoracoabdominal amplitude more than 95% is considered as apnoea whereas 50% reduction in thoracoabdominal amplitude along with 4% reduction in blood oxygen is considered as hypopnoea [5]. For both apnoea and hypopnoea, if the obstruction in airflow exists for at least 10 seconds then the event is considered to be an apnoea/hypopnoea event. The severity of apnoea can be defined using the Apnoea Hypopnoea Index (AHI), which is calculated as the number of apnoea-hypopnoea events per hour. According to the Chicago criteria AHI<5 (no apnoea), AHI=5-15 (mild), AHI=15-30 (moderate) and AHI>30 (severe) [6].

The standard technique for diagnosing apnoea is whole night polysomnography at sleep laboratories/ clinics. This process is very expensive as well as inconvenient for the patient. Therefore, there has been considerable interest in research for introducing an inexpensive home monitoring of sleep apnea. Electrocardiogram (ECG) signal contains concealed information about central autonomic control of cardiovascular function, respiration and the electric activity of heart. Also, OSA diagnosis using ECG would be an inexpensive, less invasive and more convenient approach.

Different algorithms on cheaper ambulatory ECG monitoring technology have recently been proposed to detect OSA [7]. However, the reported methods can detect if there is an OSA event during any given minute

This study aims to recognize actual OSA events from normal breathing events using power spectral density (PSD) features of ECG signal over 5 second sliding window (onset, maximum and termination of the event). Model

based scoring results have been compared with original scoring for the events.

II. METHODS

A. ECG and epoch scoring

The polysomnograms of 10 sleep apnoea patients [(mean ±SD) age 54±9 yrs, body mass index (BMI) 30±2 kg/m2] were analysed. All subjects were free of any cardiac history. Diagnosis was based on clinical symptoms and polysomnographic (PSG) outcomes. Respiratory events were scored using criteria proposed by the AASM[8]. Obstructive apnoea was defined as the absence of oronasal airflow for >10 s in the presence of persistent respiratory efforts. The average length of ECG recording was around 10.5 hours. Sampling frequency of the ECG signals was 256Hz. The OSA events were manually scored using the Polysomnography [8]. The whole ECG signal was then divided into five seconds epochs. Any epoch was marked as normal (N) if there was no apnoea event during that time. On the other hand, any 5-second epochs during the OSA event was marked as OSA maximum (M). For the epochs containing a combination of 2 seconds of normal and first/last 3 seconds of OSA event or vice versa were marked as onset (O) or termination (T). All other types of apnoea events or epochs were discarded in this study. In total, 30,753 N, 2861 O, 9376 M and 2861 T epochs were used in this study.

B. Power spectral analysis

Power spectral density (PSD) calculates the strength of the variations (energy) as a function of frequency. It represents at which frequencies variations are strong and at which frequencies variations are weak. In other words, it is the estimation of distribution of power contained in a signal over frequency range. The unit of PSD is energy per frequency and the energy within a specific frequency range can be computed by integrating PSD within that frequency range. Each epoch of ECG signal had 1280 data points. ECG signal were a time series data associated with each 5 seconds segment. In order to eliminate the effect of low frequency noise (ex: baseline drift) each epoch were passed through a high pass filter with cut-off frequency of 4 Hz. The noise was especially involved with base line wandering due to body position movement. The sequence was then zero padded to length 2048, the next power of 2 larger than the number of data points.

In this study we used the MATLAB R2006b implementation of Welch's method for calculating power spectral density (PSD). In this method the data were zero-padded to length of the FFT and divided into overlapping segments. Then PSD of each segment was computed and averaged. This averaging decreases the variance of the estimate relative to a single periodogram estimate of the entire data record [9]. We used FFT length of 128 data points which is divided into eight segments with 50% overlap between them.

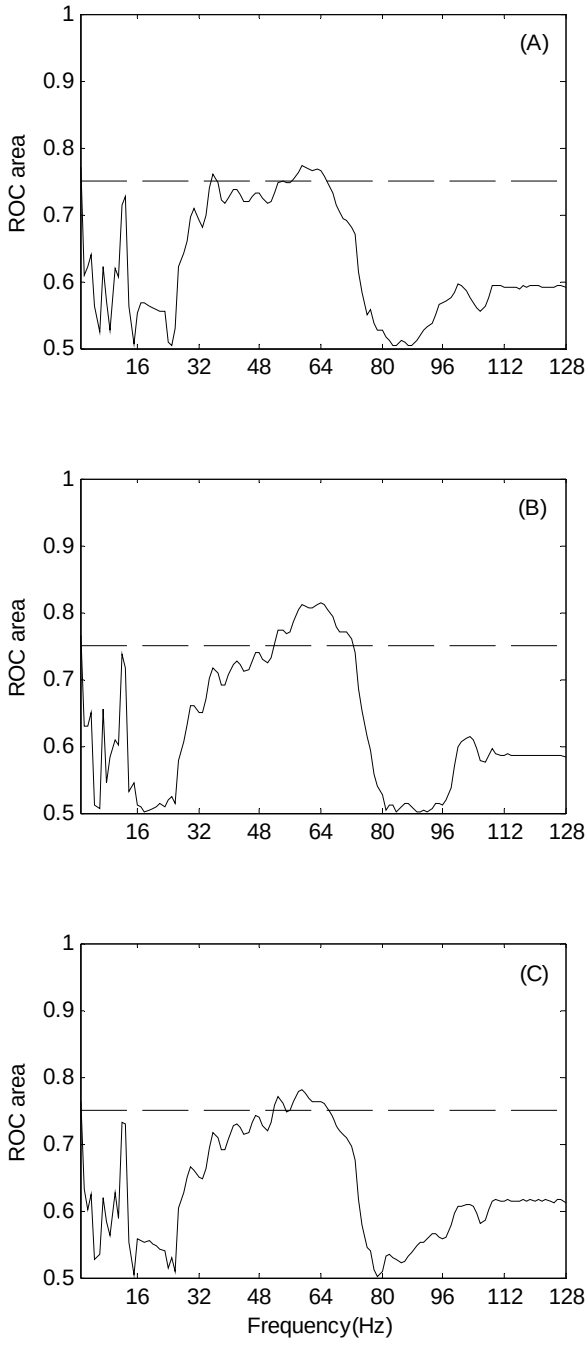


Fig. 1. ROC area between Normal/Onset (A), Normal/Maximum (B) and Normal/Termination (C) epochs. Dashed line represents the threshold value (0.75) of ROC area used for selecting significant band.

The resultant PSD vector length was 129 which also represented the number of frequency bin. The resolution of each bin was 1 Hz. Normalized PSD values were estimated by diving the absolute PSD values at each frequency bin by total PSD value (2-128 Hz).

C. Statistics and ROC analysis

MATLAB statistics toolbox was used to perform the statistical analyses. In order to provide the relative importance of features, receiver-operating curve (ROC) analysis was used [10], with the areas under the curves for each feature represented by the ROCarea. An ROCarea value of 0.5 means that, the distributions of the features are similar in two groups with no discriminatory power. Conversely, a ROCarea value of 1.0 would mean that the distributions of the features of the two groups do not overlap at all. ROC plots are used to gauge the predictive ability of a classifier over a wide range of threshold values. A threshold value is applied such that a value below the threshold was assigned into one category whereas a value equal to or above the threshold is assigned into another category. ROC curves are then plotted using results to examine qualitatively the effect of threshold variation on the classification performance. The area under ROC curve was approximated numerically using the trapezoidal rules [10] in this study. The larger the ROC area the better the discriminatory performance.

III. RESULTS AND DISCUSSION

PSD features on 5-second epochs were estimated in the frequency range of 1-128 Hz (bin size=1 Hz) for N,O,M,T classes. The frequency domain analysis was employed for the study because epoch length is too short to extract information using time domain analysis. Figure 1A represents the ROC areas between Normal and Onset epochs. It shows that the PSD values of Normal and Onset epochs were better discriminated within the frequency range of 31-66 Hz. The threshold was set at 0.75 to select the significant frequency band. A narrow band at 1 Hz was ignored. The maximum ROC area was found to be 0.77 at 59 Hz. However, the ROC area of the band 31-66 Hz was found to be 0.78, which is higher than that of a single frequency bin.

The ROC areas between Normal and OSA-maximum epochs were shown in Figure 1B. Like Figure 1A, maximum discriminating features were found to be in the higher frequency range of 35-73 Hz. The significant frequency band for Normal and OSA-maximum epochs were found from 52-72 Hz with the same threshold as applied in Figure 1. The maximum ROC area was 0.81 at 64 Hz.

In Figure 1C, ROC areas between Normal and Termination epochs were shown. The frequency range 35-72 Hz shown to have contained the most discriminative features. The significant band was selected in the frequency range of 52-66 Hz and the maximum ROC area was found to be 0.78 at 59 Hz. Table I summarizes the mean and standard deviation (SD) of powers in significant frequency bands for all types of epochs with the ROC areas.

In this paper, we described our preliminary study of how OSA events can be detected using single lead ECG. By using the PSD analysis, it is possible to detect the significant frequency bands that correlate with onset, maximum and termination of OSA event.

The significant frequency bands found in all cases were in high frequency range. From the literature [12], it was found that the chest muscle electromyography (EMG) power reflecting respiratory muscle activity dominates the QRS power in the frequency over 35 Hz. We speculate that differences in the intensity of respiratory effort during onset, maximum and termination of OSA events are reflected in higher frequency bands of ECG signals. The more specific physiological explanation of the results remains to be studied.

ROC analysis shows the average accuracy of 79%, 80% and 81% in recognizing O, M and T epochs respectively from N epochs. In order to detect a full OSA event, an heuristic rule was applied. Threshold values 6.30E-04 and 8.9E-04 were calculated from powers in non-overlapping bands 58-66 Hz and 53-58 Hz for O and T epochs respectively. At first, M epochs in ECG signal were identified using the threshold value 23E-04. Then O and T epochs were searched in all M epochs to define the full events (Fig. 2). However, the heuristic rule (N-O-M-T-N sequence) to define the OSA events that was set at 10 second (two consecutive 5 second M epochs) needs to be improved. For example, two 10 second events separated by a 5-second N epoch could have been combined to one event. Further research on designing an automated classifier using the spectral features of N, O, M and T epochs needs to be done. These results indicate the possibility of noninvasively recognizing OSA events from normal breathing events based on shorter segments of ECG signals.

Table 1. Mean and SD of total power distribution in the significan frequency bands for all groups of epochs. ROC area represents discrimination capability of the frequency band among epochs.

Frequency Band (Hz)	Epoch Type PSD(Mean ± SD) (V^2/Hz)	Epoch Type PSD(Mean ± SD) (V^2/Hz)	ROC area
58-66	Normal 0.0014 ± 0.0021	Onset (O) 0.0035 ± 0.0042	0.78
53-73	Normal 0.0036 ± 0.0053	Maximum (M) 0.0110 ± 0.0095	0.81
53-67	Normal 0.0031 ± 0.0046	Termination (T) 0.0084 ± 0.0088	0.78

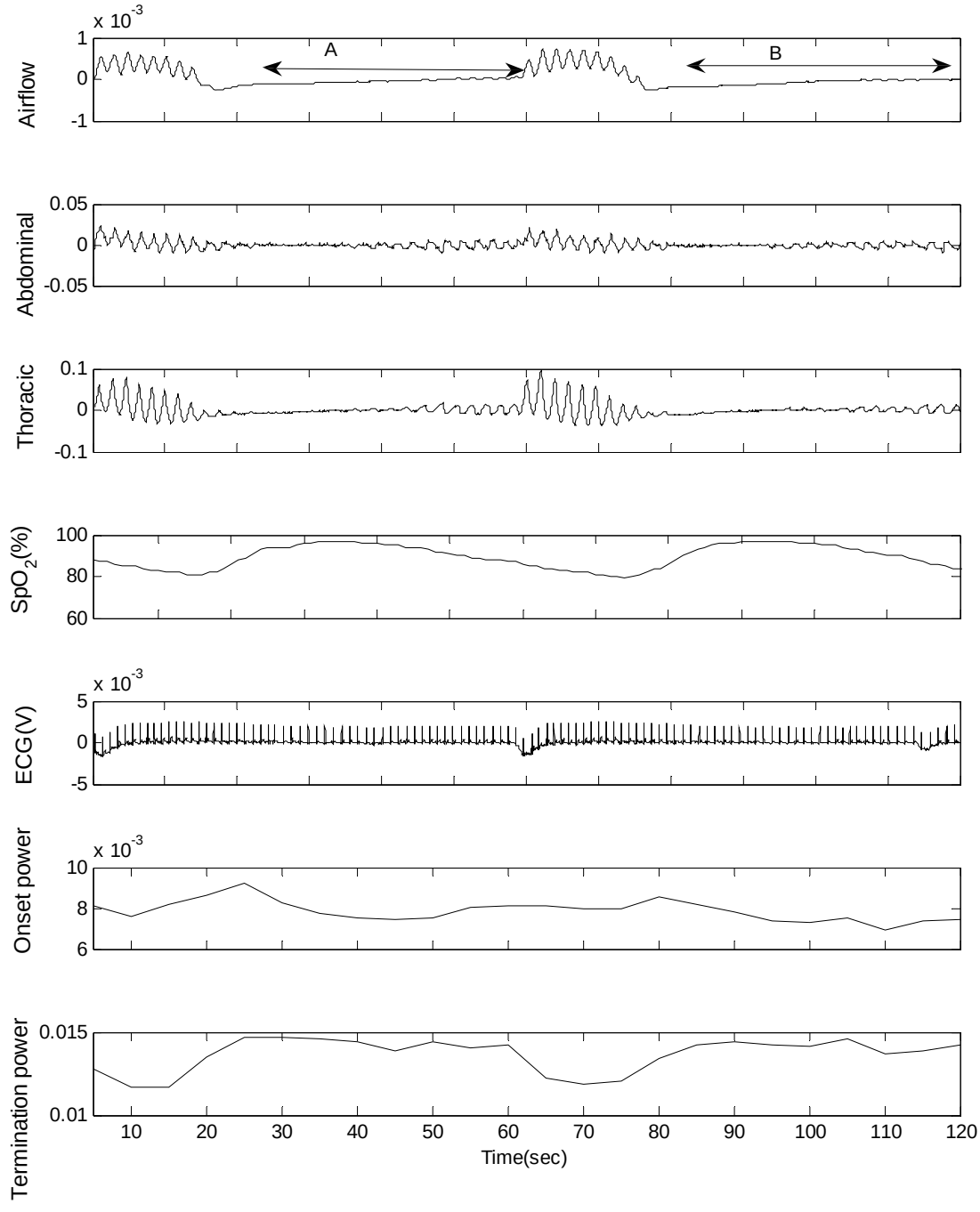


Fig. 2. An example of nasal airflow, abdominal, thoracic, pulse oximetry, ECG, PSD values of the ECG signals over 58-66 Hz band for the onset and 53-58 Hz band for the termination of OSA events.

Acknowledgements

This study was supported by an Australian Research Council (ARC) Linkage Project with Compumedics Pty Ltd (LP0454378). The authors would like to thank all members of research and innovation team of Compumedics for providing sleep studies and their valuable advices, feedback and support

References

[1] G.J. Gibson. Obstructive sleep apnoea syndrome: underestimated and undertreated. Br Med Bull 2005;72:49–65.

[2] T. Young, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993;328:1230–5.

[3] R Cartwritht,. "Obstructive sleep apnea: A sleep disorder with major effects on health", Disease-a-Month, 2001. 47(4):p. 109-147.

[4] F.J Nieto., Young T.B., Lind B.K., Shahar E., Samet J.M., Redline S., D'Agostino R.B., Newman A.B., Lebowitz M.D., and Pickering T.G., "Association of sleep disordered breathing, sleep apnea, and hypertension in a large community-based study", J. Am. Med. Assoc. , 283, p. 1829-1836, 2000.

[5] G.A Gould., Whyte K.F., Rhind M.A., Arie M.A.A., Catterall J.R., Shapiro C.M., "The sleep hypopnea syndrome", Am. Rev. Respir. Dis. , vol. 137, p. 895-898, 1998.

[6] Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep 1999; 22(5): 667–89.

[7] T. Penzel, McNames J, de Chazal P, Raymond B, Murray A, Moody G. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. Med Biol Eng Comput 2002;40:402–407.

[8] AMERICAN ACADEMY OF SLEEP MEDICINE (AASM) TASK FORCE. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Sleep 1999; 22:667–689.

[9] P.D Welch., "The Use of Fast Fourier Transform for the Esimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms", IEEE Transactions on Adio and Electroacoustics, vol. AU-15, No. 2, June 1967.

[10] J.A Hanley. and McNeil B.J., "The meaning and use of the area under receiver operating characteristic (ROC) curve". Radiology; 143: 29–36.1982.

[11] P. Petalas, P. Spyridonos, D. Glotsos,D. Cavouras, P. Ravazoula, G. Nikiforidis. Probabilistic neural network analysis of quantitative nuclear features in predicting the risk of cancer recurrence at different follow-up times. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Volume 2, Issue , 18-20 Sept. 2003 Page(s): 1024 – 1027.

[12] N. V. Thakor, Webster J. G. and Tompkins W. J., "Estimation of QRS Complex Power Spectra for Design of a QRS Filter", IEEE Transactions on Biomedical Engineering, vol. BME-31, No. 11, November 1984.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Evaluation of a High-Birefringence Linearly Chirped Grating for PMD Compensation in WDM/ IM-DD Transmission System

M. S. Islam* and S. P. Majumder**
*Institute of Information and Communication Technology
**Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
E-mail : mdsaifulislam@iict.buet.ac.bd, sai_ful89@yahoo.com

Abstract - We analytically evaluated the performance of an adjustable polarization mode dispersion compensator based on a 2-cm long high-birefringence linearly chirped fiber Bragg grating and simulated a 4-channel WDM /IM-DD system using this compensator. The device can adjust differential group delay in a linearly continuous way without affecting wavelength outside the bandwidth. The various properties of the device such as relative group delay, reflectivity, differential group delay etc. are investigated in terms of wavelength and stretching ratio. Results show that under stretched condition; the device generates a time-delay between fast and slow polarization axes that is adjustable from 0 ps to 55 ps and is tunable within 2.4 nm wavelength range. Through simulation, it is also found that the power penalty is reduced from 7.10 db to 4.5 dB at a mean DGD 40 ps for a 4-channel 10 Gb/s WDM system when no stretched is applied on the grating.

I. Introduction

Polarization mode dispersion (PMD) has become one of the most difficult issues in implementing next-generation high-bit rate 10 Gb/s and beyond transmission systems. PMD is a serious problem, because a large amount of the installed fibers exhibit PMD values that are several times that of current state-of-art fibers. Degrading effects that tended to cause catastrophic events even at lower bit rates have become critical concerns of high-performance networks. First-order PMD compensation is the simplest technique to compensate PMD and it is realized by delaying one state-of-polarization (SOP) with respect to other. There have been several experiments and/or simulation work to demonstrate first-order PMD compensation [1]-[4]. Conventionally to compensate PMD in a WDM system, channels have to be demultiplexed and compensated individually resulting in increased cost and complexity.

Chirped fiber gratings are useful for PMD and chromatic dispersion compensation [5]-[7]. The linearly high-birefringence (Hi-Bi) chirped FBG has a large refractive index difference (Δn) between its fast and slow polarization axes. The birefringence, Δn causes the orthogonal polarization modes to experience two different couplings with the grating and the Bragg reflection from the birefringence chirped grating for a given signal wavelength occurs at different locations for different polarizations. The Hi-Bi fiber provides different time delay for different SOPs and the chirp of the grating provides the selectability of varying amounts of differential polarization time delay when the FBG is stretched or compressed [8]. Z. Pan *et al* [9], experimentally demonstrated chirp-free tuneable PMD compensation for a 10 Gb/s signal by using an adjustable highly-birefringence nonlinearly-chirped FBG in a novel dual-pass configuration that significantly reduces the induced chirp of the FBG. It is shown that a 45-km link interacting with the FBG induced chirp is reduced from 4.0 to 0.5 dB. X. Kun *et al* [10], based on numerical simulations reported that the efficiency of a sampled Bragg grating PMD compensator is assessed for the 10 Gb/s NRZ transmission system with 58.6 and 106 ps differential group delay (DGD), respectively by applying transverse and uniform distributed force.

In this paper, we have developed analytical formulations and evaluate the performance of a Hi-Bi linear chirped fiber Bragg grating (LCFBG) based WDM-PMD compensation device and simulation is done to investigate the feasibility to compensate several channels at the same time for a 4-channel WDM/IM-DD transmission system using this device. The reflection point inside the grating can be moved by several millimeters by stretching the grating by few microns. It is found that a 2 cm long LCFBG allows the DGD to be adjusted from 0 ps to 55 ps in a continuous way within a 2.4 nm wavelength range by stretching it by 0.2%. Simulation results show that the power penalty for a 4-channel 10 Gbps system with 40 ps average DGD is reduced from 7.10 dB to 4.5 dB.

II. LCFBG based PMD compensator

The system model of a LCFBG based PMD compensator is shown in Fig.1. The proposed delay element consists of a 4-port polarization beam splitter (PBS) and a 2 cm length Hi-Bi LCFBG. The PBS splits the incoming optical signal into two orthogonal polarizations. The fast axis polarization signal (P_f) enters the Hi-Bi LCFBG from the longer

wavelength port and slow one (P_s) from the shorter wavelength port. The polarization states of the signal (λ_i) within the bandwidth of the grating will be reflected and differently delayed by the grating. By properly tuning the LCFBG, we can adaptively generate required DGD ($\Delta\tau$) for the PMD compensation. The two orthogonally polarized optical signals are then combined without interference and the compensated signal is directed to output port 1 of the PBS. An optical circulator can be used to separate the input signal and the PMD compensated output signal. All light (λ_0) outside the reflection bandwidth of the grating will not be affected and detected to output port 2.

II. WDM-PMD compensation system model

Fig.2 shows the typical module of the compensation scheme using the LCBG based compensator. It shows the power penalty (at the input and output) for different DGD values when the input power is evenly divided between the two PSPs. The power penalty for small (such as 10 ps) DGD is usually negligible or within the power margin of the system and therefore be tolerated. It is the high values that cause significant degradation when signal power splits almost equally between the two PSPs. At the input 4-channels (bit rate, 10 Gbps per channel) with wavelength λ_1, λ_2, λ_3 and λ_4 are multiplexed and fed to fiber link. In our simulation, we take two span of 200 km length fiber link. The LCFBG compensators are placed after 200 km fiber link length where fiber loss is totally compensated by EDFA. Finally, the channels are demultiplexed at the receiving end and eye diagram is monitored without and with compensator to assess the deterioration and amount of compensation of each channel.

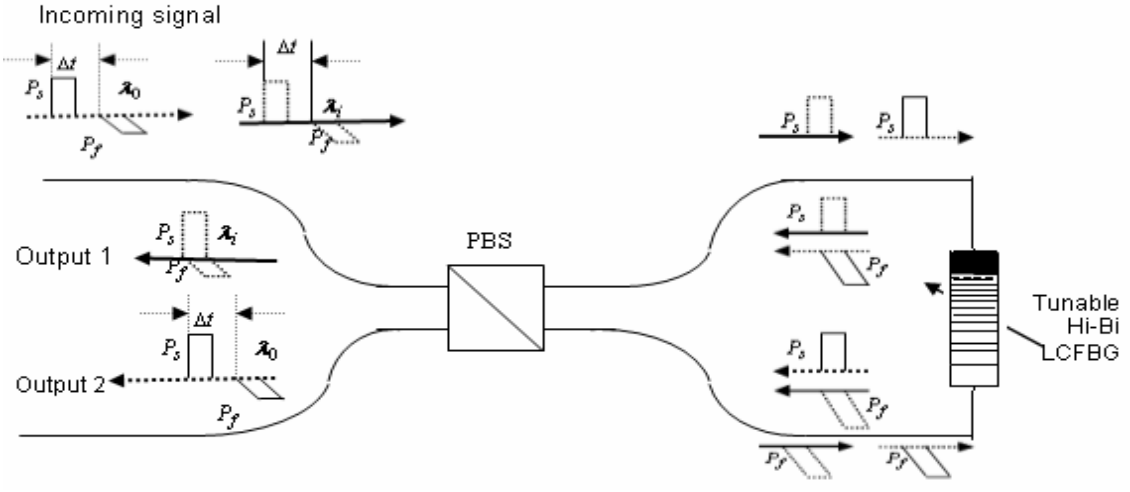


Fig.1: System model of the LCFBG based PMD compensator. The incoming signal has polarization components along both the fast- (P_f) and slow (P_s)axis. The LCFBG generates the required DGD for the Bragg reflected signal (λ_i), while it does not affect the signal (λ_0) outside the grating bandwidth

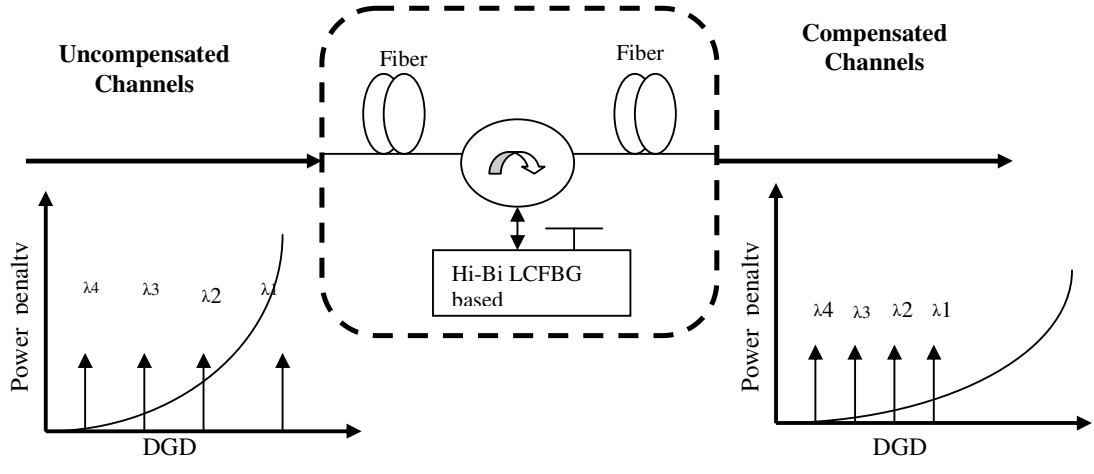


Fig.:2 Hi-Bi LCFBG based WDM-PMD compensation module

IV. Theoretical analysis

A. Characteristics of Hi-Bi LCFBG

Let us assume that a Bragg grating in a single mode fiber consists of a refractive index that is varied with the period, Λ and the modulation amplitude is added to the initial refractive index in the fiber. The perturbation to the effective refractive index n_{eff} of the guided mode(s) of interest given by,

$$\delta n(z) = \delta n_{eff}\left[1 + \nu(z)\cos(Kz + \phi(z))\right] \quad (1)$$

Where, $\delta n_{eff} = n_{av} - n_{co}$ is the average index change over one period, ν is the fringe visibility of the index change, $0 \le \nu \le 1$. $\cos(Kz + \phi(z))$ is the index perturbation, which has a constant spatial frequency with an additional position dependent phase variation $\varphi(z)$ that represents the change in periodicity.

Birefringence in optical fibers is defined as the difference in refractive index Δn between a pair of orthogonal modes (called slow- and fast modes) and results from the presence of circular asymmetries in the fiber section. The refractive index for both the slow- and fast-mode is defined as

$$n_{eff,s} = n_{eff} + \frac{\Delta n}{2}\ ; \qquad n_{eff,f} = n_{eff} - \frac{\Delta n}{2} \quad (2)$$

Where n_{eff} is the fiber effective index. To denote the refractive index of both modes from now on we will use notation $n_{eff,f(s)}$.

B. The effect of strain on Hi-Bi LCFBG

The penetration depth in reflection and the distance traversed in transmission in response to an applied axial strain or temperature changes, because there is a redistribution of the period as well a change in the refractive index due to photo-elastic or photo-thermal effect [11]. Both of these effects influence the Bragg condition (*i.e.,* $\lambda_B = 2n_{eff,f(s)}\Lambda$). Suppose at constant temperature, under the influence of axial strain (+ve) $\varepsilon(z)$, at the grating position *z*, the FBG will experience a physical elongation of grating period, Λ and a change of refractive index, $n_{eff,f(s)}$ due to the photo-elastic effect. The grating period can be written as,

$$\Lambda(z, \varepsilon(z)) = (\Lambda_0 + C_0 z)(1 + \varepsilon(z)) \quad (3)$$

Where, Λ_0 is the grating period at position, $z = 0$ without strain. Constant C_0 denotes the initial chirp of the grating and represented as, $C_0 = d\Lambda/dz\big|_{z=0}$. The induced change in the fiber index $dn_{eff,f(s)}(z)$ that is due to the photo-elastic effect is expressed as [12], $\frac{dn_{eff,f(s)}}{n_{eff,f(s)}} = -\rho_e\,\varepsilon(z)$, where we assumed the photo-elastic contributions into ρ_e, which is defined by $\rho_e = \frac{n_{eff,f(s)}}{2}\left[p_{12} - \mu(p_{11} + p_{12})\right]$ in terms of Pockel's coefficients p_{ij} and μ is the Poisson ratio. Together with the Bragg condition the resonance condition can be approximated and becomes dependent on local strain. Thus Bragg wavelength $\lambda_{f(s)}$ at grating position z becomes,

$$\lambda_{f(s)}(z) = 2n_{eff,f(s)}\left[(\Lambda_0 + C_0 z) + (\Lambda_0 + C_0 z)(1 - \rho_e)\varepsilon(z)\right] \quad (4)$$

From (4), it is straightforward to observe that the reflection wavelength shift of the grating is proportional to the applied strain. If strain gradient is added to a FBG, the Bragg wavelength varies linearly along the fiber length because of the change in the effective grating period.

C. Modeling of the fiber Bragg grating

Coupled-mode theory is a good tool for obtaining quantitative information about the diffraction efficiency and spectral dependence of fiber grating. Generally, Bragg reflection is the interaction between the optical signal and its medium. The wave formulation for forward and backward wave amplitudes *F(z)* and *B(z)* can be expressed as,

$$\begin{aligned} \frac{dF^+(z)}{dz} &= -i\hat{\sigma}_{f(s)}\,F^+(z) + i\kappa_{f(s)}(z)\,B^+(z) \\ \frac{dB^+(z)}{dz} &= -i\hat{\sigma}_{f(s)}\,B^+(z) - i\kappa^*_{f(s)}(z)\,F^+(z) \end{aligned} \quad (5)$$

Where, $F^+(z) = F(z)\,e^{(i\delta z - \varphi/2)}$, $B^+(z) = B(z)\,e^{(-i\delta z + \varphi/2)}$; $F^+(z)$ (reference) represents the forward propagating mode, $B^+(z)$ (signal) is the identical backward (counter) propagating mode, $\hat{\sigma}_{f(s)}$ and $\kappa_{f(s)}$ are the general dc (period averaged) self-coupling and ac coupling coefficients respectively and can be expressed as,

$$\begin{aligned} \hat{\sigma}_{f(s)} &= \delta_{f(s)} + \sigma_{f(s)} - \frac{1}{2}\frac{d\varphi(z)}{dz} \\ &= \delta_{f(s)} + \frac{2\pi}{\lambda_{f(s)}}\delta n_{eff,f(s)} - \frac{1}{2}\frac{d\phi(z)}{dz} \end{aligned} \quad (6)$$

$$\kappa_{f(s)} = \frac{\pi}{\lambda_{f(s)}}\delta n_{eff,f(s)}\nu \quad (7)$$

$$\delta_{f(s)} = \beta - \beta_B = 2\pi\, n_{eff,f(s)}\left(\frac{1}{\lambda_{f(s)}} - \frac{1}{\lambda_B}\right) \quad (8)$$

Where $\delta_{f(s)}$ is the detuning (which is independent of z for all gratings) from the Bragg wavelength λ_B related to period Λ_0; $\delta n_{eff,f(s)}$ is the dc index spatially averaged over a grating period. $\sigma_{f(s)}$ is the dc coupling coefficient and The derivative $1/2\left(d\varphi(z)/dz\right)$ describes the possible chirp of the grating.

The coupled mode equations for the forward and the backward propagating modes can be solved using appropriate boundary conditions. The boundary conditions assume a forward propagating mode with $F^+(0)=1$ and that the backward propagating mode, at the end of the grating, will be zero, $B^+(L_g)=0$ as there are no perturbing beyond the end of the grating.

The solution of the coupled mode equation (5) can be best expressed by,

$$\begin{bmatrix} F^+(0)=1 \\ B^+(0) \end{bmatrix} = [T] \begin{bmatrix} F^+(L_g) \\ B^+(L_g)=0 \end{bmatrix} \quad (9)$$

where T is the transfer matrix and given by,

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \quad (10)$$

Where, $T_{11} = \cosh(\gamma L_g) - i\frac{\hat{\sigma}_{f(s)}}{\gamma}\sinh(\gamma L_g)$ (11)

$$T_{12} = -i\frac{\kappa_{f(s)}}{\gamma}\sinh(\gamma L_g) \quad (12)$$

$$T_{22} = \cosh(\gamma L_g) + i\frac{\hat{\sigma}_{f(s)}}{\gamma}\sinh(\gamma L_g) \quad (13)$$

$$T_{21} = i\frac{\kappa_{f(s)}}{\gamma}\sinh(\gamma L_g) \quad (14)$$

γ is the parameter relating to coupling coefficient as $\gamma = \sqrt{\kappa_{f(s)}^2 - \hat{\sigma}_{f(s)}^2}$.

Reflectivity is the percentage of light reflected at the Bragg wavelength, the wavelength outside of the reflected bandwidth is transmitted without disturbance. For the case of uniform grating, the complex amplitude reflection coefficient at the beginning of the FBG at $z = 0$ is defined as,

$$\rho_{f(s)}(\lambda) = \frac{B^+(\hat{\sigma}_{f(s)})}{F^+(\hat{\sigma}_{f(s)})} = \frac{T_{12}}{T_{11}} \quad (15)$$

$$\rho_{f(s)}(\lambda) = -\frac{\kappa_{f(s)}\sinh(\gamma L_g)}{\hat{\sigma}_{f(s)}\sinh(\gamma L_g) + i\gamma\cosh(\gamma L_g)} \quad (16)$$

And the reflectivity *or* the power reflection coefficient, $R_{f(s)}(\lambda)$ is given by,

$$R_{f(s)}(\lambda) = \left|\rho_{f(s)}(\lambda)\right|^2 = \frac{\sinh^2(\gamma L_g)}{\cosh^2(\gamma L_g) - \hat{\sigma}_{f(s)}^2/\kappa_{f(s)}^2} \quad (17)$$

The group delay of the reflected light can be determined from the phase of the complex amplitude reflection coefficient $\rho_{f(s)}(\lambda)$. If we denote $\varphi_\rho \equiv phase(\rho_{f(s)}(\lambda))$, then the time delay τ_R for light reflected off of a grating is,

$$\tau_{R,f(s)} = \frac{d\varphi_\rho}{d\omega} = -\frac{\lambda_{f(s)}^2}{2\pi c}\frac{d\varphi_\rho}{d\lambda_{f(s)}} \quad (18)$$

Where, $\tau_{R,f(s)}$ is usually given in picoseconds, ω is the angular frequency and c is the velocity of light. Thus an optical wave traveling through a medium of length L and refractive index n will undergo a phase change, $\varphi_\rho = (2\pi n_{eff,f(s)}L_g)/\lambda$; The derivative of the phase with respect to wavelength is an indication of the delay experienced by the wavelength component of the reflected light;

$$\frac{d\varphi_\rho}{d\lambda_{f(s)}} = -\frac{2\pi n_{eff,f(s)}L_g}{\lambda_{f(s)}^2} \quad (19)$$

The reflected spectrum is characterized by a main peak at the wavelength defined as, $\lambda_{\max,f(s)} = 2(n_{eff,f(s)} + \delta n_{eff,f(s)})\Lambda(z)$. Thus, with a large refractive-index difference Δn between the fast and slow polarization axes; that results in a shift in Bragg wavelength $\Delta\lambda_B = \left|\lambda_{\max,s} - \lambda_{\max,f}\right| = 2\Delta n\,\Lambda_0$ at the same location for two polarizations. Therefore, the Hi-Bi linearly chirped FBG can be seen as two different chirped FBGs because of the birefringence. The position difference of the reflection produces a DGD:

$$DGD(\lambda) = \left|\tau_{R,f}(\lambda) - \tau_{R,s}(\lambda)\right| \quad (20)$$

V. Results and Discussion

Following the analytical formulations, at first the performance of a 2 cm long LCFBG is investigated in terms of relative group delay, DGD, stretching to determine the tuning range and maximum amount of PMD that can be compensated. To demonstrate its capability simulation is carried out for a 4-channel WDM system using this LCFBG compensator at unstretched condition. The different parameters used in the simulation are shown in Table 1. To avoid the power loss due to the leakage of the grating, the reflectivity was greater than 99% within the grating bandwidth is assumed.

Fig.3 shows the relative time delay for the fast- and slow axis polarization reflected signals inputting from the long- and short wavelength ports of the unstressed FBG respectively as a function of wavelength. From the plots, it is seen that the group delay of the grating reveals good linearity and the delay curves for the two polarization axes are shifted by 0.25 nm at wavelength 1550.5 nm relative to each other due to the high birefringence of the fiber. Since the group delay of the grating is shifted to longer or shorter wavelength by stretching or compressing the grating, the DGD between fast- and slow axes are increased or decreased for a given wavelength. The linear variation makes it easy for the adjustment of the DGD to any given value.

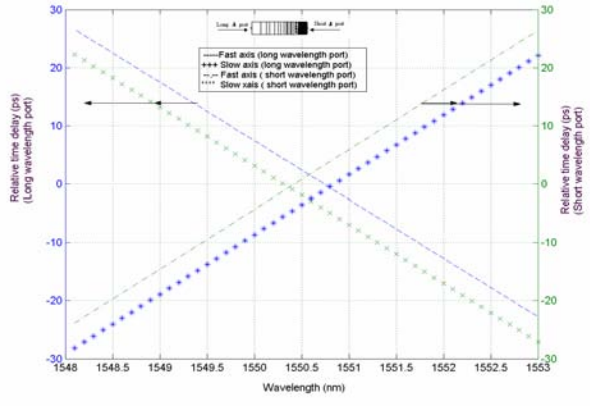


Fig.3: Reflected relative time delay for the fast- and slow axis of polarization

The reflection spectrum of the grating is shown in Fig.4. It is observed that the bandwidth of the grating is 1549 nm to 1551.02 nm at unstretched condition. When the grating is stretched, the amplitude and the group delay spectrum shifts to longer wavelength. Note that due to

tuning, though the passband shifts to shorter or longer wavelength but the polarization of the reflected signal remains same. S. Lee *et.al,* [8] experimentally achieved reflection wavelength ranging from 1547.2 nm to 1550.5 nm at unstressed condition and a tuning range of 2.32 nm for a 5 cm long nonlinearly chirped grating. The simulation work in Ref. 7 obtained a bandwidth ranging from 1548.2 nm to1550.5 nm and 2.1 nm of tuning range using a 1cm LCFBG.

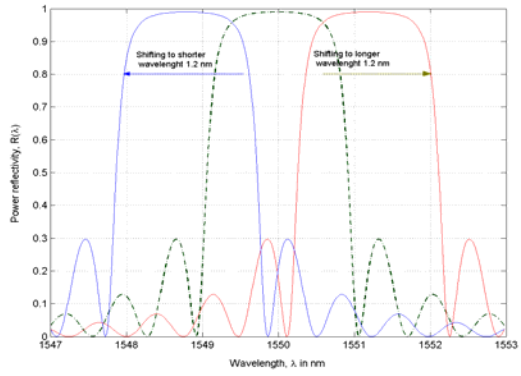


Fig.4: Reflection spectrum for the LCFBG under stretching and compression. Wavelength tuning shifts the passband to longer- or shorter wavelength regime without changing the shape of the spectrum

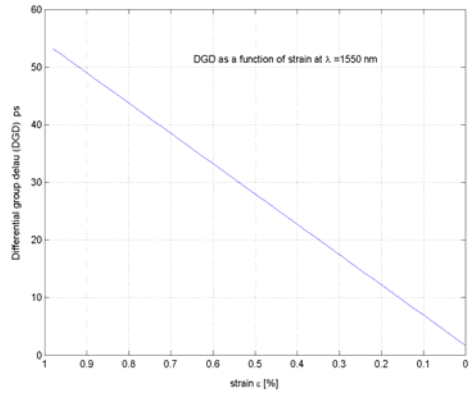


Fig.4: Variation of the DGD between two polarizations as the LCFBG stretched (the stretching ratio is the change in the length of the fiber grating divide by its original length ($\varepsilon = \Delta L/L$)

Fig. 5 shows the change in the differential time delay (ps) as a function of the grating stretching ratio. By changing the strain ε from 0% to 0.2% the DGD can be continuously adjusted from 3 ps to 58 ps. As mentioned earlier, all light outside the reflection bandwidth λ_0 can be directed to the output port without adding any PMD. For this system we have a residual DGD of 3 ps. It is also observed that a maximum DGD is varied over a 55 ps range by a 0.2 % stretch of the grating at 1552.2 nm. Ref. [15] experimentally demonstrates a variable polarization delay line exhibits an adjustable amount of PMD within its operating bandwidth that the DGD can be continuously adjusted from 0 to 80 ps (from 1330.4 nm to 1531 nm wavelength range) under mechanical strain applied by a piezo driven translator.

From the above results we found that stretching the LCFBG adjust the differential time delay without altering the polarization and thus this device has the potential for compensation variable PMD for a long-distance high-speed optical WDM transmission system. Through simulation, we investigate the impact of PMD on the WDM system in terms eye diagram without applying any stress on the LCFBG. At the receiving end we demultiplexed all the 4-channel and monitor their eye diagram without and with applying the PMD compensation scheme. A typical measurement of the eye diagram of the accumulated DGD for 4-channels before and after compensation is shown in Fig. 6. At 10 Gb/s data rate each channel produces (200 km span) about 28 ps of DGD and our designed LCFBG compensator can compensate only 25 ps DGD without any external strain. We observe that channel 4 has been improved significantly after compensation without impacting the other channels and about 6 ps DGD of each channel remain uncompensated which is also within tolerance level or power margin of the system.

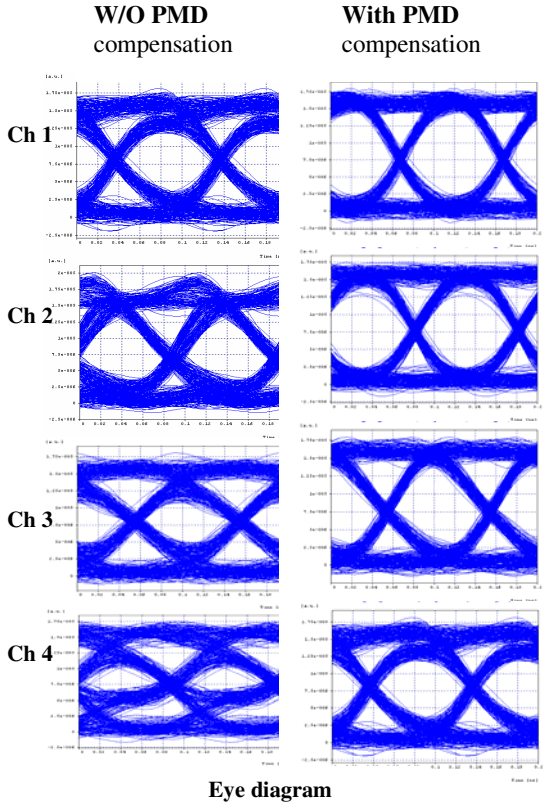


Fig.:6 Eye diagram for a 4-channel (each channel at 10 Gbps, NRZ modulation) with and without WDM-PMD compensation. The output signal is received after 400 km transmission

Finally, we calculate the average eye-opening penalty (EOP) due to PMD for the 4-channel WDM system. Here, we define the parameter of the eye-opening penalty (EOP) s,

$$EOP = -20\log\left(\frac{B}{B_0}\right) \qquad (21)$$

Where, B is the eye opening without PMD effect and B_0 is the eye-opening with PMD. Fig.7 shows a comparison of EOP penalty between our 4-channel simulated transmission system and with single change LCFBG based compensation system as a function of DGD. From the curve we found that about 2.6 dB (reduced from 7.10 dB to 4.5 dB) of EOP reduction is obtained at 40 ps by the compensation scheme. The performance of the single channel is 0.55 dB better for mean DGD > 40 ps and this happens due to inter-channel crosstalk.

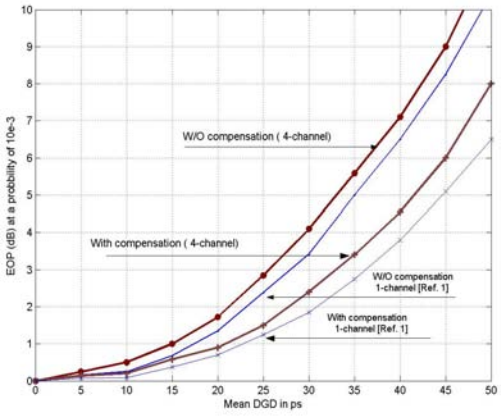


Fig.: 7: Eye opening penalty (EOP) at a probability of 10^{-3} in the dependence of the mean DGD for 4-channel and single transmission system

Parameters/description	**values**
Fiber loss (dB/km)	0.25
Dispersion (ps/km-nm)	0
Effective core are (μm^2)	80
Fiber length (2-span of 200 km)	400
PMD co-efficient (ps/$\sqrt{}$km)	2
Birefringence of the Hi-Bi LCFGB	$3.77x10^{-4}$
Bit rate of each channel (Gb/s)	10
No. of channels	4
Comp. BW w/o strain (nm)	2
Channel spacing (nm)	0.4
Working wavelength range (nm)	1549 nm -1551 nm
Receiver sensitivity (at 10^{-9})	-24
Transmitter power (dBm) /ch.	-10
Booster EDFA output (dBm)	9

Table 1: Simulation parameters of Hi-Bi LCFBG based PMD-WDM compensation scheme

VI. Conclusion

The performance analysis of a novel PMD using a 2-cm long LCFBG compensator and 4-channel PMD-WDM simulation is carried out. The various properties of reflectivity, relative group delay and DGD of the Hi-Bi LCFBG have been discussed in detail. The PMD compensation capability of the device mainly determined by the characteristics of the Hi-Bi LCFBG and improved capability can be achieved by adopting higher chirp ratio phase mask to fabricate the Hi-Bi grating. The performance of the simulated compensation scheme achieved a significant improvement in quality of eye pattern of the received signal for a link length of 400 km at a bit rate of 10 Gb/s and it compensates about 50 ps of DGD per channel. It is also observed that the amount eye power penalty for a single channel is about 0.55 dB less than that of 4-channel WDM system (at 40 ps of DGD) and that happens due to absence of channel crosstalk in a single channel system. It is an all fiber PMD compensation solution, which is inexpensive, compact, absence of nonlinear effects and feasible for continuously adjustable DGD.

References

[1] M. Wang, T. Li and S. Jian, "Tunable PMD compensator based on high-birefringence linearly chirped FBG with cantilever beam", *Optics Express,* vol. 11, no. 19, pp. 2354-2363, 2003.

[2] B. W. Hakki, "Polarization mode dispersion by phase diversity detection", *IEEE Photon. Technol. Lett*., vol. 9, pp. 121-123, 1997.

[3] F. Heismann, D. A. Fishman and D. L. Wilson, "Automatic compensation of first-order polarization mode dispersion in a 10 Gb/s transmission system", in *Proc. ECOC'98*, pp.529-530, 1998.

[4] L. Yan, X. S. Yao, M. C. Hauer, A. E. Willner, "Practical solutions to polarization-mode-dispersion emulation and compensation", *J. of lightwave Technol.*, vo. 24, no. 11, pp. 3992-4005, 2006

[5] R. Kashyap, "Fiber Bragg Grating, Chapter 7: Chirped fiber Bragg gratings", *Academic Press*, San Diego, USA (1999)

[6] "Fiber Bragg gratings," USA: Furukawa Electric North America, Lightwave Inc., 2005

[7] M. Wang, T. Li and S. Jian, "Tunable PMD compensator based on high-birefringence linearly chirped FBG with cantilever beam", *Optics Express*, vol. 11, no. 19, pp.2354-2362, 2003

[8] S. Lee, R. Khosravani, J. Peng, A. E. Wilner, V. Grubsky *et al* "Adjustable compensation of polarization mode dispersion using a high-birefringence nonlinearly chirped fiber Bragg grating," *IEEE photonics Technology Letters*, vol. 11, no. 10, pp. 1277-1279,1999

[9] Z. Pan, Y. Xie, S. Lee, A. E Willner, "Tunable compensation for polarization-mode dispersion using a birefringent nonlinearly-chirped Bragg grating in a dual-pass configuration," *U. S. Patent* 6,400,869 B2, 2002.

[10] X. Kun, F. Jia, X. Chen, Mao Jin, M. Chen, X. Li and S. Xie, "A novel adjustable PMD compensation using sampled Bragg gratings with uniform grating period," *Optics Communications,* vol. 202, no. 4-6, pp. 297-302, 2002

[11] A. D. Kersey, M. A. Devis, H. J. Patrick et al "Fiber Grating Sensors," *Journal of Lightwave Technology*, vol.15, pp. 1442-1461, 1997.

[12] D. Roylance, mechanics of materials, Wiley & sons, New York, 1996

[13] W. Du, X. Tao and H. Tam, "Fiber Bragg grating cavity sensor for simultaneous measurement of strain and temperature", *IEEE Photonics Technology Letters*, vol. 11, pp. 105-107, 1999.

[14] Ho Sze Phing, Jalil Ali, Rosley Abdur Rhaman *et al* "Fiber Bragg grating modeling, simulation and characteristics with different grating length," *Journal of fundamental sciences*, UTM, Malaysia.

[15] H. Rosenfeldt, Ch. Knothe, E. Brinkmeyer, "Component for Optical PMD-Compensation in a WDM Environment" *Proc. Europ. Conf. on Opt. Comm. (ECOC)*, Germany, vol. I(3.4.1), pp. 135-136, 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Synthesis of CMOS OTA based Communication Circuit

Arabinda Roy, Deptt. of E.T.C., Sekhar Mandal, Deptt. of C.S.T.,
Baidya Nath Ray, Deptt. of E.T.C.
Bengal Engineering and Science University,
Shibpur, Howrah, India
Prasanta Ghosh
Deptt. of E.E.C.S. Syracuse University,U.S.A.,

Keywords: Operational Transconductance Amplifier (OTA), Radio Frequency (RF), Phase Lock Loop (PLL), Adaptive Delta Modulator (ADM), Compander.

Abstract

In this paper a generalized design methodology for Operational Transconductance Amplifier (OTA) based current mode Radio Frequency (RF) communication circuit is proposed. Three novel communication circuits, Phase Lock Loop (PLL), Adaptive Delta Modulator (ADM) and data compressor have been simulated. In the first implementation, direct frequency modulation is used to realize PLL, while the second gives coded pulse modulation system, which employs sampling, quantizing and coding to convert analog waveforms into digital signals. Finally, the third circuit implements data compression in digital communication system. Performance of the circuits realized with OTAs has been demonstrated through SPICE simulation.

1 Introduction

This paper introduces a generalized synthesis procedure for application of OTA's in three important Radio Frequency (RF) communication circuits. In general, all types of communication systems can be classified into three categories, (i) Modulator, to modulate the message signal (ii) Data Converter to convert analog signal into digital signal and (iii) Data Compressor; a nonlinear device that compresses the signal amplitude, particularly in voice transmission [1]. This paper proposes a methodology to design these three categories of circuits, using circuit components built with OTA as the basic building block. The simulation results of the circuit performance have also been reported. The motivation to undertake this research stems from the following considerations. (i) Growing popularity of analog and mixed signal ICs has provided the impetus to explore innovative designs implemented with CMOS VLSI technology and (ii) Reduction of design turn around time. Further, OTA is an excellent current mode device to realize high frequency resistor-less analog designs. With this backdrop, we have embarked on design of OTA based RF communication circuits. A survey of the literature dealing with OTA based designs depicts following picture. A number of researchers have employed OTA as the basic building block in the design of non-linear networks. Edgar Sanchez-Sinencio et al. [2] have reported synthesis of a number of nonlinear circuits with OTA network. The authors have reported two synthesis techniques viz. rational approximation functions and piece-wise-linear approximation technique. Four-quadrant multiplier and a phase shifter are the two important analog functions used in data communication system. Bang - Sup - Song [3] has reported the recent development in respect of synthesis of these functions with OTA networks. It is well accepted that the design of the radio frequency (RF) section in a communication integrated circuit (IC) is a challenging problem.

Modulator circuit for analog and digital communication system (AM, ASK, FM, and FSK) has been synthesized and realized with single output OTA by Ray et al. [4]. Similar type of digital communication circuit (i.e., ASK/FSK/PSK/QAM) using multiple output OTA and a number of digitally controlled switches is proposed by Taher et al. [5]. Also, Hietala proposes PSK/GMSK polar transreceiver [6], Rahim et al. present software defined wireless receiver [7]. All those work propose novel design methodology of a particular category of communication circuit. However, no general methodology has so far been reported to realize three categories (i.e., modulation, data conversion and compression) of communication circuits (operating at radio frequencies) with OTA or with other devices. This paper bridges this gap and reports a generalized scheme for realizing the communication (analog and digital) circuits with an array of OTAs.

In the above background the section II introduces the OTA based basic building blocks used for design of the above mentioned communication circuits. Section 2 and 3 reports the design methodology of OTA based radio frequency non linear communication circuits. Finally, the simulation results are highlighted in section 4 to

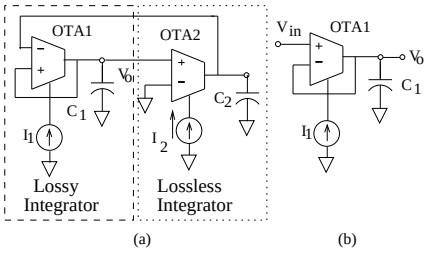


Figure 1: OTA Based (a) Oscillator and (b) Integrator.

verify the performance of the designed circuits.

2 Design of Non Linear Circuit

This section deals with design methodology of non linear analog circuit used in communication system. We start with the description of the OTA based basic building blocks.

2.1 OTA Based Basic Building Blocks

In this section we report the methodology for synthesis of non linear analog circuits. The synthesis procedure is developed using a cell library. For the sake of demonstration of our synthesis procedure we have assumed multiplier, oscillator, phase detector and integrator (a first order filter) as the basic cells available in the library.

Each of these cells is designed with a network of OTAs. OTA is an active current mode building block similar to voltage mode counterpart of an operational amplifier. OTA is voltage controlled current source with infinite input and output impedance. Output current of an OTA is given by, $I_o = g_m(V_1 - V_2)$, where, I_o is the output current, V_1 is input voltage at non-inverting input and V_2 is input voltage at inverting input. Main problems of OTA are to maintain the linearity of transconductance at high frequencies, gain-bandwidth product, slew rate and dynamic range. In the recent past a number of researchers have concentrated on those problems and proposed different techniques [8, 9, 10, 11, 12].

The OTA based oscillator and integrator are shown in Fig. 1, while the circuit for multiplier and phase detector are depicted in Fig. 2.

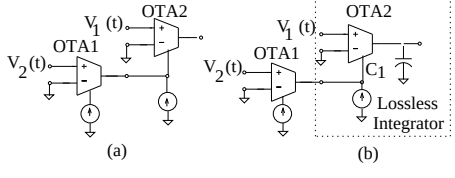


Figure 2: OTA Based (a) Multiplier and (b) Phase Detector.

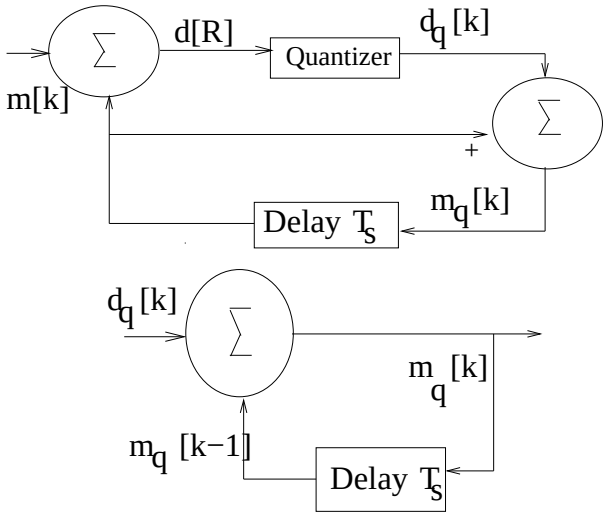


Figure 3: Block diagram of DM.

The methodology is demonstrated through design of the following non linear circuits.

2.2 Delta Modulator

Delta modulation (DM) may be viewed as two-level (1-bit) quantizer. A block diagram of DM encoder is shown in Fig. 3. From this figure, we obtain

$$m_q[k] = m_q[k-1] + d_q[k] \tag{1}$$

Hence,

$$m_q[k-1] = m_q[k-2] + d_q[k-1] \tag{2}$$

Substitution of Eqn. 1 in Eqn. 2 yields

$$m_q[k] = m_q[k-2] + d_q[k] + d_q[k-1] \tag{3}$$

Proceeding iteratively and assuming zero initial condition, (i.e., $m_q[-1] = 0$ and $d_q[0] = 0$) we arrive at

$$m_q[k] = \sum_{m=0}^{k} d_q[m] \tag{4}$$

Equation 4 represents an accumulator (adder). If the output $d_q[k]$ is represented by impulses, then the accumulator may be realized by an integrator. OTA network of linear delta modulator is shown in Fig. 4 . In that figure (Fig. 4) the portion of the circuit encircled by dashed lines represent OTA realization for the linear delta modulator. Along with that portion, circuit outside the dashed line represents continuously variable slope delta modulator (discussed in the next section). Subsequent discussions analyze linear delta modulator.

OTA1 and OTA2 serve the purpose of a difference amplifier leading to a two level quantizer. The output of the quantizer assumes the supply voltage level +V and

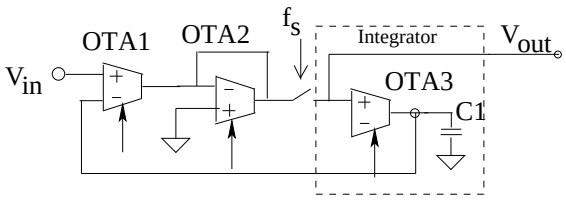


Figure 4: OTA realization of Linear DM.

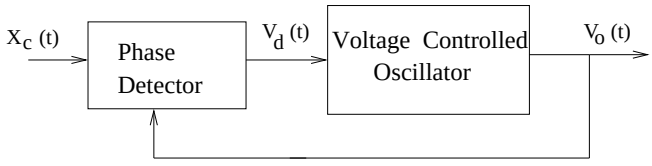


Figure 5: Block diagram of Phase Lock Loop.

-V depending on whether the input overrides the predicted signal or not. The output of the quantizer is sampled at a rate $f_s = 2kf_m$ (by the switch S_1 in Fig. 4), where f_m is the frequency of the input sine wave and k is the oversampling factor. The sampled quantizer output is fed next to the input of an integrator implemented by the OTA3 and capacitor C_1. The input to the integrator is given by the following equation.

$$s_i(t) = V \sum a_n p(t - nT_s) \tag{5}$$

where 'V' is the maximum output level of the quantizer and a_n is either +1 or -1, and $T_s = 1/f_s$. The pulse shape p(t) is given by -

$$p(t) = u(t) - u(t - \frac{T_s}{2}) \tag{6}$$

where u(t) is Heaviside step function. Assuming ideal OTA, the output of the integrator is given by

$$s_o(t) = \frac{g_{m3}}{C_1} V \sum a_n[\rho(t - nT_s) - \rho(t - nT_s - \frac{T_s}{2})] \tag{7}$$

where $\rho(t)$ is a ramp function and $T_s = 1/f_s$.
The step size δ of the staircase waveform is given by -

$$\delta = \frac{g_{m3} T_s}{2C_1} \tag{8}$$

Step size of the DM can be optimized by varying the bias current of the OTA3.

2.3 Continuously Variable Slope Delta Modulator

A linear delta modulator requires a large over sampling factor 'k' (from 8 to 16) for proper operation. By operating the modulator in an adaptive mode which increases the step size in the overload region and decreases it in the granular region the oversampling factor can be reduced down to 4 to 8 thus decreasing the output bit rate. A continuously variable slope delta modulator adapts the step size in a continuous fashion both in the granular and overload regions. Fig. 4 (including linear delta modulator encircled by dashed lines) represents continuously variable slope delta modulator. The front end of the circuit (encircled by dashed line) is an integrator followed by a rectifier and a low pass filter (LPF). When the delta modulator is overloaded the input to the integrator will be a sequence of all positive or all negative pulses. The magnitude of the output of the integrator in this case will be higher. The integrator output is rectified and filtered by LPF. The output of the low pass filter increases the transconductance g_{m3} of the integrator of the linear delta modulator block. As a result step size (δ) increases because step size (δ) directly varies with g_{m3} (equation 8). When the modulator hunts around a flat portion of the input signal the output of the switch will be alternate in polarity (i.e its magnitude changes between positive and negative values alternatively). As a result, transconductance of OTA3 (g_{m3}) will assume a low value. Consequently, from equation (8) it can be concluded that step size (δ) decreases. Thus Fig. 4 realizes continuously variable slope delta modulator.

2.4 Phase Lock Loop

A Phase lock loop is basically an oscillator whose frequency is locked onto some frequency component of an input signal $x_c(t)$. Fig. 5 depicts the block diagram of PLL.

The linear model for phase detector (PD) can be written as follows.

$$v_d = K_d.\theta_e + V_{off} \tag{9}$$

Where, K_d is the PD gain, θ_e is the phase error of the VCO output relative to the input signal, and V_{off} is the offset voltage or free running voltage. Next, we have discussed the realization procedure of the phase detector.
Phase error θ_e between the two signals, $x_c(t) = cos(\omega_c t + \theta_1)$ and $x_0(t) = cos(\omega_0 t + \theta_2)$ can be written as,

$$\theta_e = \theta_1 - \theta_2 = cos^{-1}V_1 - cos^{-1}V_2 + \omega_0 t - \omega_c t \tag{10}$$

The above equation can be expressed as

$$cos(\theta_e) = V_1.V_2 + sin(\omega_c t + \theta_1).sin(\omega_0 t + \theta_2) - cos(\omega_c t - \omega_0 t) \tag{11}$$

Substitution of equation (11) in (9) produces

$$V_d(t) = K_d[cos^{-1}[V_1V_2 + sin(\omega_c t + \theta_1)sin(\omega_0 t + \theta_2) - cos(\omega_c t - \omega_0 t)]] + V_{off} \tag{12}$$

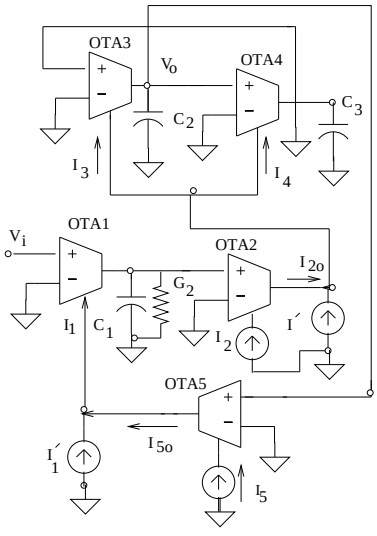


Figure 6: OTA realization of Phase Lock Loop.

If it is assumed $K_d\theta_e = \bar{K}_d cos\theta_e$ then it can be written

$$V_d(t) = \bar{K}_d[V_1V_2 + sin(\omega_c t + \theta_1)sin(\omega_0 t + \theta_2) - cos(\omega_c t + \omega_0 t)] + V_{off} \quad (13)$$

Equation (13) prevails that, with proper selection of K_d and V_{off} of equation (9), behavior of phase detector can be modelled with multiplier and integrator. OTA realization for multiplier and phase detector have been discussed earlier (Fig. 1 and 2). Figure 6 depicts OTA realization of PLL connecting different blocks. VCO is realized by OTA3, OTA4, C_3 and C_4. OTA1 and OTA5 form phase detector and OTA2 realizes voltage to current converter. State equation is given by

$$\begin{bmatrix} \dot{V}_{C2} \\ \dot{V}_{C3} \end{bmatrix} = \begin{bmatrix} 0 & \frac{g_{m3}}{C_2} \\ \frac{g_{m4}}{C_3} & 0 \end{bmatrix} \cdot \begin{bmatrix} V_{C2} \\ V_{C3} \end{bmatrix} \quad (14)$$

The radian frequency is given by

$$\omega_0 = \sqrt{\frac{g_{m3}g_{m4}}{C_2C_3}} = k\sqrt{\frac{I_3I_4}{C_2C_3}} \quad (15)$$

Where k is the transconductance gain factor i.e., $g_m = kI$, when I = bias current. If $I_3 = I_4 = I$, then VCO can be tuned by I' and I_{20}. Thus $I = I_3 = I_4 = \frac{I'+I_{20}}{2}$. VCO frequency can be written as, $\frac{k(I'+I_{20})}{2\sqrt{C_2C_3}}$. The input signal for lock in detection is to be fed as the input to the OTA1, whose bias current I_1 is controlled by the VCO output V_0 through OTA5. Output of OTA1 can be written as, $-\frac{g_{m1}V_i}{G_2+sC_1}$.

Where G_2 represents the input conductance of the OTA2.

Since g_{m1} (transconductance of OTA1) is controlled by I_{50} the output current of OTA5, then small signal incremental analysis yields, $g_{m1} = kg_{m5}, V_0 = k^2I_5V_0$ (as $g_{m5} = kI_5$). k is the transconductance gain related to bias current .

Again,

$$V_{C1} = -\frac{g_{m1}V_i}{G_2 + sC_1} = -\frac{k^2I_5V_iV_0}{G_2 + sC_1} \quad (16)$$

Assume that the input is locked to the VCO. Let, $V_i = Asin\omega_0 t$ and $V_0 = Bcos(\omega_0 t + \phi)$ then V_{C1} can be written as

$$V_{C1} = -\frac{k^2I_5AB}{2(G_2 + sC_1)}[Sin(2\omega_0 t + \phi) - Sin\phi] \quad (17)$$

If the double frequency term is rejected by the low pass filter formed by C_1 and G_2, then the low frequency output voltage across C_1 is given by,

$$V_{C1} \simeq \frac{k^2I_5AB}{2sC_1}Sin\phi = k_dSin\phi \quad (18)$$

[As G_2 is very low and $k_d = \frac{k^2I_5AB}{(2sC_1)}$.]

Output of OTA1 i.e., V_{C1} controls the output current of OTA2 i.e., I_{20} which is in appropriate magnitude and polarity to drive the VCO frequency towards the input signal frequency. Let ω_{fr} denotes the free running frequency of the VCO with $I_{20} = 0$ and $I' = I_{fr}$. Thus,

$$\omega_{fr} = \frac{kI_{fr}}{2\sqrt{C_2C_3}} \quad (19)$$

When I_{20} is nonzero it is to be noted that $I_3 = I_4 = \frac{I_{fr}}{2} \mp \frac{I_{20}}{2}$.

$$g_{m3} = k(\frac{I_{fr}}{2} \mp \frac{I_{20}}{2} = g_{m0} + -\Delta g_m \quad (20)$$

where $g_{m0} = \frac{kI_{fr}}{2}$ and $\Delta g_m = \frac{kI_{20}}{2}$. The VCO instantaneous frequency,

$$\omega_0 = \omega_{fr} + \frac{g_{m3}}{\sqrt{C_2C_3}} = \omega_{fr}[1 \pm \frac{\Delta g_m}{g_{m0}}] \quad (21)$$

Again I_{20} can be written as
$I_{20} = g_{m2}\ V_{C1(dc)} = kI_2k_d\ Sin\ \phi$. Thus $\omega_0 = \omega_{fr}(1 \pm \frac{kI_2}{2I_{fr}})k_dSin\phi$. Where $\frac{\delta g_m}{g_{m0}} = \frac{kI_2k_dSin\phi}{I_{fr}}$ and Again. $\omega_{fr}\frac{\Delta g_m}{g_{m0}}$ is the capture range of the phase detector circuit.

2.5 Compressor

For transmission of speech signal using waveform coding techniques the same quantizer has to accommodate input signals with widely varying power level. The range of voltages covered by voice signals, from loud talk to weak talk, is in the order of 1000 to 1. For acceptable voice transmission signal-to-quantization noise ratio should remain essentially constant for wide range

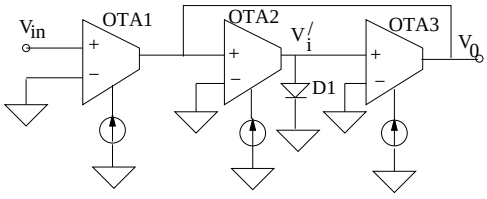


Figure 7: OTA realization of Compressor.

of input power levels. If the step size of the quantizer increases non uniformly then the dynamic range of the quantizer is improved. By using a device called compressor the desired form of the nonuniform quantization can be achieved. Characteristic of a compressor can be mathematically represented by logarithmic function [1]. Figure 7 shows OTA realization of compressor.

Analysis of OTA-Compressor circuit is as follows.

At the output node of OTA2 of that circuit,

$$g_{m2}V_0 = (e^{\frac{qv_i'}{KT}} - 1)I_s \tag{22}$$

and

$$e^{\frac{qv_i'}{KT}} = \frac{1+g_{m2}v_0}{I_s}, v_i' = 26(10^{-3})ln[\frac{1+g_{m2}v_0}{I_s}] \tag{23}$$

The above equations are satisfied at room temperature. At the output of the OTA1 it can be written,

$$g_{m1}v_i = g_{m3}v_i' = g_{m3}[26.(10^{-3})ln(\frac{1+g_{m2}v_o}{I_s})] \tag{24}$$

If, it is assumed, $g_{m1} = g_{m3}$, then v_o can be written as,

$$v_o = (\frac{I_s}{g_{m2}})(e^{\frac{10^3.v_i}{26}}) - \frac{1}{g_{m2}} \simeq (\frac{I_s}{g_{m2}})(e^{\frac{10^3.v_i}{26}}) \tag{25}$$

which approximates a compressor.

3 Simulation Result

All the circuits discussed above have been simulated with the Microsim PSPICE CAD tool. A simple CMOS linear OTA with differential input has been proposed by Szczepanski [13]. Such an OTA with g_m varying 70 to 120 μA/V is employed in our design application. Figure 8 shows the input and output signal of the delta modulator. Frequency of the input signal is 10 MHz and step size is 250 mv. Step size can be tuned by varying OTA3 and C_1 in Fig. 4. The output of adaptive delta modulator is depicted in Fig. 9. Like delta modulator the input signal frequency is 10 MHZ and the sampling frequency is 90 MHZ. It is clear from Fig. 9 that step size varies from 0.4 volt to 0.9 volt. From the simulated result of the PLL it has been observed that the capture

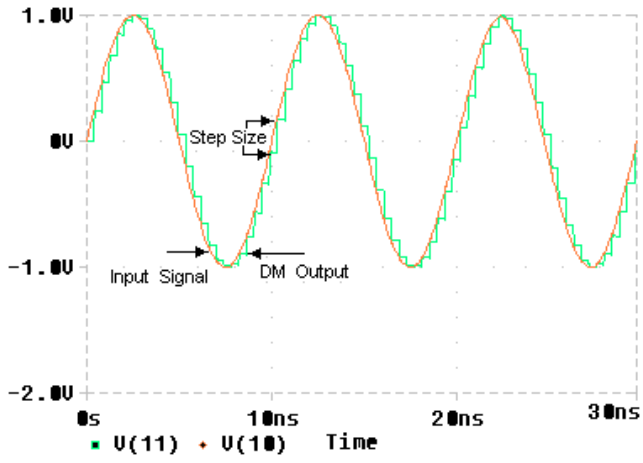


Figure 8: Output of Delta Modulator. Frequency of the Input Signal = 10 MHZ and Step Size = 250mv.

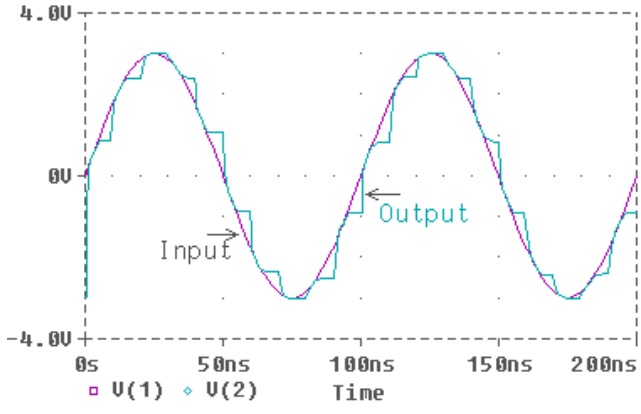


Figure 9: Output of Adaptive Delta Modulator.

range and lock range can be tuned by varying g_{m2} of Fig. 6. Lock range varies from 200 KHz to 300 KHz for the variation of g_{m2} from 70 μA/V to 120 μA/V. For that variation of g_{m2} capture range is also varied from 200 KHz to 300 KHz. Fig. 10 depicts the simulated input-output characteristic of the compressor. A ramp signal which varies from 0V to 4V during the time interval 0 to 2ms is applied at the input of the compressor. Output characteristic curves for three different values of g_{m2}, for 80, 90 and 100 μA/V, are shown. From Eqn. 27 and Fig. 10, it can be shown that the simulated result validates the theory of the compressor discussed in Section 3.4.

4 Conclusion

The potential of the use of current mode device,OTA, for RF circuit has been discussed. From considerations of cell based design of non linear circuit, OTA is preferred for its simplicity. Though three types of communication circuits are presented in this paper, using

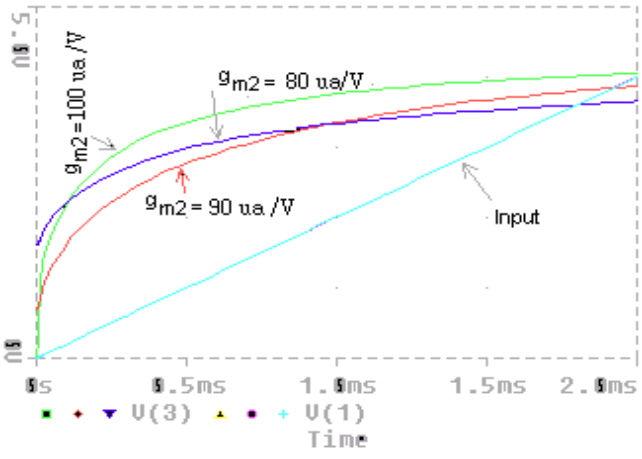


Figure 10: Input Output Characteristic of Compressor. Input Signal is Ramp Function. Three different output characteristics are shown, as it depends on g_{m2} of OTA2 (Fig. 7 and Eqn. 27).

OTA as basic building block and with the cell library concept other types of non linear circuits can easily be designed.

References

[1] S. Haykin, *Digital Communication.* Singapore: John Wiley and Sons, 1988.

[2] E. S. S. et. al., "Operational transconductance based nonlinear function synthesis," *IEEE Journal of Solid State Circuits, Vol - 24, Dec, PP - 1576-1586*, 1989.

[3] B. S. Song, "Cmos rf circuits for data communication applications," *IEEE journal of solid state circuits, Vol - Sc - 21, No -2. April, PP - 350-377.*, 1992.

[4] B. N. Ray, P. Palchaudhuri, and P.K.Nandi, "Design of ota based field programmable analog arary," *13 th International Conference on VLSI Design; Calcutta , India.*, 2000.

[5] M. Taher and Abuelma'atti, "New ask/fsk/psk/qam wave generator using multiple-output operational transconductance amplifiers," *IEEE Tran on Circuit and System -I Fundamental Theory and Application, Vol. 48 , No. 4, PP - 487-489 December*, 2001.

[6] A. W. Hietala, "A quad band 8 psk / gmsk polar transreciever," *IEEE Journal of Solid State Circuits , Vol. 41 , No. 5, PP - 1131 - 1141 May*, 2006.

[7] R. Bagheri, "An 800mhz - 6ghz software defined wireless receiverin 90 nm cmos," *IEEE Journal of Solid State Circuits , Vol. 41 , No. 12, PP - 2860 - 2879 December*, 2006.

[8] J. R.-A. S. Baswa, A.J. Lopez-Martin and R. Carvajal, "Low-voltage micropower super class ab cmos ota," *ELECTRONICS LETTERS,19th February 2004 Vol. 40 No. 4*, 2004.

[9] J. R.-A. Antonio J. Lpez-Martn, Sushmita Baswa and R. G. Carvajal, "Low-voltage super class ab cmos ota cells with very high slew rate and power efficiency," *IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 40, NO. 5, MAY, pp. 1068-1077*, 2005.

[10] F. A. P. Barqui and A. Petraglia, "Linearly tunable cmos ota with constant dynamic range using source-degenerated current mirrors," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–II: EXPRESS BRIEFS, VOL. 53, NO. 9, SEPTEMBER , pp. 797-801*, 2006.

[11] S. O. Cannizzaro, G. P. Alfio Dario Grasso, Rosario Mita, and S. Pennisi, "Design procedures for three-stage cmos otas with nested-miller compensation," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–I: REGULAR PAPERS, VOL. 54, NO. 5, MAY, pp. 933- 940*, 2007.

[12] X. Zhang and E. I. El-Masry, "A novel cmos ota based on body-driven mosfets and its applications in ota-c filters," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–I: REGULAR PAPERS, VOL. 54, NO. 6, JUNE, pp. 1204-1212*, 2007.

[13] S. Szczepa'nski, A. Wyszyn'ski, and R. Schaumann, "Highly linear voltage controlled cmos trans conductor," *IEEE Trans. Circuit and System, Vol. 40, April pp. 258-262*, 1993.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Optical Directional Coupler Switch Using Domain Inversion Technology

Ridwan Rafi Hossain, Mohammad Kibria Chowdhury, Ehsanul Matin Shujon, and M.S. Islam[+]

Department of Electrical and Electronic Engineering, Bangladesh University
of Engineering and Technology, Dhaka, Bangladesh
[+]E-mail:islams@eee.buet.ac.bd

Abstract-The switching voltage reduction in optical directional coupler switch utilizing alternating Δβ is fully analyzed. It is shown that increasing the number of sections in electrode can reduce the switching voltage significantly. But high-speed operation of the switch is limited as the number of electrode-section increases. In this paper, the improvements of above problems are discussed. By the introduction of domain inversion technology in lithium niobate ($LiNbO_3$), lower switching voltage and high-speed operation of the optical directional coupler switch can be achieved.

I. Introduction

For an optical switch, achieving of perfect cross state and bar state to reduce the crosstalk and reduction of the switching voltage are required. A typical switching voltage at present for a wavelength of 1.5μm is about 5V. As it is difficult to fabricate exactly identical waveguides in the conventional directional coupler optical switch, so achieving perfect cross state is hard. The most popular type optical switch, alternating $\Delta\beta$ with electrodes divided into two sections, can achieve perfect cross state and bar state even though the two waveguides are not exactly identical [1]. However the switching voltage is high and high speed operation cannot be achieved.

In this paper, reduction of switching voltage by dividing the coupler into multiple sections is analyzed. The switching voltage and interaction length of optical switch with alternating $\Delta\beta$ of five and six sections are calculated. Advantages and disadvantages of the division are discussed. Furthermore we propose the introduction of domain inversion technology in $LiNbO_3$ (hereafter referred as LN) for alternating $\Delta\beta$ type optical switch. The new design has advantages in forming multiple sections, high-speed switching and minimizing switching voltage.

II. The Principle of Directional Coupler Switch

The schematic of a conventional directional coupler switch is shown in Fig. 1. Here two waveguides with two parallel strip electrodes are formed in $LiNbO_3$,. The principle of operation of this switch is given in Fig. 2. If the interaction length, L equals the minimum conversion length, L_o, the propagation constants β_1 and β_2 are equal. With no voltage applied between two electrodes the light is guided into one waveguide and outputs from another. This state is defined as cross state and we associate with the symbol ⊗. But if appropriate voltage is applied between two electrodes to satisfy $(\beta_1 - \beta_2)/\kappa = 2\sqrt{3}$, the light outputs from the same waveguide. This state is defined as bar state and we associate with symbol ⊖.

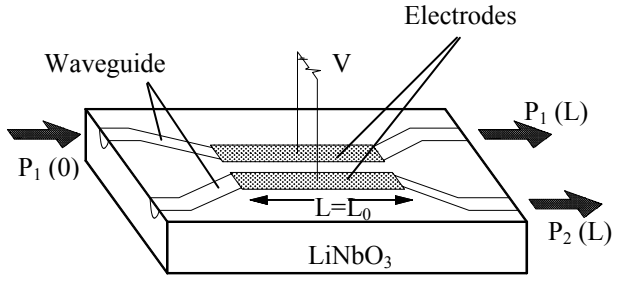


Fig. 1. The schematic diagram of an optical directional coupler switch.

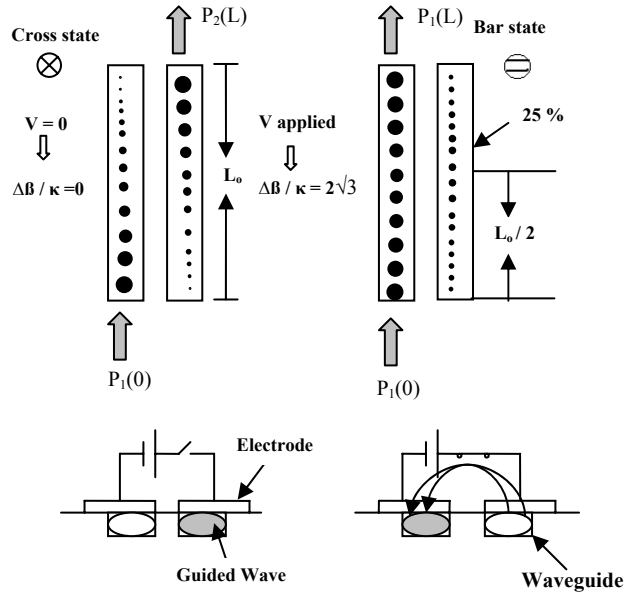


Fig. 2. The principle of operation of an optical directional coupler switch (single section).

If $P_1(L)$ and $P_2(L)$ are the received power in case of bar state and cross state, respectively then we can write,

$$P_1(L) = 1 - \frac{1}{1+\left(\frac{\Delta\beta}{2\kappa}\right)^2} Sin^2 \sqrt{1+\left(\frac{\Delta\beta}{2\kappa}\right)^2}\ \frac{\pi}{2L_0} L \qquad (1)$$

$$P_2(L) = \frac{1}{1+\left(\frac{\Delta\beta}{2\kappa}\right)^2} Sin^2 \sqrt{1+\left(\frac{\Delta\beta}{2\kappa}\right)^2}\ \frac{\pi}{2L_0} L \qquad (2)$$

Where L is the distance traveled by the transmitted wave. The variation in received power for $\Delta\beta/\kappa = 0$ and $\Delta\beta/\kappa = 2\sqrt{3}$ is shown in Fig. 3.

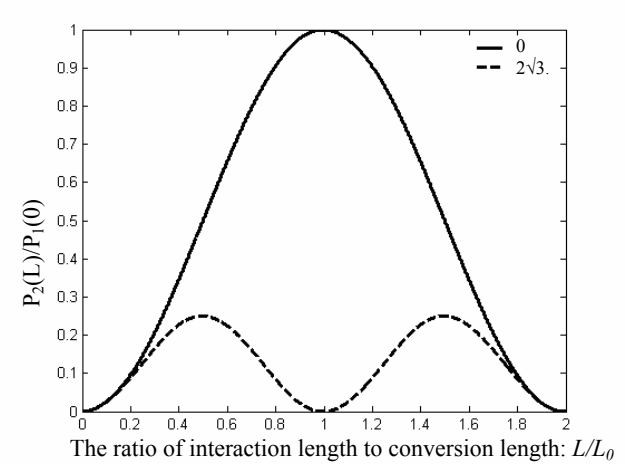


Fig. 3. $P_2(L)/P_1(0)$ versus the ratio of interaction length to conversion length, L/L_0.

III. The Construction of Alternating Δβ

The schematic of alternating $\Delta\beta$ type optical switch with two and five sections are shown in Fig. 4 and Fig. 5, respectively. Alternating voltages is applied to each divided electrode referred as $+\Delta\beta$ and $-\Delta\beta$ [1,2]. By adjusting the applied voltage perfect cross state and bar state can be achieved.

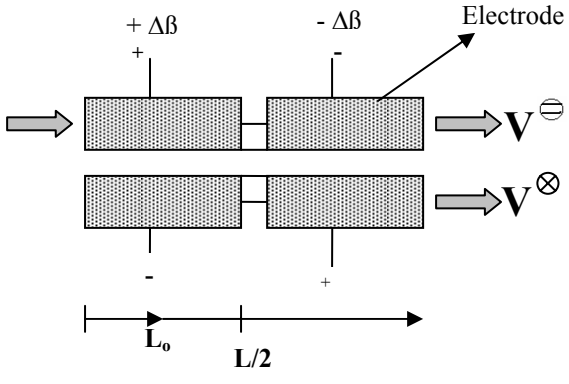


Fig. 4. The schematic of alternating $\Delta\beta$ type optical switch with 2-section.

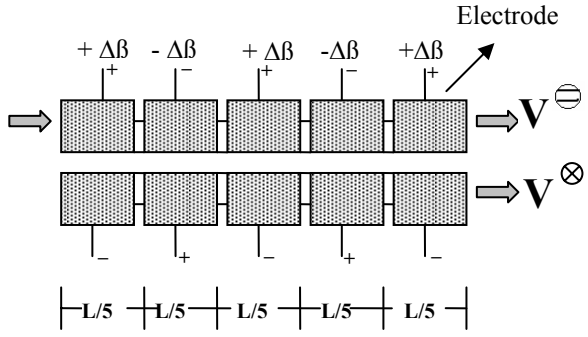


Fig. 5. The schematic of alternating $\Delta\beta$ type optical switch with 5-section.

The state of this type of switching device can be represented graphically in a switching diagram. Figure 6 shows the switching diagram with five sections. The horizontal axis represents $\Delta\beta L/\pi$, which increases proportionally to the applied voltage. The vertical axis represents L/L_o, where L is the interaction length and L_o is the minimum conversion length. V_S represent the switching voltage for a particular L/L_o. Clearly, many such combinations can be drawn on the figure for different values of L/L_o and for the same value of L/L_o. By adjusting the interaction length or applied voltage, any state in the switching diagram can be reached, except for the two axes. The region except two axes is the switching region. From Fig. 6 it is seen that perfect cross state and bar state can be achieved by adjusting applied voltage provided the L/L_o is between 4 and 5. In this figure the voltage corresponding to cross state and bar state are represented as $V^{\otimes}$ and $V^{\ominus}$, respectively. Switching voltage, V_S is calculated as $V^{\otimes}$ - $V^{\ominus}$ [3].

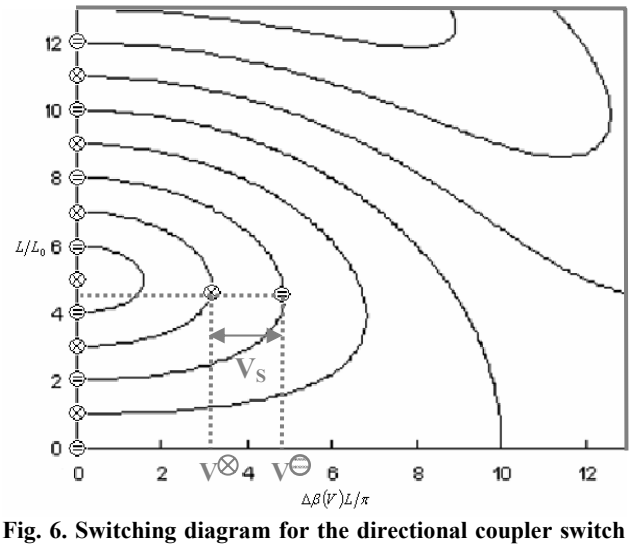


Fig. 6. Switching diagram for the directional coupler switch with 5-section of alternating $\Delta\beta$.

IV. Minimization of Switching Voltage

Calculation of minimum switching voltage for four sections and two sections was performed by Minakata, *et*

al. [3]. Here, we have calculated the minimum switching voltage for five and six sections.

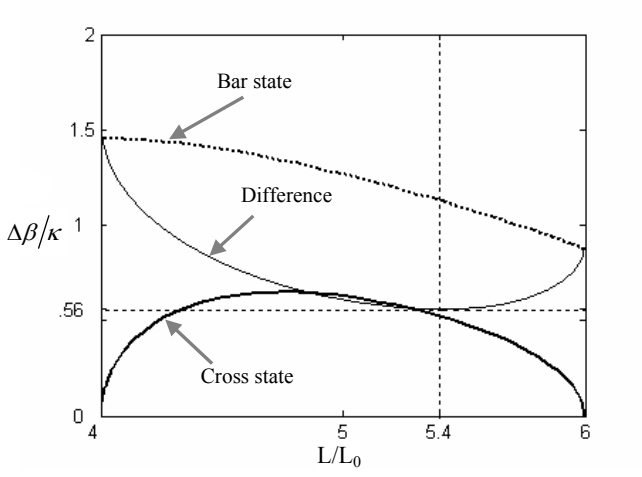


Fig. 7. The calculated state of alternating $\Delta\beta$ optical switch with 5-section.

Equations (3) and (4) represent the cross state and bar state, respectively for five sections.

$$\frac{4}{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\sin^2\left(\frac{\pi}{20}\frac{L}{L_0}\sqrt{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\right)=\sin^2\left(\left(\frac{2\nu+1}{5}\right)\frac{\pi}{2}\right) \quad (3)$$

[Where ν = 0, 1, 2, 3…]

$$\frac{4}{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\sin^2\left(\frac{\pi}{20}\frac{L}{L_0}\sqrt{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\right)=\sin^2\left(\frac{2\pi}{5}\right) \quad (4)$$

In Fig. 7 we have calculated the minimum switching voltage for five sections. The dotted line is for bar-state and the solid line is for cross state. The difference of the cross state and the bar state gives the switching voltage. The values of $\Delta\beta$/k and L/L_0 give the minimum switching voltage are 0.56 and 5.4, respectively.
Similarly, equations (5) and (6) represent the cross state and bar state, respectively for six sections.

$$\frac{4}{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\sin^2\left(\frac{\pi}{24}\frac{L}{L_0}\sqrt{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\right)=\sin^2\left(\left(\frac{2\nu+1}{6}\right)\frac{\pi}{2}\right) \quad (5)$$

[where ν= 0,1,2,3,…]

$$\frac{4}{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\sin^2\left(\frac{\pi}{24}\frac{L}{L_0}\sqrt{4+\left(\frac{\Delta\beta}{\kappa}\right)^2}\right)=\sin^2\left(\frac{\pi}{3}\right) \quad (6)$$

The values of $\Delta\beta$/k and L/L_0 for the minimum switching voltage are 0.48 and 6.3, respectively (Fig. 8).

Therefore, we see that as section number increases, necessary interaction length increases for corresponding minimum switching voltage, which is much lower than switching voltage required with single section. But the electrodes with multiple sections will degrade high-speed operation.

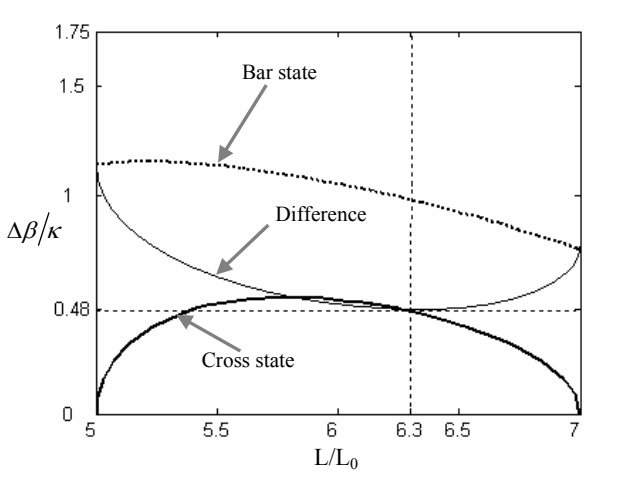


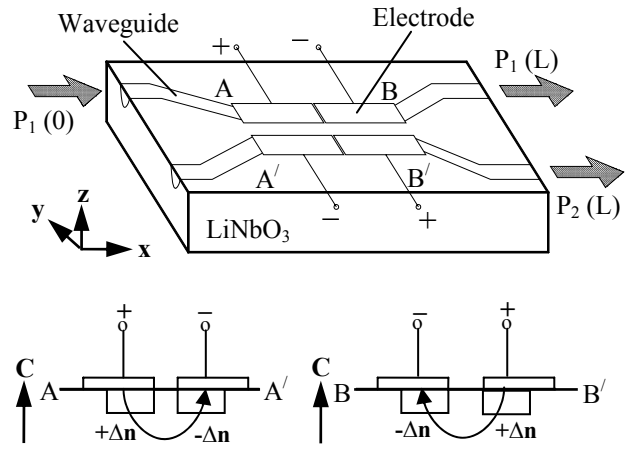
Fig. 8. The calculated state of alternating $\Delta\beta$ optical switch with 6-section.

The calculated results of switching voltage and interaction length for various sections is summarized in Table-1. The calculation shows that the switching voltage with five sections electrode is 40% of the one with two sections and 16% of one with single section electrode. The switching voltage with six sections electrode is 35% of the one with two sections and 14% of one with single section electrode. Lower switching voltage can be achieved by increasing the number of sections. Practically it is difficult to fabricate the electrode with multiple sections. Furthermore the electrode with multiple sections will degrade high-speed operation as mentioned earlier. To solve the problems, we propose a new method-introduction of the polarization inversion to the coupler region in $LiNbO_3$ by domain inversion technology.

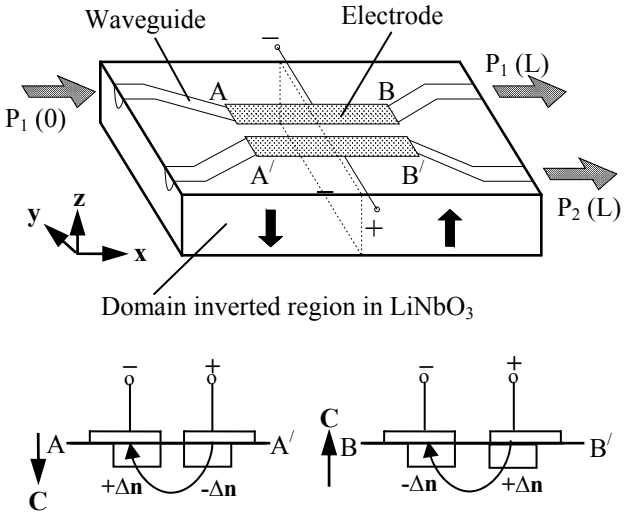
Table-1: Relationship between number of sections, the interaction length and switching voltage.

Number of sections, N	***Interaction length, L***	***Switching voltage, V_0***
1	L_0	V^1
2*	2.7 L_0	0.40 V^1
3	3.6 L_0	0.28 V^1
4*	4.5 L_0	0.22 V^1
5	5.4 L_0	0.16 V^1
6	6.3 L_0	0.14 V^1

V^1: The switching voltage for single section, *From [3].



(a) Optical switch with alternating Δβ.



(b) Optical switch with domain inversion.

Fig.9. Comparison between optical switch with alternating Δβ and optical switch with domain inversion.

V. The Concept of Domain Inversion

The optical switch as shown in Fig.9 is discussed to show the concept of domain inversion technology. A z-cut LN crystal is used as the base material, domain inverted region is formed in the half of the interaction length of waveguide with a single electrode as shown in Fig.9(b).The inversion makes the electrooptic constants in the left side and right side of the crystal to be opposite. The changes of refraction index by applying the voltage, deduced in two regions via the electrooptic effect are the same as the ones in the optical switch with alternating Δβ (Fig.9.(a)). Therefore the theory of an optical switch with alternating Δβ can be applied, without dividing electrode and perfect cross-state and bar-state can be achieved. Moreover, because the multiple sections can be easily achieved by the domain inversion technology, it is possible to decrease the switching voltage significantly. Introduction of domain inversion technology would enhance the degree of freedom in the design of the optical switch. The high-speed operation would be realized by using single electrode as the traveling wave electrode [4].

VI. Conclusions

The switching voltage and interaction length of optical switch with alternating $\Delta\beta$ of five and six sections are calculated. Advantages and disadvantages of the division are discussed. It is concluded that the lower switching voltage can be achieved by increasing the number of electrode sections. For example, the switching voltage with five-section electrode is 40% of the one with two-section and 16% of one with single-section electrode. The switching voltage with six-section electrode is 35% of the one with two-section and 14% of one with single-section electrode. In the proposed directional coupler optical switch with domain inversion, the alternating Δβ can be achieved with domain inversion technology without dividing electrode. Introduction of the domain inversion technology would enhance the degree of freedom in the design and make it possible to achieve both perfect cross-state and bar-state, lower switching voltage and high-speed operation.

References

[1] H. Kogelink and R.V. Schmidt, “Switched directional couplers with alternating Δβ”, IEEE J. of Quantum Electronics, vol.12, no.7 pp.396-401, July 1976.

[2] R.V. Schmidt and H. Kogelnik, “Electro-optically switched coupler with stepped Δβ reversal using Ti-diffused LiNbO3 waveguides”, Appl. Phys. Lett., vol.28, no.9 pp.503-506, May 1976.

[3] M. Minakata, T. Sakano and S. Kawakamin. ”Miniaturized directional coupler-type optical switches”, Trans. Inst. Electron. Inform. Commun. Eng., vol.J71-C, no.5 pp.666-671, May 1988.

[4] S. K. Korotky, G. Eisenstein, R. S. Tucker, J. J. Veselka and G. Raybon, “Optical intensity modulation to 40 GHz using a waveguide electro-optic switch”, Appl. Phys. Lett., vol. 50, no. 23, pp. 1631-1633, June 1987.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A New Switching Algorithm for TSISS Network

S. M. Raiyan Kabir [1], Rezwanur Rahman [2], Anita Quadir [3], S. P. Majumder [4]
[1] Department of Electrical and Electronic Engineering, United International University, Bangladesh
[2] Department of Aerospace Engineering and Mechanics,
The University of Alabama, Tuscaloosa, Alabama, USA.
[3] Department of Computer Science and Engineering, BRAC University, Bangladesh
[4] Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Bangladesh
E-mail: raiyan.kabir@gmail.com, djrezwan@yahoo.com, aniitaa.quadir@gmail.com,
spmajumder@eee.buet.ac.bd

Abstract— Time Slot Interchange and Signal Separator (TSISS) network was introduced to reduce the time slot requirement of TST architecture. In the switching algorithm of the TSISS network, an adjacency matrix was formed and Breath-first search (BFS) or Depth-first search (DFS) algorithms was applied on the matrix to find optimal input and output time slots. As time complexity of the algorithm for formation of adjacency matrix and also time complexity of BFS or DFS are non-linear, switching time will increase non-linearly with the increment of number of subscribers. In this paper a new switching algorithm is proposed which forms a distance metric, instead of adjacency matrix. It also utilizes a linear minimum selection algorithm instead of using BFS or DFS to determine the optimal input output time slots. All parts of the proposed algorithm have linear or constant time complexity. As a result, the maximum time complexity of the proposed algorithm is linear. So, switching time will increase linearly with the number of subscribers.

I. Introduction

The demand for faster and more efficient switching in telecommunication industry is increasing every day. As time switches ensure maximum utilization of resources using a technique called time division multiplexing (TDM), they are good candidates for fulfilling present and future demand [1]. But TDM switches has a short coming; that is, if the bandwidth of every call is constant and more calls are needed to be switched, the duration of time slots has to be reduced [2].

Time-Space-Time (TST) switches can be expanded without increasing the operational frequency of the network. According to [3], [4]; in conventional TST networks, $m \geqslant 2n-1$ time-slots per frame are needed to build a strict-sense nonblocking switch for n input per frame.

To address this issue Time Slot Interchange and Signal Separator (TSISS) network was introduced in [2]. TSISS network is a strict-sense non-blocking switch [2] with n time-slots per switching block. The switching algorithm used in the TSISS network utilized an algorithm for formation of an adjacency matrix and Depth-first search (DFS) [5] or Breath-first search (BFS) [5] for selecting the optimum input and output time slots from it. In [2], the time complexity of the algorithm was not addressed. Both BFS and DFS have time complexities of $O(|V|+|E|)$[5] that is $O(n^2)$[6]. The time needed to make switching decision, will increase nonlinearly with number of subscribers.

This paper, proposes a new switching algorithm with lower time complexity for the TSISS network to reduce the time needed for taking switching decisions.

II. Overview of the TSISS Network

Fig. 1(a) illustrates the TSISS network. It is a combination of several delay buffers and time multiplexed space switches [2]. As shown in Fig. 1(a), the TSISS network is divided into several parallel frames. Frames are the parallel parts of the TSISS network that can be used as an independent TDM switch. When several Frames are connected together, a multi-frame TSISS network is formed.

Delay requirement for the TSISS network is very low. The total memory required for the switch is $m_{total} = n \times x \times r$ bits. where, n is the number of subscribers, x is the length of a time slot in bits and r is the number of frames.

The cross connects are the parts that take the outputs of the delay buffers and distribute them among the output demultiplexers of different frames. As there are n lines per frame and r number of frames, $n \times r$ cross connects are used.

The joining network joins a specific output of all cross connects and connects them to the demultiplexer of a specific frame. Joining technique of the joining network is illustrated in Fig. 1(b). Extended description of different parts of the TSISS network can be found in [2].

III. Definition of Time-slots

A. Input-Output Time-slots

Input time-slots are the time slots that the TSISS network receives from input multiplexers. Input time-slot allocation to subscribers is controlled by the decision of switching algorithm.

Output time-slots are the time slots that the TSISS network gives as output to the demultiplexer. The demultiplexers demultiplex the output time slots according to the instructions of the switching algorithm.

A subscriber of one frame can connect to a different subscriber of any frame. The sets of input time slots and output time slots of the same frame are disjoint.

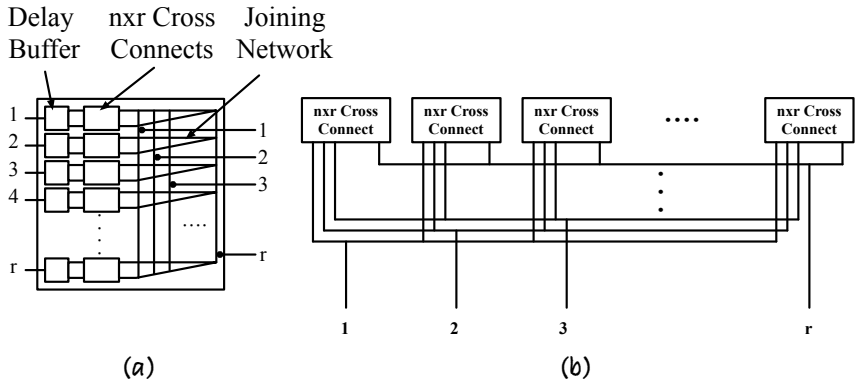


Fig. 1. (a) The TSISS network, (b) Joining network of the TSISS Network

B. Switchable Time-slots

Unused time slots are defined as switchable time slots. To make a successful connections, this type of time slots should be available in both input time slots' set of the source frame and output time slots' set of the destination frame.

IV. Switching Algorithm

Switching algorithm is the intelligence that decides which input and output time slots will be optimal for the call at present condition.

According to [2] all connections in the input frame is sampled periodically and the period of one sample is distributed among all subscribers connectable to the input frame. *i.e.*, if there are n subscribers in the input frame time slot allocation can be expressed as Equ 1[7]. This is also true for output time-slots (Equ 1).

$$\begin{aligned} I &\equiv i \bmod \mathbf{n} \\ O &\equiv j \bmod \mathbf{n} \end{aligned} \tag{1}$$

Where, i and j are indices of input and output time slots respectively.

When an input time slot from I is transmitted to the output frame, if necessary the time slot can be placed in any free time slot in between two members of O.

A. Switchable Combination Set

In [2], the switchable input and output time slots' sets were defined by evaluating the equivalency of all input and output time slots using Equ 2.

$$p \oplus h = q \oplus h = h \tag{2}$$

where, $p, q \subset \{0, 1\}$ which refer to the binary states of a specific input and output time slots respectively. $h \subset \{0, 1\}$ is used to evaluate the equivalency of p and q.

Two $\mathbf{XOR}$ terms, $p \oplus h$ and $q \oplus h$ should be equal to h to become equivalent according to Equ 2. Logically it can be expressed as Equ 3.

$$\{p \oplus h\} \wedge \{q \oplus h\} \wedge h = 1 \tag{3}$$

The index of input and output time-slots can be deduced by Equ 4.

$$\begin{aligned} k &: \{p\} \rightarrow \mathbb{N} \\ l &: \{q\} \rightarrow \mathbb{N} \end{aligned} \tag{4}$$

Where, $k(p)$ and $l(q)$ are the index number sets of input and output sets respectively. $\mathbb{N}$ is the set of natural numbers.

A cartesian product or relation set $\mathbf{R} \subseteq \mathbf{k(p)} \times \mathbf{l(q)}$, can be obtained from the switchable input and output sets by Equ 5.

$$\mathbf{R} \subseteq \{(\mathbf{k(p)}, \mathbf{l(q)}) | \mathbf{p_{k(p)}} \sim \mathbf{q_{l(q)}}\} \tag{5}$$

where, equivalency of $p_{k(p)}$ and $q_{l(q)}$ is tested with Equ 3.

B. Decision Set

A distance or delay metric $d : \mathbf{R} \rightarrow \mathbb{N} \cup \{0\}$ [8] can be defined with Equ 6.

$$d(\mathbf{R}) = \begin{cases} k(p) - l(q) & k(p) > l(q) \\ 0 & k(p) = l(q) \\ n - k(p) + l(q) & k(p) < l(q) \end{cases} \tag{6}$$

where, $(k(p), l(q)) \in \mathbf{R}$ and n is the maximum number of subscribers in a frame.

Another cartesian product or relation set of delay $\mathbf{D}$ can be defined by Equ 7.

$$\mathbf{D} \subseteq \{(\mathbf{k(p)}, \mathbf{d(R)}) | \mathbf{k(p)}, \mathbf{d(R)} \in \mathbb{N}\} \tag{7}$$

C. Decision Making Algorithm

A vector space can be spanned with $\mathbf{D}$. Now, a set of delays A and a set of switchable input time slots B can be taken by Equ 8 [7].

$$\begin{aligned} A &= \cup_0^{max(k(p))} \{d(\mathbf{R})\} \\ B &= \cup_0^{max(k(p))} \{k(p)\} \leftrightarrow {}_{k(p)}\mathbf{D}_{\mathbf{d(R)}} \end{aligned} \tag{8}$$

Now, another set $A' \subset A$, can be taken, as given in Equ 9, which selects the set of minimum delays from the set of delays A with linear minimum selection algorithm [9]. From the inverse relation, another set B' can be obtained as given in Equ 9 using the same linear minimum selection algorithm. B' is the set of indices of input time slots

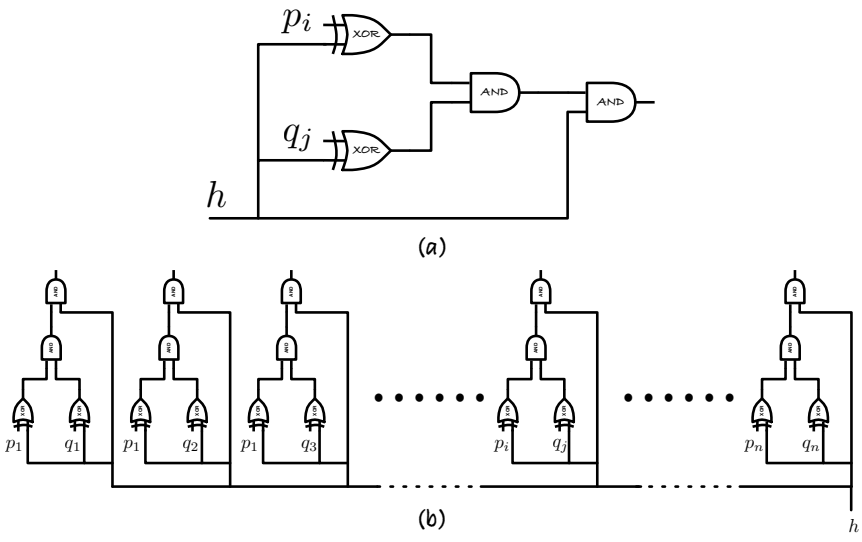


Fig. 2. (a) Logic circuit for equivalency testing, (b) Parallel logic circuit arrangement to reduce time complexity

with minimum delay. If A' is not a singleton set (all non-distinguishable elements), the minimum element of B' is taken as the optimal input time-slot for the connection.So, the corresponding output time slot is the optimal output time slot.

$$\begin{array}{rcl} A' & = & \cup\{min(d(\mathbf{R}))\} \\ B' & = & \{k(p)|_{k(p)}\mathbf{D^{-1}}_{\mathbf{min(d(R))}}\} \end{array} \quad (9)$$

V. Calculation of Time Complexity

As the algorithm is divided into 3 stages, time complexity of all the stages will be calculated separately. In this section time complexity of any operation is denoted with $\mathbf{T(operation)}$ symbol.

A. Time Complexity of Switchable Combination Set

According to Equ 4 and Equ 5, $\mathbf{R} \subseteq \mathbb{N} \times \mathbb{N}$. As $\{h\} \in \{h=0\} \vee \{h=1\}$ is a singleton set, the time complexity of Equ 5 can be given by Equ 10.

$$\mathbf{T}(\{\mathbf{p_{k(p)}} \oplus \mathbf{h}\} \wedge \{\mathbf{q_{k(q)}} \oplus \mathbf{h}\} \wedge \mathbf{h}) = \mathbf{O(n^2)} \quad (10)$$

The number of time slots in all blocks are equal in the TSISS network [2]. Equ 5 evaluates all the p's and q's of the input and output frames. Logic circuit for equivalency testing is given in Fig. 2(a). Total number of n^2 successive logic operations should be performed if only one circuit is used. As a result, the time complexity is $O(n^2)$.

A parallel logic circuit as shown in Fig. 2(b) can be implemented. To evaluate all elements in parallel, Equ 5 can be expressed as Equ 11. In Equ 11, all sets are singleton sets. The time complexity of evaluation of Equ. 11 is of order zero, given in Equ 12 [7].

$$\mathbf{R} \subseteq \left\{ \begin{array}{l} (k(p),l(q))|\bigvee_{k(p)=1}^{n}(\{p_{k(p)} \oplus h\} \\ \wedge(\bigvee_{l(q)=1}^{n}\{q_{l(q)} \oplus h\}) \wedge h = 1) \end{array} \right\} \quad (11)$$

$$\begin{array}{ll} \mathbf{T}(\{\mathbf{p_{k(p)}} \oplus \mathbf{h}\}) & = O(1) \\ \mathbf{T}(\{\mathbf{q_{k(q)}} \oplus \mathbf{h}\}) & = O(1) \\ \mathbf{T}(\{\mathbf{p_{k(p)}} \oplus \mathbf{h}\} \wedge \{\mathbf{q_{k(q)}} \oplus \mathbf{h}\}) & = O(1) \\ \therefore \mathbf{T}(\{\mathbf{p_{k(p)}} \oplus \mathbf{h}\} \wedge \{\mathbf{q_{k(q)}} \oplus \mathbf{h}\} \wedge \mathbf{h}) & = O(1) \end{array} \quad (12)$$

B. Time Complexsity of Decision Set

The decision set $\mathbf{D}$ is calculated with Equ 6 and Equ 7. The time complexity for the three conditions of Equ 6 are given by Equ 13.

$$\begin{array}{ll} \mathbf{T}(\mathbf{k(p)} - \mathbf{l(q)}) & = O(n) \\ \mathbf{T}(\mathbf{0}) & = O(n) \\ \therefore \mathbf{T}(\mathbf{n} - \mathbf{k(p)} + \mathbf{l(q)}) & = O(n) \end{array} \quad (13)$$

So, time complexity for the calculation of decision set $\mathbf{D}$ is of $O(n)$.

In [2], instead of calculating distance metric, an adjacency matrix was formed with Equ 14. As Equ 14 searches n^2 elements of $\{k(p)\} \times \{l(q)\}$ in $\mathbf{R}$, its time complexity is $O(n^2)$ [9].

$$d(k(p),l(q)) = \left\{ \begin{array}{ll} 1 & (k(p),l(q)) \in \mathbf{R} \\ 0 & otherwise \end{array} \right. \quad (14)$$

C. Time Complexity of Decision Making Algorithm

In the decision making algorithm, linear minimum selection algorithm is used firstly for selecting the set of minimum delays A' according to Equ 9. The time complexity of the linear minimum selection is $O(n)$ [9]. B' is taken with the inverse relationship with time complexity $O(n)$. If A' is not a singleton set, minimum of B' is taken with the same linear selection algorithm. So, the time complexity is $O(n)$. As a result, total time complexity is given by Equ 15.

$$\begin{array}{ll} \mathbf{T}(\mathbf{min(d(R))}) & = O(n) \\ \mathbf{T}(\mathbf{k(p)D^{-1}_{min(d(R))}}) & = O(n) \\ \mathbf{T}(\mathbf{min(B')}) & = O(n) \\ \therefore \mathbf{T}(\mathbf{decision\ making}) & = O(n) \end{array} \quad (15)$$

Table 1
Comparison of results

Stage	Previous Algorithm	Proposed Algorithm
Switchable Combination Set	$O(n^2)$	$O(1)$
Decision Set	$O(n^2)$	$O(n)$
Decision Making Algorithm	$O(n^2)$	$O(n)$
Maximum Time Complexity	$O(n^2)$	$O(n)$

In [2], BFS or DFS was used to find out the optimal input output time slots. So, time complexity of BFS and DFS is $O(n^2)$ [6]. The time complexity of the decision making part of the switching algorithm in [2] is $O(n^2)$.

VI. Result Comparison

Results of the above calculations of time complexity in different stages are summarized in Table 1.

As it is shown in Table 1, time complexity of first stage in the proposed algorithm is $O(1)$ where, the time complexity of the same part in the switching algorithm of [2] was $O(n^2)$.

In the second and third stages the proposed algorithm has time complexity of $O(n)$ where the previous algorithm had $O(n^2)$. As a result, maximum time complexity of the proposed algorithm is $O(n)$ where the previous algorithm had $O(n^2)$.

VII. Conclusion

The proposed switching algorithm uses the same cartesian product set of input and output time slots as used by the previous algorithm in [2]. Again, the proposed algorithm does not require any change in the TSISS network architecture. As a result, the TSISS network holds its strict-sense nonblocking characteristics. Hence, according to the previous section, the maximum time complexity of the network has improved remarkably in the proposed algorithm.

References

[1] A. A. Als, F. Z. Ghassemlooy, G. Swift, P. Ball, and j. Chi, "Performance of passive recirculating fiber loop buffer within an otdm transmission link," *Optical Communications*, vol. 209, pp. 137–147, 2002.

[2] S. P. Majumder, S. M. R. Kabir, R. Rahman, M. F. Imtiaz, and M. Moniruzzaman, "A new architecture of tdm switch," in *Proceedings of IEEE PACRIM 2005*, 2005, pp. 372–375.

[3] C. Clos, "A study of non-blocking switching networks," *Bell System Technical Journal*, vol. 32, no. 2, pp. 406–424, March 1953.

[4] D. K. Hunter, A. Kent, J. G. Williams, and C. M. Hall, *Encyclopedia of Information Technology*. New York/Basel: Marcel Dekker, 2000, vol. 42.

[5] T. H. Cormen, C. E. Leiseron, and R. L. Rivest, *Introduction to Algorithms*, 1st, Ed. Prentice-Hall, 1990.

[6] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*. Prentice-Hall, 1974.

[7] K. H. Rosen, *Discrete Mathematics and Its Application*, 4th ed. McGraw-Hill, 1981.

[8] M. A. Armstrong, *Basic Topology*. Spriger-Verlag New York Inc., 1983.

[9] D. E. Knuth, *The Art of Computer Programming*, 2nd, Ed. Pearson Education, 1998, vol. 3/Sorting and Searching.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

On Location Tracking and Load Balancing in Cellular Mobile Environment-A Probabilistic Approach

Sulata Mitra and Sipra DasBit
Dept. of Computer Sc.& Tech., Bengal Engineering science university,India
e-mail:mitra_sulata@hotmail.com

***Abstract*—The present work addresses probabilistic solution of two fundamental issues of cellular mobile environment maintaining a common set of information. One of the issues is to predict the location of a mobile user whereas the other is to predict the traffic load of each area and accordingly distribute channels among different areas. For location management the entire area covered by cellular architecture is considered as a hierarchy of location areas considering mobile switching centre as the root of the hierarchy and thus a new tree like data structure is introduced. When a call arrives, mobile switching centre computes the location probability of the called mobile unit at all the cells under it with the help of a database and tree-like data structure is kept in mobile switching centre itself. Finally mobile switching centre performs the appropriate searching to find the best probable cell(s) where the desired mobile unit may be traced. The scheme is made more realistic later exploiting the fact that most of the users confine their movement within a group of cells. It is done by making leaf level of the hierarchy dynamic by switching the membership of a cell from one group to the other depending upon the change of probability of finding a user at a cell. For predicting traffic load in each cell, mobile switching centre scans the same database used for predicting the location of a mobile user and finds out frequency of locating the user in each cell. With the help of this information and a heuristic function the cell wise channel requirement is computed. Finally experiments are carried out to see the variation of system cost and delay with time considering call arrival pattern as poisson distribution and movement pattern by Gaussian distribution.**

I. Introduction

In the present work probabilistic solution of two fundamental issues in cellular mobile environment has been considered. One of the issues is to locate the desired mobile unit (MU) and the other is to estimate channel requirement of the cells in the environment. To start with a generalized location tracking scheme [1] considering the equal probability of getting a MU in any cell is considered. Later the scheme is made more realistic [2] exploiting the fact that most of the users confine their movement within a group of cells. For both the schemes the entire cellular area is considered as a hierarchy of location areas (LAs) considering mobile switching centre (MSC) as the root and cells as the leaves of the hierarchy. In the scheme [1], MSC maintains a database to keep the information required to locate the desired MU. When a call arrives for an MU, MSC computes the location probability of the called MU in the cells under it and performs the best first search along the hierarchy to detect the best probable cell(s) for the called MU. But the cells having highest location probability for a particular MU remains distributed in the leaf level of the hierarchy. Thus the entire tree has to be searched for the best probable cells. The scheme is modified [2] by considering the real life movement pattern of the MUs. As most of the MUs restrict their movement within a group of cells, location probability is the highest at those cells. These highest probable cells may not be contiguous. The modification of the scheme is done by making the leaf level of the hierarchy dynamic. Each cell at the leaf level of the hierarchy maintains the MU_identifications (MU_ids) along with the frequency of locating the MUs visiting within it and sends such information periodically to the MSC. Each LA in the hierarchy belongs to a particular probability class. So the highest probable cells may belong logically under the same LA depending upon their probability values. Such dynamic change of logical position of the cells helps to reduce the search space at the instant of call arrival. Moreover the original scheme is purely centralized and depends too much on MSC. Whereas in the modified one, the information required to locate the desired MU remains distributed among MSC and base station (BS). MSC computes the leaf level structure of all the MUs in the environment periodically after receiving information from BSs. After receiving a call for an MU, MSC performs the steepest ascent depth first search to detect the most probable cell(s).
It is observed that under the constraint of scarcity of channel in this environment unless channels are assigned efficiently amongst different area, efficiency in location management can not be expected. So in this work attempts has been made to propose a scheme to estimate resources such as channels and distribute them in a balanced way. The problem is formally known as load balancing. To estimate the channel requirement for each cell, MSC scans the same database used for location management scheme. With the help of the information available in the database and a heuristic function, MSC assigns channels to the cells.

II. Previous work

In this section previous works related to the location management problem and load balancing problem are addressed.

2.1 Location management: Two different approaches can be used to solve the location management problems in the mobile environment - one is deterministic and the other is probabilistic. Most of the previous works are based on the

deterministic strategy. In this paradigm a set of schemes has been proposed using selective updation/paging. In the time based scheme [3] the MU sends its profile to the location management database after a certain period of time. Whereas in the movement based scheme [4] the MU counts the number of boundary that it crosses during movement and it updates the location management database when the count value is equal to certain threshold. But in case of the distance based scheme [5] the MU updates only after crossing a certain specified threshold distance. To reduce such redundant updation, zone based scheme [6] is proposed. To implement the said scheme, the entire cellular network is partitioned into LA and the users update whenever a LA boundary is crossed. When a call arrives, the current LA of the MU i.e. last updated position is paged. In profile based scheme [7] each MU sends its profile periodically to the respective location servers (LSs). In this scheme the system cost is reduced by selectively choosing the user profile for updation, based on dynamic categorization of the user periodically. To overcome the difficulty of sending profile, another scheme [8] is proposed, which maintains an n level hierarchy of location servers. To facilitate searching to track an MU, information for searching is distributed among different components of the hierarchy. But this scheme has huge updation cost due to periodic updation at all the levels of the hierarchy.

2.2 Load balancing: Load balancing can be done either by assigning the frequency channels among the cells or by borrowing channels from the cells that have excess channel.
In fixed assignment scheme [9], a set of frequencies is statically allocated to each cell and the same frequencies are reused in another cell sufficiently far apart such that the co-channel interference is negligible. Though this scheme is simple but if the number of calls exceeds the allocated set of channels for the cell, the excess calls are blocked until channels become available.
To cope up with the nonuniformly loaded traffic condition the dynamic channel assignment scheme (DCA) [10] is proposed. Here there is a global pool of channels from where channels are allocated on demand. The flexible channel assignment schemes [11,12] combines the concept of both the fixed and dynamic strategies, whereby there is a fixed set of channel for each cell, but channels are also allocated from a global pool in case of shortage.
But flexible schemes are basically reduced to dynamic strategies on high channel demand. Though the dynamic as well as flexible schemes are expected to cope up better with traffic overloads to a certain extent, but on high demands, the computational overheads for both the schemes deceive the purpose of the scheme.
To combat the problem of temporal as well as spatial non-uniformity of channel demand, many channel borrowing schemes [13,14] have been proposed so far.
In the centralized dynamic load balancing scheme [13] a set of fixed channels are allocated to each cell. These channels are again assigned on demand to a user in the cell. As soon as a cell becomes hot, it can borrow a channel from a compact pattern cell that has excess channel with it. However, the disadvantage here is that it is a centralized scheme. So it too much depends on central server in the MSC.
In another work [14], to minimize the load on central server, the author proposed a distributed load balancing scheme where each BS within a cell is able to run the channel borrow algorithm whenever the cell becomes hot. The proposed scheme led to a significant reduction in the call blocking probability at the cost of huge message exchange.
To reduce the number of message exchange, another load balancing technique [15] is proposed which is a combination of dynamic channel assignment and channel borrow technique in cellular network. All the cells of the system are divided into six groups, each group under one MSC. Initially a fixed number of channels are assigned to each cell under each MSC. In case of a sudden demand of channel under one MSC due to the occurrence of an unknown event, the cell who needs more channel can borrow channel from one of its compact pattern cells belonging to the same group with it or from other MSC by exchanging a negligible number of messages among the different components of the network.
Hence summarily it can be said, in spite of the fact that all the schemes mentioned here differ in strategy as far as cost reduction is concerned, more or less they belong to the category of deterministic strategy. Moreover no attempt has been reported so far to address the two important problems in this environment such as location update and load balancing in an integrated way.

III. Proposed location management scheme

In this section one probabilistic location management scheme (variant 1) is proposed. Further the scheme is modified (variant 2) to make it more realistic. In this section two variants of the probabilistic location management scheme is considered for discussion.

3.1 Variant 1: In the proposed scheme an n level hierarchy of LAs is considered. Fig.1 shows an instance of two level hierarchies of LAs. The root level of the hierarchy is the MSC whose children LAs are at level 1. The LAs at level 2 in turn be the children of the LAs at level 1 and so on. LAs at leaf level represent the cells. So the logical architecture may be considered as hierarchy of some intermediate level node (LAs) as well as leaf level node (cells). At every level of hierarchy a probability value P_i^j, where j be the level number and i be the node number (LA/cell) in the j^{th} level, is attached along with each node. The number of cells in the leaf level under each of the parent node (level n) must be same at any instant of time to facilitate a faster searching process. Uneven distribution of cells under each LA at level n may cause the cumulative probability value of a LA having large number of cells with low probability value to be greater than that of the LA having less number of cells with high probability value. In that case the less desired LA will be taken for expansion while the more desired LA will be chosen for expansion only after one further step causing undesired delay in searching.
The MSC maintains an indexed database (Fig.2) to keep the record of visiting MUs in the form of (MU_id,Cell_id). The data f_{C_i} under each cell C_i in the database corresponding to a particular MU_id indicates the frequency of locating the given MU at the cell C_i. When a call arrives at MSC, it searches the database for the desired MU. Then with the help of best first/AND-OR best first search technique, MSC finds the paging area (PA) of the called MU. PA is defined as a cell or a collection of cells having the equal highest location probability among all the cells under it for the called MU. The cells under a PA may not be contiguous. Whenever an MU is detected at a cell C_i, the frequency associated with cell C_i corresponding to that MU is updated in the database. **3.1.1 Heuristic function for location probability distribution:** This function is actually the location probability P_i^{n+1} of the leaf

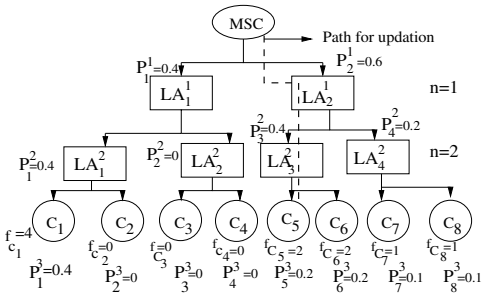


Fig. 1. An instance of two level hierarchy of LAs

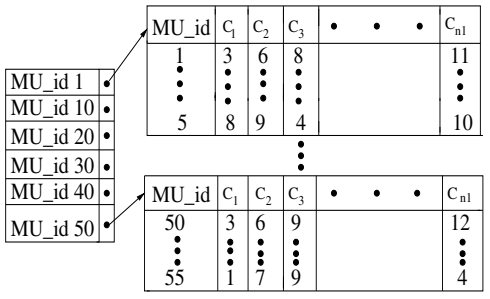


Fig. 2. Database at MSC

level node i (i.e. the cell C_i) for a specific MU. It has been defined as the ratio of number of times (frequency) the MU has been detected in the leaf level node i after paging (i.e. the number of success at i^{th} leaf level node) to total number of times it has attended calls (i.e. total number of success), over a certain interval of time (τ). Thus ${P_i}^{n+1} = f_{Ci}/T$, where f_{Ci} is the frequency associated with cell C_i for that MU and T be the total call arrives for the given MU at all the cells in the leaf level (Fig.1), T=$\sum_{i=1}^{n1} f_{Ci}$, where n1 be the number of cells in the leaf level. In Fig.1 n1=8. Hence ${P_i}^{n+1} = f_{Ci}/(\sum_{i=1}^{8} f_{Ci})$. The probability value ${P_i}^{n+1}$ attached with cell C_i varies from MU to MU. The probability corresponding to the particular MU at any parent node is the sum of probability of all its children nodes.

3.1.2 Locating an MU: In this section the searching method to locate an MU in the proposed work has been discussed. When a call comes at MSC for a particular MU, MSC searches the database (Fig.2) for the desired MU. If the result of search is null, MSC sends page to all the cells under it and inserts the data corresponding to the new MU in its database. Otherwise MSC computes the location probability of the called MU at all the nodes (LA/cell) under it and with the help of best first search technique, MSC determines the PA.

For example, in Fig.1 ${LA_2}^1$ is expanded due to high probability i.e. depth first search is started. Then the probability of ${LA_3}^2$, ${LA_4}^2$ and ${LA_1}^1$ are compared; out of which ${LA_1}^1$ and ${LA_3}^2$ are expanded as their probability values are equal and higher than ${LA_4}^2$. In this case the AND-OR best first search is proposed where all the equiprobable LAs are expanded simultaneously. Now among the children nodes ${LA_1}^2$, ${LA_2}^2$, C_5 and C_6, ${LA_1}^2$ is expanded further as ${P_1}^2$ is the highest. Again among the children nodes of ${LA_1}^2$ i.e.C_1 and C_2, C_1 is the best probable cell.

3.1.3 Updation of the database: Let us consider the database shown in Fig.2. Whenever an MU has been detected at a cell C_i at time t+1, the frequency associated with cell C_i for that MU is increased by 1 i.e. T=$\sum_{i=1}^{8} f_{Ci} = f_{Ci} + 1$. The total number of calls received by the MU is also increased by 1, i.e.T=$\sum_{i=1}^{8} f_{Ci} + 1$. So the new probability value at C_i for that MU becomes, ${P_i}^{n+1} = f_{Ci}/T$. As probability corresponding to a particular MU at a node is the sum of probabilities of all of its children nodes (i.e. immediate successors), the probability values of that MU at all the predecessor nodes are updated similarly.

For example, the database shown in Fig.2 is updated for a given MU whose f_{Ci} and ${P_i}^j$ are shown in Fig.1 at any certain instant of time t. As $\sum_{i=1}^{8} f_{Ci} = f_{C1}+f_{C5}+f_{C6}+f_{C7}+f_{C8}$=4+2+2+1+1=10, T=10.

Now let at the instant (t+1), a call arrives for a given MU and the corresponding MU is detected at cell C5. So T=10+1=11.

The new frequencies at all the nodes in the path for updation in Fig.1 are as follows:

$f_{C5_{new}} = f_{C5_{old}} + 1 = 2 + 1 = 3$, thus $P^3_{5_{new}} = 3/11$

$f_{LA^2_{3_{new}}} = f_{LA^2_{3_{old}}} + 1 = 4 + 1 = 5$, thus $P^2_{3_{new}} = 5/11$

$f_{LA^1_{2_{new}}} = f_{LA^1_{2_{old}}} + 1 = 6 + 1 = 7$, thus $P^1_{2_{new}} = 7/11$

Now f_{C_i} of all other nodes remain unchanged. But due to the increase of T, the probabilities will be modified as follows:

$f_{C1_{new}} = f_{C1_{old}} = 4$, but $P^3_{1_{new}} = 4/11$

$f_{C6_{new}} = f_{C6_{old}} = 2$, thus $P^3_{6_{new}} = 2/11$

$f_{C7_{new}} = f_{C7_{old}} = 1$, but $P^3_{7_{new}} = 1/11$

$f_{C8_{new}} = f_{C8_{old}} = 4$, but $P^3_{8_{new}} = 1/11$

Thus the probability values of the cells C_1, C_6, C_7 and C_8 are reduced from 4/10, 2/10, 1/10 and 1/10 to 4/11, 2/11, 1/11 and 1/11 respectively. Similarly the new probability values of the other parent nodes can be calculated in the same way. Hence the probability of the successful nodes and the rest of the nodes are modified which improves the search result towards more successful direction.

But in this scheme the highest probable cells for an MU are not contiguous. As a result the entire tree has to be searched to detect the best probable cells which increase the search space as well as the search time. In the modified scheme discussed in section 3.2, MSC computes the logical position of the cells at the leaf level for each MU separately. As a result the logical position of the highest probable cell for a particular MU is contiguous which helps to reduce the search space.

3.2 Variant 2: It is an extension of the scheme presented in variant 1. It becomes more realistic considering the fact that most of the MUs limit their movement within a few cells. Here also a hierarchy of LAs having (Fig.3) MSC as the root of the hierarchy is considered. Each LA belongs to a specific probability class indicating the range of probability of locating an MU in that area. Unlike variant 1 the leaf level of the hierarchy is dynamic depending upon the location probability of the MUs in the cells at the leaf level. The probability of locating the MUs at various cells is computed using the same heuristic function as discussed in section 3.1.1. The LAs at any level of the hierarchy is subdivided into the children LAs provided it crosses a threshold range of probability say 0.25 and lower range crosses a threshold limit say 0.5. The rest of the LAs are not divided further as the cells under these LAs contain cells of low probability. The selection of parent node (LA) for a leaf level node (cell) is made periodically depending upon the probability range in which the cell presently belongs.

A change in probability value of a cell w.r.t. an MU causes the cell to switch its membership from one LA to the other; MSC

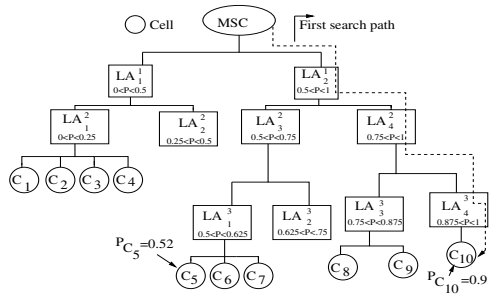


Fig. 3. An instance of search space

maintains this logical structure of LAs and cells for individual MU. The structure is dynamically updated with the help of some periodic information from BSs. Thus the information required to track an MU remains distributed among MSC and BSs. As the task of tracking an MU is shared by MSC and BSs, the scheme is a distributed one.

In the hierarchical structure shown in Fig.3 $LA_i{}^j$ is defined as i^{th} node in the j^{th} level. The probability range of the node ${LA_1}^1$ is 0 to 0.5 and that of the node ${LA_2}^1$ is from 0.5 to 1. So ${LA_1}^1$ consists of those cells at its leaf level having probability value in the range 0 to 0.5 corresponding to a particular MU. In other words, the probability with which a call arrives for an MU when the MU is under $LA_1 1$ is in the range 0 to 0.5. Again ${LA_1}^1$ is subdivided into two children classes ${LA_1}^2$ and ${LA_2}^2$ having range of probabilities 0 to 0.25 and 0.25 to 0.5 respectively. Hence the total probability space (0 to 1) is evenly divided and assigned to each LA of a particular level. The distribution of cells in the leaf level is different for different MUs. Further expanding ${LA_1}^2$ and ${LA_2}^2$ having the probability range 0 to 0.25 and 0.25 to 0.5 is useless as they contain cells of low probability. For example, in Fig.3 let the probability of finding out an MU in cell C_5 becomes 0.7 instead of 0.52 at any instant of time. Then C_5 will leave ${LA_1}^3$ and becomes a member (child) of ${LA_2}^3$. 3.2.1 Locating an MU: In this section the various operations performed jointly by BS and MSC to locate an MU is discussed.

3.2.1.1 Operation performed by BS: Base stations of each cell maintain the frequency of successful calls established for each MU visiting the cell. Whenever a call is established for an MU in a cell C_i, the corresponding BS increases the frequency of locating that MU within it by one. BSs send frequency values to MSC after every unit of time.

3.2.1.2 Operation performed by MSC: MSC computes the location probability (as discussed in section 3.1.1) associated with an MU and a cell C_i after receiving the information of frequency f_{Ci}. MSC in turn modifies the respective leaf level structure i.e. the dynamic portion of the structure maintained at MSC. As soon as a call arrives for a particular MU, MSC starts searching the probability space (i.e. the hierarchical structure) using 'Steepest Ascent Hill Climbing' - a variant of depth first search technique. In this case searching is done down the hierarchy (from MSC to cells). The child node with highest probability value is chosen and expanded. This process is repeated until the called MU is found out or all the path along the structure is exhausted. This searching process is advantageous because it will guarantee to reach the cell with highest probability in comparison to depth first and best first search because search space is reduced considerably. For example, let Fig.3 shows the structure representing the search space for a particular MU where the first search path is $MSC- > {LA_2}^1- > {LA_4}^2- > {LA_4}^3- > C_{10}$. So C_{10} is selected as the highest probable cell where the called MU may be found. If no such cell is found under ${LA_4}^3$ search has to be again performed among the nodes unexpanded yet, following the same technique.

IV. Proposed load balancing scheme

MSC uses this scheme to assign the channel dynamically to the cells under it. MSC computes the call arrival probability as well as the number of required channels of all the cells under it using a heuristic function as discussed in section 4.1 and accordingly distributes the channels to its cells. MSC performs such computation and assignment at every instant of database updation as discussed in section 3.1.3.

4.1 Heuristic function for call arrival probability: The database used in variant 1 (Fig.2) may be used to compute the heuristic function which is the call arrival probability ${Q_i}^{n+1}$ at the leaf level node i (i.e. cell C_i). It has been defined as the ratio of number of times the various MUs are detected in the i^{th} node after paging i.e. $\sum f_{Ci}$ to the total number of times the MUs are detected at all the cells in the leaf level i.e. $\sum_{i=1}^{8}\sum_{k=1}^{s1} f_{Cik}$, where s1 be the total number of MUs. Thus ${Q_i}^{n+1} = f_{Ci}/(\sum_{i=1}^{8}\sum_{k=1}^{6} f_{Cik}$, as s1=6 in Fig.4. This function helps the MSC to compute the required number of channels CH_{C_i} of all the cells C_i under it as $CH_{C_i} = ({Q_i}^{n+1}) * x$, where x be the total number of available channels with MSC.

For example, in Fig.4, the total frequency of locating the six MUs in the cell C_1 is f_{C_i}=3+1+2+1+2+5=14 whereas the total frequency of locating all the MUs in the environment within the cells C_1, C_2, C_3 and C_4 is $f_{C_1} + f_{C_2} + f_{C_3} + f_{C_4} = 91$. Thus ${Q_1}^{n+1} = 14/91$ and CH_{C_1}=14/91*x. So, if x=13, CH_{C_1}=2.

V. Simulation

Detailed simulation experiments are carried out for our probabilistic location management and load balancing scheme. Gaussian distribution is used in the experiments to simulate the real life movement pattern of various MUs considering call arrival pattern as poisson distribution in the cells under a system. In this section, the simulation environments are discussed briefly along with the experimental results.

5.1 Simulation environments: For the purpose of simulation, the Gaussian distribution (Fig.5) is considered as it meets the assumptions that most of the MUs are normally bound within a moderate number of cells, whereas a few of them may move within a wide area containing a large number of cells.

The Gaussian distribution is a plot of probability density function (pd_func) at Y-axis with respect to the corresponding values in X-axis. The pd_func is given by the relation f=$e^{-(x1-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$, where f is the pd_func corresponding to different values of variable x1, μ=mean and σ=standard deviation. In normalized form, the pd_func can be written as,

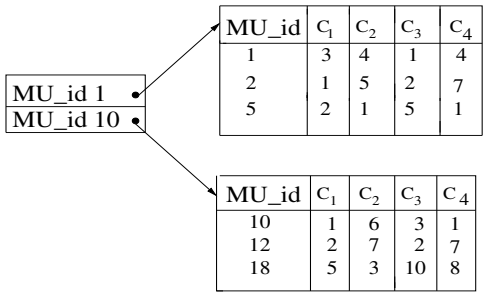


Fig. 4. Database at MSC for six MUs and four cells under MSC

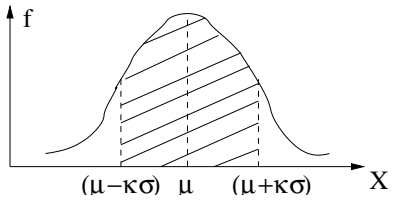


Fig. 5. Gaussian distribution

$f_1 = f\sqrt{(2\pi\sigma^2)} = e^{-(x1-\mu)^2/2\sigma^2}$. Now the probability that a value will lie in the range $a \leq x1 \leq b$ is obtained by the area covered by that region under the given probability density curve i.e. $P(a \leq x1 \leq b) = \int_a^b f\ dx1$. The probability that a value will lie in $(\mu - k\sigma) \leq x1 \leq (\mu + k\sigma)$ region is obtained as, $P_{\pm k\sigma} = P(\mu - k\sigma \leq x1 \leq \mu + k\sigma) = erf(k/\sqrt{2})$, where $erf(u) = 2/\sqrt{(\pi)} \int 0^u e^(u^2)\, du$, $u = (x1 - \mu)/(\sqrt{2}\sigma)$ and $P_{\pm k\sigma}$ denotes the shaded area in Fig.5.

5.2 Simulation Parameters: In Fig.5 the X-axis denotes the number of cells (cumulative scale) in a system, i.e. the variable values are given by number of cell. Y-axis denotes probability density function for MUs. So, $P_{\pm k\sigma}$ denotes the number of MUs that will have a movement pattern bound under cells $(\mu - k\sigma)$ to $(\mu + k\sigma)$ out of total number of MUs. Thus the cumulative number of MUs that will have a moving space $(\mu - k\sigma)$ to $(\mu + k\sigma)$ will be equal to $P_{\pm k\sigma}$ *(total number of MUs). Let total number of MUs is 100000 and number of cells is 32. Then, for μ=16 and σ=4 we have the results as shown in the TABLE.

Each k-values denotes a group consisting of a certain number of cells obtained from Gaussian distribution. **5.3 Simulation results:** In this section the performance of both the variant of probabilistic location management issue and the load balancing issue is presented considering the call arrival pattern as poisson.

TABLE

k	Number cells of covered by a certain number of MUs	$P_{\pm k\sigma}$	Number of MUs (cumulative)	Number of MUs in the corresponding group of cells
0.5	4	0.383	38300	38300
1.0	8	0.683	68300	30000
1.5	12	0.866	86600	18300
2.0	16	0.955	95500	8900
3.0	24	0.977	97700	2200
4.0	32	1.0	100000	2300

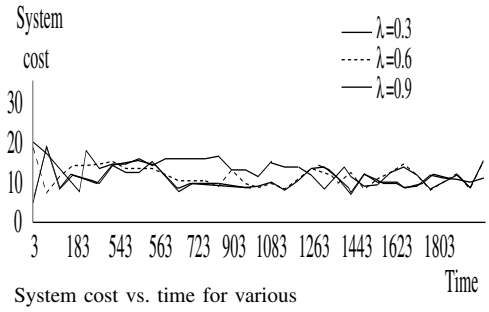


Fig. 6. System cost vs. time for various

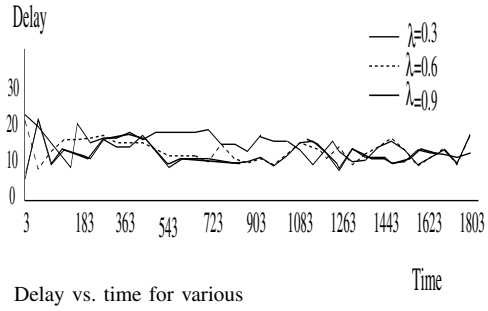


Fig. 7. Delay vs. time for various

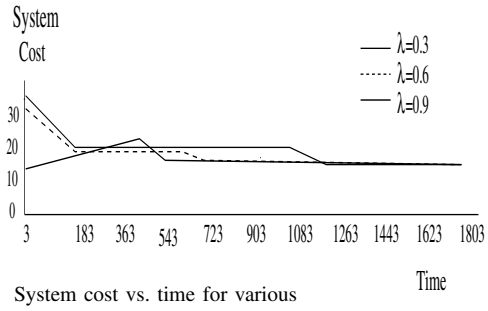


Fig. 8. System cost vs. time for various

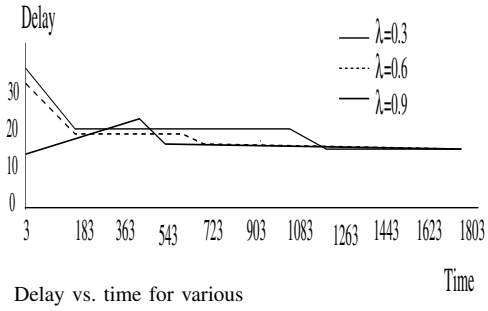


Fig. 9. Delay vs. time for various

5.3.1 Performance of variant 1: Experiments are conducted to note the variation of the system cost and time delay to search a particular MU as a function of time for different values of call arrival probability (λ). Fig.6 and Fig.7 show the performance of the variant 1 graphically. Fig. 6 shows the variation of cost and Fig.7 shows the variation of delay to locate a particular MU as a function of time for various λ. From the plot it can be observed that as time grows the average value of the system cost as well as delay remains constant.

5.3.2 Performance of variant 2: In this case the experiments are carried out to see the variation of system cost as well as delay with time for different values of λ. Fig.8 and Fig.9 show the performance of the variant 2 graphically. Fig.8 shows the

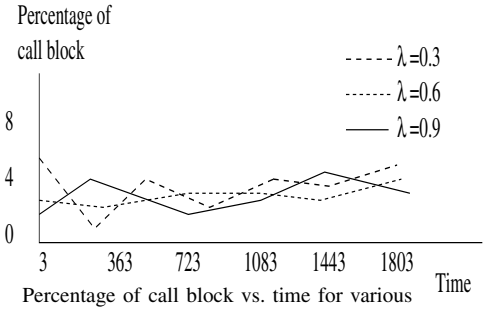


Fig. 10. Percentage of call block vs. time for various

variation of cost and Fig.9 shows the variation of delay as a function of time to locate a particular MU for various λ. From the curves it can be observed that both the cost and delay reduces with λ and becomes entirely constant after a certain interval of time.

5.3.3 Performance of load balancing scheme: In this case the detailed simulation experiments are carried out to note the percentage of call block. Fig.10 shows the variation of call block as a function of time for various λ. From the curve it can be observed that as time grows the average value of the percentage call block remain constant.

VI. Conclusion

In this paper the main theme is to address two fundamental issues in mobile cellular environment in an integrated way. The merit of the approach lies in the fact that no extra overload such as maintenance of database etc. is needed for load balancing other than which is required in location management scheme.
The present work may be extended by introducing the other issues such as Query processing in this environment as an element of the integrated effort.

References

[1] S.DasBit, P.Raha, and S.Mitra, 'A probabilistic location management strategy in cellular mobile environment', IEEE TENCON 2002.
[2] S.DasBit, P.Raha, and S.Mitra, 'A distributed probabilistic location management strategy in cellular mobile environment', Workshop on Communication Network and Media, Eurasia-ICT, 2002.
[3] C.Rose, 'Minimising the average cost of paging and registration:a timer based method", Wireless network, vt-2, pp.109-116, 1996.
[4] I.F.Akyildiz and J.Ho, 'Movement based location update and selective paging for PCS networks', IEEE/ACM Transaction on networking, vol-4, pp.629-638, 1995.
[5] M.U.Madhow and K.Steiglitz, 'Optimization of wireless resources for personal commmunication mobility tracking', IEEE/ACM Transaction on Networking, vol.3, pp.698-707, 1995.
[6] S.K.Das and S.K.Sen, 'A new location update strategy for cellular networks and its implementation using a generic algorithm', IEEE/ACM Conference on Mobile Computing and Networking/Mobicom, pp.185-194, 1997.
[7] S.DasBit and S.Mitra, 'A varying per user profile based location update strategy for cellular network', International Conference ICCT/WCC, pp.754-760, 2000.
[8] S.Mitra and S.DasBit, 'A location management strategy in cellular mobile environment using distributed searching', IEEE 3Gwireless, pp.350-355, 2001.
[9] S.M.Elnoubi, R.Singh, and S.C.Gupta, 'A new frequency channel assignment in high capacity mobile communication systems', IEEE Transaction on Vehicular Technology, vol.VT-31, 1982.
[10] D.C.Cox and D.O.Reudink, 'Increasing channel occupancy in large scale mobile radio systems:dynamic channel reassignment', IEEE Transaction on Vehicular Technology, vol.22, no.4, pp.218-222, 1973.
[11] J.Tajima and K.Imamura, 'A strategy for flkexible channel assignment in mobile communication systems', IEEE Transaction on Vehicular Technology, vol.VT-37, !988.
[12] S.DasBit and S.Mitra, 'Load balancing in a cellular mobile environment-A Database Approach', TENCON, vol.2, pp.195-200, 2000.
[13] S.K.Das, S.K.Sen, and R.Jayaram, 'A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment', IEEE/ACM Conference on Mobile Computing and Networking, 1996.
[14] S.K.Das and R.Jayaram, 'An efficient distributed channel management for cellular mobile networks', IEEE International Conference ICUPC, pp.654-660, 1997.
[15] S.Mitra and S.DasBit, 'Load balancing strategy using dynamic channel assignment and channel borrowing in cellular mobile environment', International Conference ICPWC, pp.278-282, 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Analysis of Star-Tree and Ring-Bus Millimeter-wave Fiber-Radio Networks Incorporated with Cascaded WDM Optical Interfaces

Masuduzzaman Bakaul[1], Ampalavanapillai Nirmalathas[1], Christina Lim[2], Dalma Novak[3], and Rod Waterhouse[3]

[1]National ICT Australia (NICTA), Dept. of Electrical and Electronic Engineering,
The University of Melbourne, VIC 3010, Australia.
[2]Centre for Ultra-Broadband Information Networks (CUBIN), Dept. of Electrical and Electronic Engineering,
The University of Melbourne, VIC 3010, Australia.
[3]Pharad LLC, Glen Burnie, MD, USA
Ph: +61-3-8344 6061, Fax: +61-3-8344 6678, Email: mbakaul@ee.unimelb.edu.au

Abstract— **The performance of wavelength-division-multiplexed (WDM) optical interfaces that support 37.5 GHz-band 25 GHz-separated wavelength-interleaved dense-WDM (WI-DWDM) signals is investigated for networks in both star-tree and ring/bus architectures. The interface offers consolidated base station (BS) architecture by enabling a wavelength reuse technique with transparent optical-add-drop-multiplexing (OADM) to the BS. The results show that the proposed interface can be a suitable candidate in future WI-DWDM mm-wave fiber-radio networks, configured in either star-tree or ring/bus architecture**.

I. INTRODUCTION

Millimeter-wave (mm-wave) fiber-radio systems are promising technologies for the transport of future broadband wireless access (BWA) services [1]. In these networks multiple remote antenna base stations (BSs), suitable for providing untethered connectivity for BWA applications, can be directly serviced by a central office (CO) via an optical fiber feeder network [2]. The introduction of wavelength interleaving (WI) enables these systems to be consistent with dense-wavelength–division-multiplexed (DWDM) feeder networks [3], while the reuse of downlink optical carriers for uplink data transport eliminates the need for a light source at the BS [4]. In [5], a multifunctional wavelength-division-multiplexed (WDM) optical interface was presented that effectively adds and drops the desired channels to and from the wavelength-interleaved-DWDM (WI-DWDM) feeder network, in addition to offering a simplified and consolidated BS by enabling a wavelength reuse technique. The interface is comprised of narrow band fiber Bragg gratings (FBGs) and multiport optical circulators (OCs), which unfortunately introduce additional crosstalk and chromatic dispersion in the link. The effects of these impairments caused by single as well as cascaded WDM optical interfaces (WOIs) were investigated in [6]. The cascaded interface was comprised of two WOIs connected by a small piece of patchcord, without any fiber between them. These analyses were particularly important in quantifying the power penalties introduced by each stage of cascade, in addition to the impacts on the performance of drops and add channels. However, in a practical WI-DWDM fiber-radio network, either in star-tree or rings/bus architectures that are considered very effective in delivering future broadband wireless services to customers [7-9], the accumulated effects of cascaded interfaces could be quite severe and may limit the network-dimensioning enormously. This paper further extends the previous investigations for cascaded interfaces and evaluates the performance of WI-DWDM fiber-radio networks both in star-tree and rings/bus architectures incorporated with cascaded WDM optical interfaces and quantifies their dimensioning.

II. PROPOSED WDM OPTICAL INTERFACE

Fig. 1shows the schematic of the proposed WDM optical interface with the optical spectra obtained from

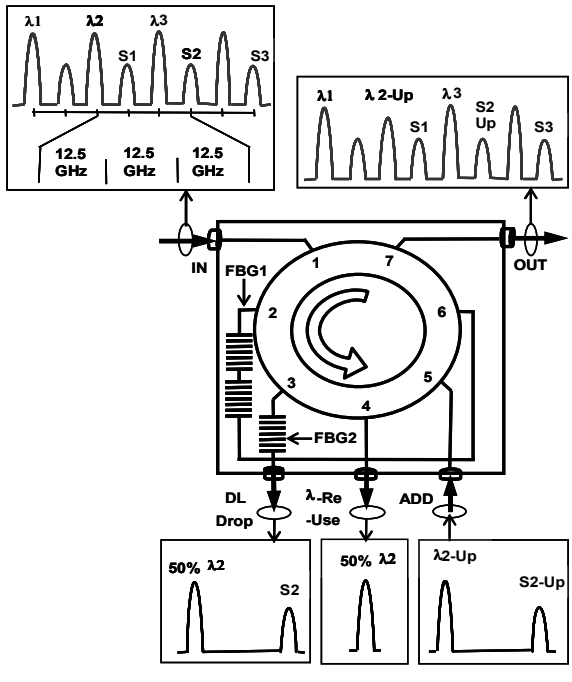


Fig. 1: Proposed WDM optical interface enabling the wavelength recovery and optical add-drop functionality for a wavelength-interleaved DWDM fiber-radio system.

corresponding input, output, drop and add ports of the interface shown as insets [5]. The input spectrum shows three 37.5 GHz-band wavelength-interleaved signals with a DWDM channel separation of 25 GHz, generated in optical single sideband with carrier (OSSB+C) modulation format. The optical carriers namely λ_1, λ_2, λ_3 and their respective modulation sidebands at S_1, S_2, S_3 of the optical mm-wave channels are interleaved in such a way that after interleaving the adjacent channel spacing, irrespective of carrier or sideband, becomes 12.5 GHz.

The interface consists of a 7-port OC connected to a two-notch FBG (FBG1) between port-2 and port-6 and a single-notch FBG (FBG2) at port-3 of the OC with a notch bandwidth of $\leq$ 12.5 GHz each. The FBG1 is designed in such a way that it reflects 100% of a specific downlink optical carrier (for instance, λ_2) with its modulation sideband (S_2), from the input WI-DWDM mm-wave fiber-radio signals. The reflected signal is received at port-3 while the transmitted signals (the through channels) are routed to port-6 of the OC where they will exit the interface via port-7 (OUT). FBG2 at port-3 was designed to reflect only 50% of the carrier at λ_2 while the remaining 50% of the carrier and the corresponding sideband, S_2 of the downlink signal will be dropped at port-3 (DL Drop) that can be detected using a high-speed photodetector (PD). The reflected 50% carrier at λ_2 is recovered at port-4 (λ-Re-Use) of the OC and will be reused at the BS as the optical carrier for the uplink path.

In the uplink direction, a dispersion-tolerant OSSB+C formatted optical signal is generated using the recovered optical carrier and the uplink radio signal at the same RF frequency as the downlink mm-wave signal. The optically modulated uplink signal is then added to the interface via port-5 of the OC. The added signal will be routed to port-6 where it will be reflected by FBG1 and combines with the remaining wavelength-interleaved channels (the through channels) before being routed out of the interface via port-7 (OUT). The output spectrum along with the spectra of the downlink drop, the recovered wavelength reuse carrier and the uplink signal generated by using the recovered optical carrier are shown in the inset of Fig. 1. The proposed interface thus enables the BSs of fiber-radio systems to the WI-DWDM fiber-feeder networks by dropping and adding the desired signals, in addition to providing the optical carrier for the uplink path, which simplifies the systems by removing the uplink light-source completely.

III. Experimental Verification Of The Proposed Interface In Cascaded Setup

Fig. 2 shows the setup of the experiment that verifies the proposed interface in cascade. In the downlink direction, three narrow-linewidth lasers λ1 (1556.2 nm), λ2 (1556.4 nm) and λ3 (1556.6 nm) were combined and applied to a dual-electrode Mach-Zehnder modulator (DE-MZM). A 37.5-GHz mm-wave signal was generated by mixing a 37.5-GHz local oscillator (LO) signal with 155 Mb/s data in binary phase shift keyed (BPSK) format. The mixer output was then amplified and applied to the

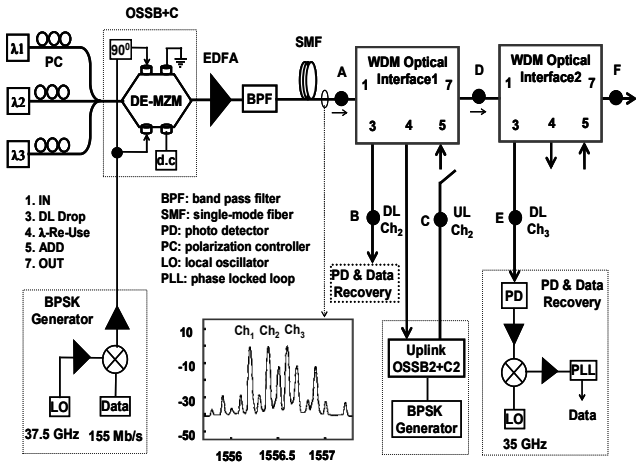


Fig. 2: Experimental setup that characterizes the effects of optical impairments in single as well as in cascaded WOIs

DE-MZM that generates OSSB+C modulated optical mm-wave signals, with three optical carriers and their respective sidebands interleaved. The interleaved signal was again amplified by an erbium doped fiber amplifier (EDFA) and passed through an optical band pass filter (BPF) prior to being transported over 10 km of singlemode fiber (SMF) to the two concatenated WDM optical interfaces, WOI_1 and WOI_2. The signal entering concatenated interfaces is shown in the inset of Fig. 2, where the three interleaved carriers with their respective sidebands are denoted as Ch_1, Ch_2 and Ch_3 for simplicity. Similar to Fig. 1, each interface in Fig. 2 is shown as a block with five ports, namely, the input (IN), the downlink drop (DL Drop), the wavelength reuse drop (λ-Re-Use), the add (ADD) and the output (OUT) port. During the experiment WOI_1 was assigned to drop and add Ch_2 while WOI_2 was to drop and add Ch_3. In the uplink direction, the OSSB+C formatted uplink (UL) Ch_2 was generated by modulating the recovered λ-Re-Use carrier with another 37.5 GHz-band UL mm-wave signal carrying 155 Mb/s BPSK data. The UL Ch_2 was then routed to WOI_1 via the ADD port. The effects of impairments on the WI-DWDM channels due to traversing the cascaded interfaces were characterized by recovering the transmitted channels at positions A, B, C, D, E, and F indicated in Fig. 2. To make the measurements comparable, the same photodetection and

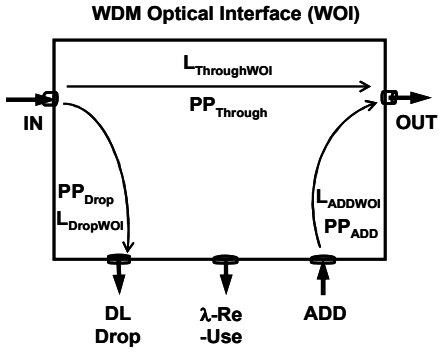


Fig. 3: Illustration of parameters of WDM Optical Interface used in the modelling of networks based on power budget calculation.

Symbol	Description	Value
$L_{ThroughWOI}$	Insertion loss (IL) of the through channels	3.2 dB
$L_{DropWOI}$	IL of the downlink (DL) drop channel including uplink (UL) carrier recovery	5.8 dB
L_{ADDWOI}	IL of the UL add channel	1.3 dB
$PP_{Through}$	Power penalty (PP) experienced by the through channels	0.4 dB
$PP_{IN\text{-}DL\ Drop}$	PP experienced by the DL drop channel	-0.3 dB
$PP_{ADD\text{-}OUT}$	PP experienced by the UL add channel	0.65 dB
L_{MOD}	IL of OSSB+C modulator in CO	15.9 dB
G_{BAMP}	Amplification by boost EDFA in CO	22.5 dB
L_{SMF}	Attenuation of signal in 10 km SMF	2.2 dB
T_{LSCO}	Power launched from the light-source	0.3 dBm
L_{MUX}	Insertion loss of the optical combiner	4.9 dB
PD_{SEN}	Sensitivity of the Photodetector	-14.2

Table 1: WDM Optical Interface parameters used in analysing the performance of the networks

data recovery circuit was used for the different channels at different positions with the characteristic parameters unchanged. The desired channels at points A, D and F were recovered by using a tunable double-notch FBG alongwith a 3–port OC. The various BER curves and optical spectra recovered at various positions of the setup with different combination of channels confirm the viability of the proposed interface for WI-DWDM mm-wave fiber radio systems, both in single as well as in cascaded configurations [6], the details of which is excluded from this discussion to avoid any repetition. However, in order to model various fiber-radio networks incorporating the proposed interfaces, a set of experimental data is recovered as per the block diagram shown in Fig. 3, which is summarized in Table 1.

IV. MODELING OF STAR-TREE NETWORKS INCORPORATING CASCADED WOIS

A generic start-tree configured WI-DWDM fiber-radio network incorporating WOIs is shown in Fig. 4. Fiber links from the CO form the 'star' part of the architecture, while the 'tree' part from the remote nodes (RNs) feeds different BSs through the respective WOIs. A unique wavelength is used to feed each BS connected by a common arm of star, with the possibility of wavelengths being reused within different arms. WOIs can be used in cascade in the RNs to enable optical-add-drop-multiplexing (OADM) functionality to the BSs, in addition to provide optical carrier in the uplink path. A single DWDM optical carrier will be used for both upstream and downstream transmission from and to a BS, and the rf signals on any DWDM carrier are those transmitted and received by the specific BS.

In the CO, a large number DWDM optical carriers are used to generate optical single sideband with carrier (OSSB+C) modulated optical mm-wave signals,

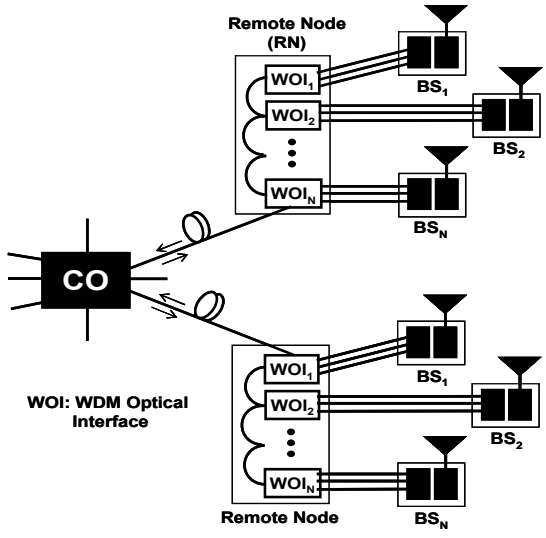


Fig. 4: Generic star-tree architecture for WI-DWDM fiber-radio network incorporating WDM optical interfaces.

combined using a suitable multiplexer and amplified before launching onto the fiber. The amplified signals will be then transported to the RNs where the composite signal will be demultiplexed by using concatenated WOIs and drop the desired downlink signals as well as the uplink optical carriers to the respective BSs. In the uplink direction, each BS will generate OSSB+C modulated optical mm-wave signal by reusing the recovered optical carrier and route it to the fiber network through the respective WOI in the RN. The fiber network then enable the uplink signals to be transported to the CO for further processing.

To model such a network based on power budget estimation, one branch of the star is simplified as Fig. 5 that clearly shows the components and subsystems contributing in power budget estimation. Shown in Fig. 5, the link is assumed to support N BSs through a single RN, where each of the BS is represented by the relevant WOI. The power budget and the power margin in the downlink direction for the Mth BS ($1 \leq M \leq N$) can be calculated by:

$$PR_{BSM} = T_{LSCO} - L_{MUX} - L_{MOD} + G_{BAMP} - L_{SMF} - (M-1)L_{ThroughWOI} - L_{DropWOI} \quad \text{.....(1)}$$

$$PM_{BSM} = PR_{BSM} - Sensitivity_{BSM} \quad \text{.....(2)}$$

where PR_{BSM} and PM_{BSM} are the received optical power and the power margin at photodetector (PD) of Mth BS

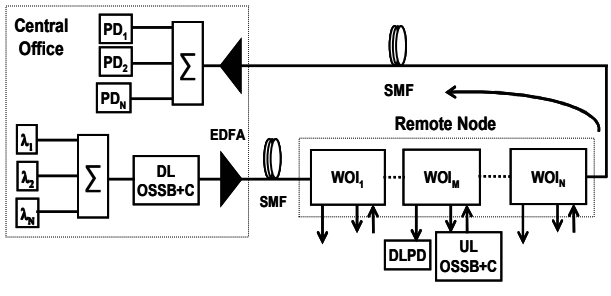


Fig. 5: A simplified branch in star-tree architecture showing the relevant components and subsystems in the CO and RN.

(BS_M), $Sensitivity_{BSM}$ is the sensitivity at the PD of BS_M, T_{LSCO} is the optical power from the respective light-source in the CO, L_{MOD} is the loss in OSSB+C modulator, G_{BAMP} is the gain from the boost-amplifier in the downlink path, L_{SMF} is the loss in fiber span between the CO and the RN, $L_{ThroughWOI}$ is the through channel loss of WOI, and $L_{DropWOI}$ is the drop-channel loss of WOI while recovering the desired downlink by the respective WOI. In this calculation, the losses in the connecting patchcords between the WOIs and the BSs are ignored due to very shorter distances. In order to solve these equations, a set of experimental data are used, which are shown in Table 1. These data are found from experimental demonstration shown in [6], which is discarded here to avoid repetition.

By using the values noted in Table 1, Equation (1) can be simplified as:

$$PR_{BSM} = -6.0 - (M-1)3.2 \quad \ldots\ldots\ldots\ldots\ldots \quad (3)$$

Therefore, received optical power at the PD of BS_1 (where $M = 1$),

$$PR_{BS1} = -6.0 \text{ (dBm)}$$

The power margin at the PD of BS_1 can be calculated by using the sensitivity of the recovered signal (shown in Table 1), which is -14.2 dBm. Therefore, power margin at the PD of BS_1,

$$PM_{BS1} = 8.2 \text{ (dB)}$$

If the power penalty is considered to add up linearly with increasing number of WOIs, and the WOIs are considered to be identical, then the number BSs supported by the link can be calculated by:

$$PM_{BS1} = (N-1)(PP_{Through} + L_{ThroughWOI}) \ldots\ldots\ldots \quad (4)$$

where N is the number of WOIs in cascade in the RN, PM_{BS1} is the power margin at the PD of BS_1, $PP_{Through}$ is the power penalty experienced by the through channels for traversing each stage of WOI, and $L_{ThroughWOI}$ is the insertion loss experienced by the through channels in a WOI.

By using the values of the parameters noted in Table 1, and Equation (4), number of WOIs in cascade can be calculated by:

$N = 1 + PM_{BS1} / (PP_{Through} + L_{ThroughWOI}) = 1 + 8.2/(0.4+3.2) = 3.28 \approx 3$ units

However, in the experiment if the lossy multiport OCs in the WOIs are replaced with standard OCs having typical through channel insertion loss (typical through loss 1dB/WOI), and typical drop channel insertion loss (typical loss 1dB/WOI), the number of units in cascade will increase to 8. Also, if the insertion of the OSSB+C generator in CO can reduced to 9 dB, the number of units in cascade will increase to 13. These results indicate that the WOI proposed in [5] can be a suitable candidate in future WI-DWDM mm-wave fiber-radio networks, configured in star-tree architecture, where cascaded

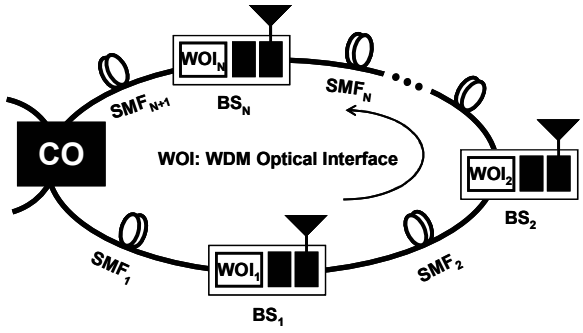


Fig. 6: Generic ring/bus architecture for WI-DWDM fiber-radio network incorporating WDM optical interfaces.

interfaces will be used in the RNs to enable OADM functionality to the BSs, in addition to provide optical carriers for the upstream transmission.

V. Modeling of Ring/Bus Networks Incorporating Cascaded WOIs

A generic ring/bus configured WI-DWDM fiber-radio network incorporating WOIs is shown in Fig. 6. This architecture allows the CO to distribute wavelengths to remote antenna BSs that are placed along the ring, with a WOI enabling OADM functionality to the relevant BS, in addition to delivering the optical carrier for upstream transmission. Each of the BSs fed from the CO has its own unique wavelength to be used for both uplink and downlink communication. In the CO, a large number optical carriers are used to generate OSSB+C modulated optical mm-wave signals, combined using a suitable multiplexer and amplified before launching onto the fiber ring. The amplified signals will be then transported along the ring where the relevant WOI will recover the downlink signal relevant to the BS and enables the through channels to be routed to the next BSs. The WOI also provides uplink optical carrier to the respective BS by recovering 50% of the optical carrier from the recovered downlink signal. In the uplink direction, each BS generates OSSB+C modulated optical mm-wave signal by reusing the recovered optical carrier and routes it to the fiber ring via the relevant WOI. The uplink signal then passes through the remaining BSs with the through channels along the ring and transported to the CO for further processing.

This architecture is typically unidirectional and the BSs

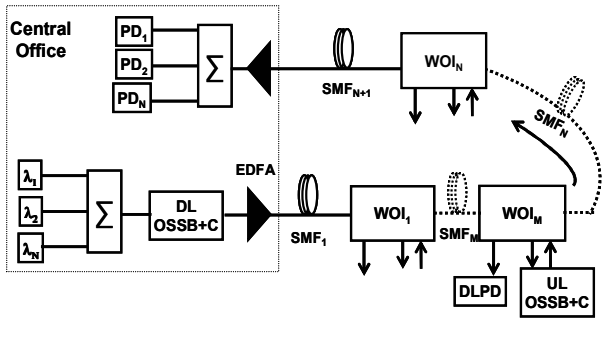


Fig. 7: Simplified optical link in ring/bus architecture showing the relevant components and subsystems in the CO and BS.

in the ring are separated typically by equal distances. It has the potential for fault restoration using second protection ring allowing a fiber break between nodes or a failure of node to be bypassed [10]. It also enables easy implementation of rf carrier reuse between the BSs, in addition to allowing dynamic frequency allocation, since frequency assignment schemes can be controlled from the CO [11].

To model such a network based on power budget estimation, the generic architecture shown in Fig. 6 can be redrawn as Fig. 7, where the components and subsystems contributing in power budget calculation are clearly shown. Similar to star-tree network, the ring is assumed to support N BSs through a single CO, where each of the BS is represented by the relevant WOI. Shown in Fig. 7, each of the WOIs is followed by a span of fiber to be connected with the neighboring WOI, which forms the ring under investigation. For simplicity, all the fiber spans are considered to be equal having a transmission attenuation of 0.2 dB/km.

The power budget and the power margin in the downlink direction for the Mth BS ($1 \leq M \leq N$) can be calculated by:

$$PR_{BSM} = T_{LSCO} - L_{MUX} - L_{MOD} + G_{BAMP} - M*L_{SMF} - (M-1)L_{ThroughWOI} - L_{DropWOI} \quad \text{.....(5)}$$

$$PM_{BSM} = PR_{BSM} - Sensitivity_{BSM} \quad \text{.....(6)}$$

where PR_{BSM} and PM_{BSM} are the received optical power and the power margin at the PD of Mth BS (BS_M), $Sensitivity_{BSM}$ is the sensitivity at the PD of BS_M, T_{LSCO} is the optical power from the respective light-source in the CO, L_{MOD} is the loss in OSSB+C modulator, G_{BAMP} is the gain from the boost-amplifier in the downlink path, L_{SMF} is the attenuation in each of the fiber span between two consecutive WOIs, $L_{ThroughWOI}$ is the through channel loss of WOI, and $L_{DropWOI}$ is the drop-channel loss of WOI while recovering the desired downlink by the respective WOI.

Equation (5) can be simplified by using the values of the parameters from the Table 1. The experiment uses 10 km SMF between to CO and the BS. To use the results from the experiment for this analysis, we consider the first span of fiber 10 km, while the others are 1 km each. After such considerations, Equation (5) can be simplified as:

$$PR_{BSM} = -6.0 - (M-1)3.4 \quad \text{..... (7)}$$

Therefore, received optical power at the PD of BS_1 (where $M = 1$):

PR_{BS1} = -6.0 (dBm).

The power margin at the PD of BS_1 can be calculated by using the sensitivity of the recovered signal, which is -14.2 dBm. Therefore, power margin at the PD of BS_1,

$$PM_{BS1} = 8.2 \text{ (dB)}$$

If the power penalty is considered to add up linearly with increasing number of WOIs and the WOIs are considered to be identical, then the number BSs supported by the link can be calculated by:

$$PM_{BS1} = (N-1)(PP_{Through} + L_{ThroughWOI}) + N(1km \times 0.2\ dB/km) \quad \text{..... (8)}$$

where N is the number of WOIs in cascade spaced by 1 km of SMF, PM_{BS1} is the power margin at the PD of BS1, $PP_{Through}$ is the power penalty experienced by the through channels for traversing each stage of WOI, and $L_{ThroughWOI}$ is the insertion loss experienced by the through channels in a WOI.

By using the values noted in Table 1, number of units in cascade can be calculated by:

$$8.2 = (N-1) \times 3.6 + 0.2N$$

$$\Rightarrow \quad N = 11.8/3.8 = 3.1 \approx 3 \text{ units}$$

However, if the lossy multiport OCs in the WOIs in the experiment are replaced with standard OCs having typical through channel insertion loss (typical through loss 1dB/WOI) and typical drop channel insertion loss (typical loss 1dB/WOI), the number of units in cascade will increase to 7. Also, if the insertion of the OSSB+C generator in CO can be reduced to 9 dB, the number of units in cascade will increase to 11. Therefore WDM optical interface proposed in [5] can be a suitable candidate in future WI-DWDM mm-wave fiber-radio networks, configured in ring/bus architecture, where the interfaces will be used along the fiber ring to enable OADM functionality to the BSs, in addition to provide optical carriers for the upstream transmission.

VI. CONCLUSION

The performance of the proposed WDM optical interface in WI-DWDM fiber-radio networks both in star-tree and ring/bus architectures is investigated. The investigation relied on the data recovered from real-time experiments. The results show that the proposed interface can be a suitable candidate in future WI-DWDM mm-wave fiber-radio networks, configured in either star-tree or ring/bus architecture, which are considered very effective in delivering future broadband wireless services to customers.

REFERENCES

[1] C H. Schmuck, R. Heidemann, and R. Hofstetter, "Distribution of 60 GHz signals to more than 1000 base stations," *Electron. Lett.,* vol. 30, pp. 59–60, Jan. 1994.

[2] H. Ogawa, D. Polifko, and S. Banba, "Millimeter wave fiber optics systems for personal radio communication," *IEEE Trans. Microwave Theory Tech.,* vol. 40, pp. 2285-2293, 1992.

[3] C. Lim, A. Nirmalathas, D. Novak, R. S. Tucker, R. B. Waterhouse, "Technique for increasing optical spectral efficiency in millimeter-wave WDM fiber-radio," *Electron. Lett.,* vol. 37, no. 16, pp. 1043–1045, 2001.

[4] A. Nirmalathas, D. Novak, C. Lim, R. Waterhouse, "Wavelength Reuse in the WDM Optical Interface of a Millimeter-Wave Fiber-Wireless Antenna Base Station,"

IEEE Trans. Microwave Theory Tech., vol. 49, pp. 2006-2012, 2001.

[5] M. Bakaul, A. Nirmalathas, and C. Lim*,* "Multifunctional WDM optical interface for millimeter-wave fiber-radio antenna base station," *J. of Lightwave Technol.,* vol. 23, no. 3, pp. 1210-1218, 2005.

[6] M. Bakaul, A. Nirmalathas, C. Lim, D. Novak, and R. Waterhouse, "Performance characterization of single as well as cascaded WDM optical interfaces in millimeter-wave fiber-radio networks," IEEE Photon. Technol. Lett., vol. 18, no. 1, pp. 115-117, 2006.

[7] G. H. Smith, D. Novak, and C. Lim, "A millimeter wave full-duplex fiber-radio star-tree architecture incorporating WDM and SCM," *IEEE Photon. Technol. Lett.,* vol. 10, pp. 1650-1652, Nov. 1998.

[8] R. Heinzelmann, T. Kuri, K. I. Kitayama, A. Stohr, D. Jager, "Optical add-drop multiplexing of 60 GHz millimeterwave signals in a WDM radio-on-fiber ring," *Proc. Conference on Optical Fiber Communication (OFC'00)*, Washington DC, USA, vol. 4 , pp. 137-139, 2000.

[9] C. Lim, A. Nirmalathas, D. Novak, and R. B. Waterhouse, "Capacity analysis for WDM fiber-radio backbones with star-tree and ring architecture incorporating wavelength interleaving," *J. Lightwave Technol.,* vol. 21, no. 12, pp. 3308-3315, 2003.

[10] B. S. Johansson, C. R. Batchellor, and L. Egnell, " Flexible bus: a self restoring optical ADM ring architecture," *Electron. Lett.,* vol. 32, pp. 2338-2339, 1996.

[11] M. Berg, S. Pettersson, and J. Zander, " A radio resource management concept for bunched personal communication systems, " *Royal Institute of Technology*," Stockholm, 1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Base Transit Time of a Bipolar Junction Transistor Considering Majority-carrier Current

Md Iqbal Bahar Chowdhury and M. M. Shahidul Hassan
Departemnt of Electrical and Electronics Engineering, United International University,
Dhaka-1209, Bangladesh. E-mail: ibchy@eee.uiu.ac.bd .
Department of Electrical nd Electronic Engineering, Bangaldesh University of Engineering and Technology,
Dhaka-1000, Bangaldesh. E-mail: shassan@eee.buet.ac.bd .

***Abstract*— In this paper an analytical expression for base transit time τ_B for an npn bipolar transistor considering majority-carrier current density is obtained. For finding τ_B, expressions for minority-carrier electron current density J_n, majority-carrier hole current density J_p and electron concentration, $n(x)$ are analytically derived. In the model energy-bandgap-narrowing effects due to heavy doping, velocity saturation as well as doping and field dependent mobility are considered. It is found that, in the low-injection condition, τ_B depends on J_p and its value is found greater than the value if J_p is neglected. In finding τ_B, previous works neglected J_p. The closed-form analytical expressions offer a clear physical insight into device operations at various bias conditions.**

I. Introduction

One of the limitations of high-speed operation of bipolar transistors is base transit time, τ_B. Therefore, rigorous efforts are made in the literature to accurately model this τ_B. Conventionally, derivation of analytical models [1]–[7] for τ_B were made by neglecting majority-carrier current in the quasi-neutral base. Liou *et al.* [8], [9] in their works considered the role of J_p on τ_B; but their model is based on simulation results of J_p instead of a closed-form analytical expression, uses position-independent transport parameters and applies iterative approach. In this paper, an analytical expression for majority-carrier current, J_p is derived for the first time in the literature. Using this expression, this model shows that τ_B is indeed affected in low injection (LI) conditions. The results show that τ_B under LI condition considering J_p is higher than that without considering J_p. This increase in τ_B is a strong function of peak doping density, slope of the doping profile and the base width.

II. Model Derivation

A. Modeling $n(x)$, $J_n(x)$ and $J_p(x)$

The electron and hole current densities in the quasi-neutral base of an npn bipolar transistor for an arbitrary base doping concentration $N_A(x)$ are given by

$$-J_n = qn(x)\mu_n(x)E_n(x) + qD_n(x)\frac{dn(x)}{dx} \quad (1)$$

$$-J_p = qp(x)\mu_p(x)E_p(x) - qD_p(x)\frac{dp(x)}{dx} \quad (2)$$

We define the directions of $J_n(x)$ and $J_p(x)$ in (1) and (2) so that they have positive values. Electric field, E_n acting on the electrons in the base is differed from the electric field, E_p acting on the holes [10] due to bandgap narrowing effect of heavy doping and are expressed as

$$E_n = E_p - \frac{V_T}{n_{ie}^2(x)}\frac{dn_{ie}^2(x)}{dx} \quad (3)$$

$$E_p = \frac{V_T}{N_A}\frac{dN_A(x)}{dx} \quad (4)$$

where $\mu_n(x)$ and $D_n(x)$ are minority electron mobility and diffusivity, respectively, $\mu_p(x)$ and $D_n(x)$ are majority hole mobility and diffusivity, respectively, $n_{ie}(x)$ is the effective intrinsic carrier concentration and V_T is the thermal voltage and is given by Einstein relation as

$$V_T = \frac{D_n(x)}{\mu_n(x)} = \frac{D_p(x)}{\mu_p(x)} = \frac{KT}{q} \quad (5)$$

The effective intrinsic carrier concentration $n_{ie}(x)$ depends on the doping profile $N_A(x)$ through bandgap narrowing effect, whereas electron mobility $\mu_n(x)$ depends on the doping profile as well as on the electric field. Although various models have been proposed in the literature [11], [12] for $n_{ie}(x)$ and $\mu_n(x)$, they can not be applicable in the differential equation to solve for minority carrier. Since base doping density $N_A(x)$ varies from 5×10^{16} to 2×10^{18} cm^{-3} in practical use [13], we use the low-field doping density dependent mobility μ_{n0} and $n_{ie}(x)$ model well approximated therein [13].

$$\mu_{n0} = \mu_n(0)\left(\frac{N_A(x)}{N_{ref}}\right)^{-\gamma_1} \quad (6)$$

$$n_{ie}^2(x) = n_{i0}^2\left(\frac{N_A(x)}{N_{ref}}\right)^{\gamma_2} \quad (7)$$

with

$$\mu_n(0) = 20.72\ cm^2\ (V.s)^{-1}$$
$$n_{i0} = 1.4 \times 10^{10}\ cm^{-3}$$
$$N_{ref} = 1.0 \times 10^{17}\ cm^{-3}$$
$$\gamma_1 = 0.42$$
$$\gamma_2 = 0.69$$

The empirical expression for field-dependent mobility $\mu_n(x)$ used in [7], originally suggested by Kull *et al.* [14] and later modified by Chen and Kuo [15] is used in this work.

$$\mu_n = \frac{v_s}{(1-\frac{1}{\alpha})|E_n| + E_c} \tag{8}$$

where $E_c = \frac{v_s}{\mu_{n0}}$ is the critical field, v_s is the saturation velocity and $\alpha = 4.43$.

The doping density profile used for this work is exponential and is given by [13]

$$N_A(x) = N_A(0)e^{-\frac{\eta x}{W_B}} \tag{9}$$

where $N_A(0)$ is the peak base doping concentration and $\eta = ln(N_A(0)/N_A(W_B))$ is the slope of base doping profile, where $N_A(W_B)$ is the doping density at $x = W_B$.

Combining equations (3), (4), (7) and (9) electric fields can be expressed as,

$$E_n = -\frac{\eta V_T}{W_B}(1-\gamma_2) \tag{10}$$

$$E_p = -\frac{\eta V_T}{W_B} \tag{11}$$

Use of equations (5), (6), (8) and (9) results in an expression for electron diffusivity as follows

$$\frac{1}{qD_n} = \frac{\eta G}{W_B} + Fu^{\gamma_1} \tag{12}$$

where

$$\begin{aligned} G &= \frac{a(1-\gamma_2)}{qv_s} \\ F &= \frac{1}{qD_n(0)}\left(\frac{N_A(0)}{N_{ref}}\right)^{\gamma_1} \\ u(x) &= e^{-\frac{\eta x}{W_B}} \\ a &= 1-\frac{1}{\alpha} \end{aligned}$$

Defining the ratio of μ_p to μ_n to be a constant, $r = \frac{\mu_p(0)}{\mu_n(0)}$, where $\mu_p(0)$ and $\mu_n(0)$ are the hole and electron mobility, respectively without considering doping and field dependency, using the quasi-neutrality condition, $p(x) = N_A(x) + n(x)$ and then combining with the equations (10), (11) and (12), the expressions for $J_n(x)$ and $J_p(x)$ become

$$-\frac{J_n}{qD_n} = -\frac{\eta(1-\gamma_2)}{W_B}n(x) + \frac{dn(x)}{dx} \tag{13}$$

$$-\frac{J_p}{qD_n} = -\frac{r\eta}{W_B}n(x) - r\frac{dn(x)}{dx} \tag{14}$$

Present day BJTs are fabricated with thinner base width and therefore, carrier recombination in the quasi-neutral base can be safely neglected [16] and we can define the constant total current density J as

$$-J = -J_n(x) - J_p(x) \tag{15}$$

Combining equations (13) and (14) with (15), we have a first-order differential equation for minority carrier concentration $n(x)$ as follows

$$-\frac{J}{qD_n(x)(1-r)} = \frac{dn(x)}{dx} - \frac{\eta b}{W_B}n(x) \tag{16}$$

where,

$$b = \frac{1-\gamma_2+r}{1-r} \tag{17}$$

The boundary condition required to solve the equation (16) can be obtained from the following relation [17], [18].

$$n(0) = \frac{\frac{n_{ie}^2(0)}{N_A(0)}e^{\frac{qV_{BE}}{KT}}}{\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{n_{ie}^2(0)}{N_A^2(0)}e^{\frac{qV_{BE}}{KT}}}} \tag{18}$$

Integrating the differential equation (16) from $x = 0$ to $x = x$ using equation (12) and (18), minority electron concentration $n(x)$ can be obtained as

$$n(x) = JA_1 + JA_2u^{\gamma_1} + Bu^{-b} \tag{19}$$

where

$$\begin{aligned} A_1 &= \frac{G}{b(1-r)} \\ A_2 &= \frac{FW_B}{\eta(b+\gamma_1)(1-r)} \\ B &= n(0) - JA_1 - JA_2 \end{aligned}$$

Since the electron velocity saturates at v_s at the base-collector junction, the electron current density $J_n(x)$ at $x = W_B$ is given by

$$J_n(W_B) = qv_sn(W_B) \tag{20}$$

Solving equation (20), results in the total current density J as

$$J = A_{low}n(0) \tag{21}$$

where

$$\begin{aligned} A_low &= \frac{m_1u_w^{-b}}{m_2A_1 + m_3A_2u_w^{\gamma_1} + m_1(A_1+A_2)u_w^{-b}} \\ m_1 &= s+b+\gamma_2-1 \\ m_2 &= 1-\gamma_2-s \\ m_3 &= 1+\gamma_1-\gamma_2-s \\ s &= \frac{v_sW_B}{\eta D_{nw}} \\ u_w &= e^{-\eta} \\ D_{nw} &= D_{n0}\left(\frac{N_A(W_B)}{N_{ref}}\right)^{-\gamma_1} \end{aligned}$$

Finally, using equation (21), expressions for electron concentration, electron current density and hole current density can be rearranged as

$$n(x) = J(A_1 + A_2u^{\gamma_1} + A_3u^{-b}) \tag{22}$$

$$J_n(x) = \frac{J}{1-r}\frac{A_{11} + A_{12}u^{\gamma_1} + A_{13}u^{-b}}{bA_1 + A_2(b+\gamma_1)u^{\gamma_1}} \tag{23}$$

$$J_p(x) = \frac{rJ}{1-r}\frac{A_1 + (1-\gamma_1)A_2u^{\gamma_1} + (b+1)A_3u^{-b}}{bA_1 + A_2(b+\gamma_1)u^{\gamma_1}} \quad (24)$$

where

$$\begin{aligned} A_{11} &= (1-\gamma_2)A_1 \\ A_{12} &= (1+\gamma_1-\gamma_2)A_2 \\ A_{13} &= (1-b-\gamma_2)A_3 \\ A_3 &= \frac{1}{A_{low}} + A_1 + A_2 \end{aligned}$$

B. Modeling of τ_B

Base transit time τ_B for an npn BJT can be defined as

$$\tau_B = q\int_0^{W_B}\frac{n(x)}{J_n(x)}dx \quad (25)$$

Therefore, combining equations (22) and (23) with the above equation, we have

$$\tau_B = -K\int_1^{u_w}\frac{(A_1u^b + A_2u^{b+\gamma_1} + A_3)(D_1 + D_2u^{\gamma_1})}{u(1 + B_1u^b + B_2u^{b+\gamma_1})}du \quad (26)$$

where

$$\begin{aligned} B_1 &= \frac{A_{11}}{A_{13}} \\ B_2 &= \frac{A_{12}}{A_{13}} \\ D_1 &= bA_1 \\ D_2 &= A_2(b+\gamma_1) \\ K &= \frac{q(1-r)W_B}{\eta A_{13}} \end{aligned}$$

Unfortunately, closed form expression of the integral equation (26) is not analytically tractable. Considering the inequalities, $\frac{1}{B_1u^b+B_2u^{b+\gamma_1}} < 1$ and $\frac{B_1}{B_2}u^{-\gamma_1} < 1$ and taking the first order approximation of the denominator polynomial, a closed form expression of τ_B can be obtained as follows

$$\tau_B = \frac{K}{B_2^2}\sum_{m=1}^{2}\frac{(-1)^m}{B_2^{m-1}}(N_{a,mi} + N_{b,mi} + N_{c,mi}) \quad (27)$$

where

$$\begin{aligned} N_{a,mi} &= \sum_{i=1}^{3}\frac{A_{mi}}{a_{mi}}(u_w^{a_{mi}} - 1) \\ N_{b,mi} &= \sum_{i=1}^{3}\frac{B_{mi}}{b_{mi}}(u_w^{b_{mi}} - 1)|_{m\neq1,i\neq3} + \eta B_{m3} \\ N_{c,mi} &= \sum_{i=1}^{3}\frac{C_{mi}}{c_{mi}}(u_w^{c_{mi}} - 1)|_{m\neq1,i\neq2} + \eta C_{m2} \end{aligned}$$

where

$$\begin{aligned} A_{mi} &= -mB_1C_{2i-1} \\ B_{mi} &= B_2C_{2i-1} - mB_1C_{2i} \\ C_{mi} &= B_2C_{2i} \\ a_{mi} &= -m(b+\gamma_1) + g_i - \gamma_1 \\ b_{mi} &= -m(b+\gamma_1) + g_i \\ c_{mi} &= -m(b+\gamma_1) + g_i + \gamma_1 \\ g(1) &= 0, \quad g(2) = b, \quad g(3) = b+\gamma_1 \\ C(1) &= bA_1A_3, \quad C(2) = (b+\gamma_1)A_2A_3 \\ C(3) &= bA_1^2, \quad C(4) = (2b+\gamma_1)A_1A_2 \\ C(5) &= 0, \quad C(6) = (b+\gamma_1)A_2^2 \end{aligned}$$

III. Results

A. Carrier and Current Density Profile

Fig. (1) shows the electron concentration profile for three peak doping densities, $N_A(0) = 5\times10^{17}\ cm^{-3}$, $N_A(0) = 1\times10^{18}\ cm^{-3}$ and $N_A(0) = 2\times10^{18}\ cm^{-3}$ for $V_{BE} = 0.65V$ with and without considering J_p. From this figure, it is evident that $J_p \to 0$ assumption underestimates electron concentration throughout the base region and this underestimation is more pronounced, as we move closer to the B-C junction, regardless of $N_A(0)$.

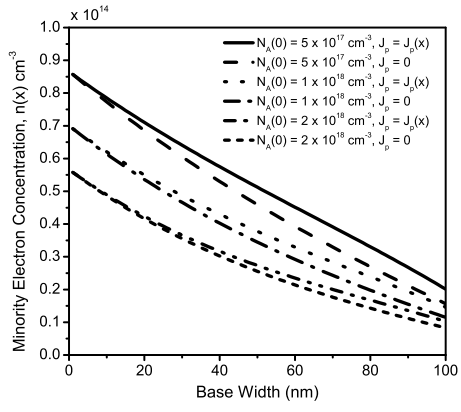


Fig. 1. Electron Concentration, n(x) with and without considering J_p

Hole current densities, $J_p(x)$ for three peak doping densities, $N_A(0) = 5\times10^{17}\ cm^{-3}$, $N_A(0) = 1\times10^{18}\ cm^{-3}$ and $N_A(0) = 2\times10^{18}\ cm^{-3}$ with $V_{BE} = 0.65\ V$ are shown in the Fig. 2. The profiles shows that J_p is almost constant near the B-E junction, whereas rapidly decreases near the B-C junction, which is consistent with the simulation results mentioned in [8], [9]. Since the total current density is the sum of electron and hole current densities and is assumed to be constant throughout the quasi-neutral base, the minority carrier electron current density should show the inverse profile of hole current density. This fact is verified in the Fig. 3. Moreover, from these two figures, it is seen that J_p goes to

zero and hence, J_n becomes constant before approaching the B-C junction at $x = W_B$ as $N_A(0)$ decreases. J_p going down to zero before $x = W_B$ is also observed in [8], [9].

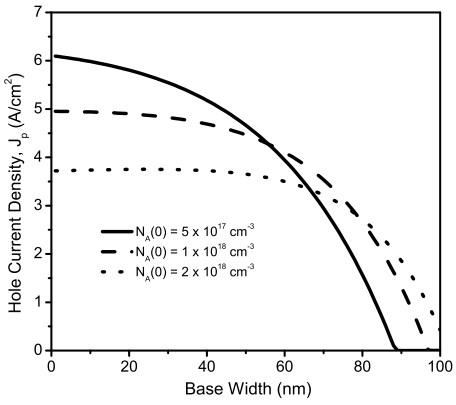


Fig. 2. Hole current density for $_A(0) = 5 \times 10^{17}$ $^{-3}$, $_A(0) = 1 \times 10^{18}$ $^{-3}$ and $_A(0) = 2 \times 10^{18}$ $^{-3}$

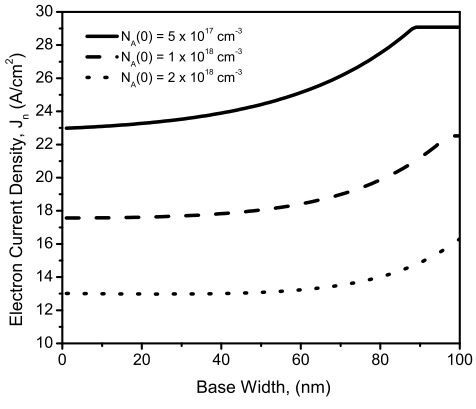


Fig. 3. Electron current density considering $_p$ for $_A(0) = 5 \times 10^{17}$ $^{-3}$, $_A(0) = 1 \times 10^{18}$ $^{-3}$ and $_A(0) = 2 \times 10^{18}$ $^{-3}$

Fig. 4 depicts the exponential variation of total current density when V_{BE} is varied maintaining LI condition. This is confirmed by the fact that total current is the function of $n(0)$ only, which in turn exponentially depends on V_{BE}. Again, it is evident from the figure that, $J_p \rightarrow 0$ assumption results in underestimation of total current density.

B. Base Transit Time

In this paper, an approximate closed form expression for τ_B is formulated. τ_B calculated using this expression, using numerical methods and using $J_p \rightarrow 0$ approximation used conventionally are plotted in Fig. 5. The approximation closely follows the numerical results as seen in the figure, showing the accuracy of the present model. The figure also points out the inaccuracies involved, if $J_p \rightarrow 0$ assumption is used. As is explained in the previous works [7], [9], the figure

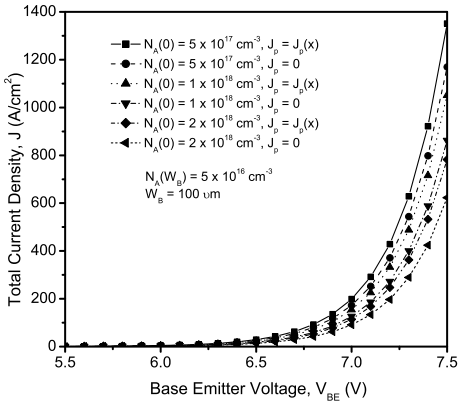


Fig. 4. Total Current Density vs. Base Emitter Voltage with and without considering $_p$ $_A(0) = 5 \times 10^{17}$ $^{-3}$, $_A(0) = 1 \times 10^{18}$ $^{-3}$ and $_A(0) = 2 \times 10^{18}$ $^{-3}$

also portrays that τ_B decreases if peak doping concentration, $N_A(0)$ increases.

Fig. 6 plots the relative error of our approximation with respect to the numerical analysis and with respect to $J_p \rightarrow 0$ assumption. It is seen that the present model shows error within $(1 - 2\%)$ compared to the numerical analysis and is less error-prone for higher peak doping densities. The figure also gives evidences why we should consider nonzero (J_p) for modeling, as an error of $(5 - 9\%)$ results in if $J_p \rightarrow 0$ assumption is used.

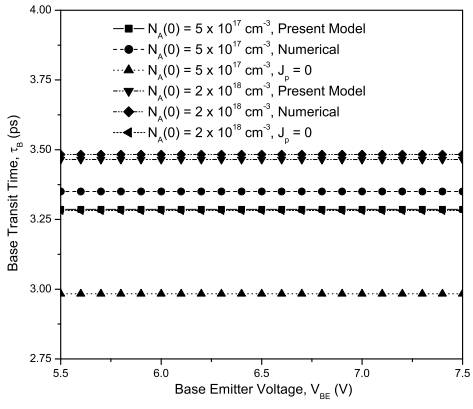


Fig. 5. Base Transit Time with and without considering $_p$ for $_A(0) = 5 \times 10^{17}$ $^{-3}$ and $_A(0) = 2 \times 10^{18}$ $^{-3}$

For a given slope of doping profile, base transit time increases with peak doping concentration. Fig. 7 verifies this fact for varying $N_A(0)$ with different η. The figure verifies that the present model more and more closely follows the numerical analysis for higher $N_A(0)$ and η. The relative error between the present model and the $J_p \rightarrow 0$ assumption decreases, as seen in the figure, if $N_A(0)$ and η increases. Fig. 8 and 9 show the base transit time calculated using the present approximate

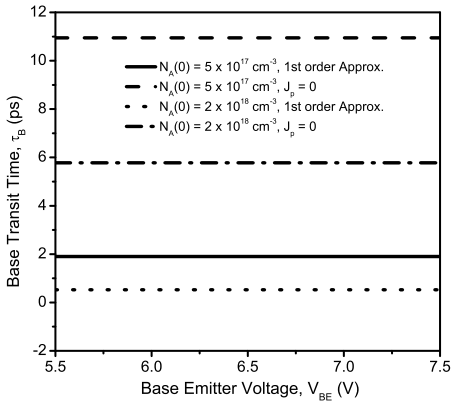


Fig. 6. Relative error in base transit time of the present model and conventional approximation $_p = 0$ w.r.t. numerical methods with $_A(0) = 5 \times 10^{17}$ $^{-3}$ and $_A(0) = 2 \times 10^{18}$ $^{-3}$

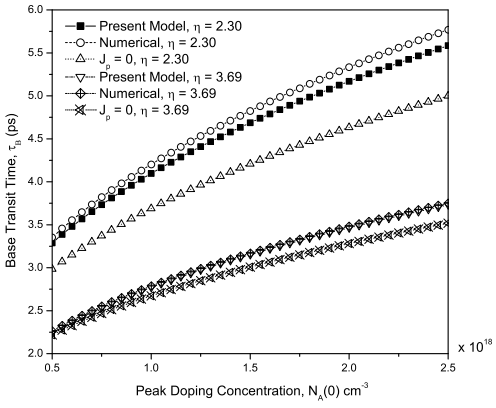


Fig. 7. Base Transit Time vs. Peak doping density for $= 2\ 30$ and $= 3\ 69$

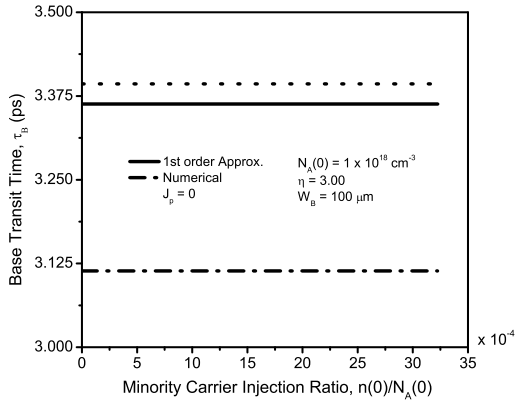


Fig. 8. Base Transit Time vs. minority carrier injection ratio

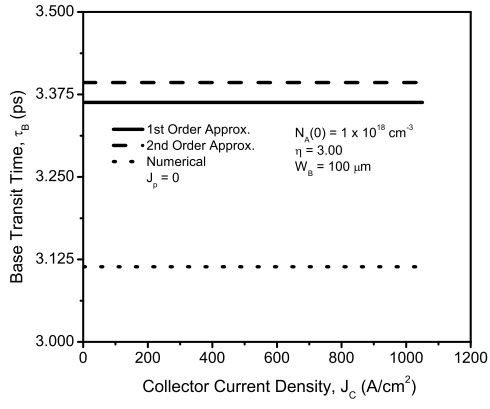


Fig. 9. Base Transit Time vs. collector current density

model, using numerical analysis and $J_p \rightarrow 0$ assumption for minority carrier injection ratio variation and collector current density variation, respectively. These figures also verify that our model closely follow the numerical analysis and that our model is less erroneous than the conventional $J_p \rightarrow 0$ model.

IV. Conclusion

Analytical expressions for electron concentration profile, electron current density and base transit time considering majority carrier current density, J_p have been derived in this paper. The model also includes bandgap narrowing effects due to heavy doping and considers velocity saturation and doping and field dependent mobility. It has been shown that considering J_p in the quasi-neutral base, under low injection, results in higher base transit time, τ_B. This is due to the retarding electric field caused by J_p. Considering J_p also results in position dependent J_n, thereby making difficult to obtain an analytically tractable closed-form expression for τ_B. Therefore, this work approximates an analytical expression for τ_B and shows that this model is within $1 - 2\%$ of the numerically calculated values.

References

[1] J. J. H. van der Beisen, "A simple regional analysis of transit times in bipolar transistors", *Solid State Electron.*, vol. 29, no. 5, pp. 529-534, May 1986.

[2] J. Weng, "A physical model of the transit time in bipolar transistors", *Solid State Electron.*, vol. 36, no. 9, pp. 1197-1201, Aug. 1993.

[3] K. Suzuki, "Analytical base transit time model for high-injection regions", *Solid State Electron.*, vol. 37, no. 3, pp. 487-493, Mar. 1994.

[4] M. Pingxi, L. Zhang, and Y. Wang, "Analytical model of collector current density and base transit time based on iteration method", *Solid State Electron.*, vol. 39, no. 11, pp. 1683-1686, Nov. 1996.

[5] M. Pingxi, L. Zhang, and M. Ostling,, "A new set of initial conditions for fast and accurate calculation of base transit time and collector current density in bipolar transistors", *Solid State Electron.*, vol. 41, no. 11, pp. 2023-2026, Nov. 1998.

[6] M. Z. R. Khan, M. M. Shahidul Hassan, T. Rahman and A. K. M. Ahsan, "Expression for base transit time in bipolar transistors", *Int. J. Electron.*, vol. 92, no. 4, pp. 215-229, Apr. 2005.

[7] M. M. Shahidul Hassan and Md. Waliullah Khan Nomani, "Base-transit-time model considering field dependent mobility for BJTs operating at high-level injection", *IEEE Trans. Electron Devices*, vol. 53, no. 10, pp. 2532-2539, Oct. 2006.

[8] J. J. Liou and Y. Yue, "High-level free-carrier injection in advance bipolar junction transistors (invited)", *Solid State Electron*, vol. 29, no. 5, pp. 174-182, May 1994.
[9] Y. Yue, J. J. Liou, A. Ortiz-Conde and F. Garcia Sanchez, "Effects of high-level free-carrier injection on the base transit time of bipolar junction transistors", *Solid State Electron.*, vol. 39, no. 1, pp. 27-31, Jan. 1996.
[10] R. J. van Overstraeten, H. J. Deman and R. P. Mertens, "Transport equations in heavy doped silicon", *IEEE Trans. Electron Devices*, vol. ED-20, no. 3, pp. 290-298, Mar. 1973.
[11] W. L. Engl, *Process and Device Modeling*, Amsterdam, The Netherlands: Elsevier, 1986.
[12] H. C. Graaff and F. M. Klaassen, *Compact Transistor Modeling for Circuit Design*, Vienna, Austria: Springer-Verlag, 1990.
[13] K. Suzuki, "Optimum base-doping profile for minimum base transit time considering velocity saturation at base-collector junction and dependence of mobility and bandgap narrowing on doping concentration", *IEEE Trans. Electron Devices*, vol. 48, no. 9, pp. 2102-2107, Sep. 2001.
[14] G. M. Kull, W. Nagel, S. W. Lee, P. Lloyd, E. J. Prendergast and H. Dirks, "A unified circuit model for bipolar transistors including quasi-saturation effect", *IEEE Trans. Electron Devices*, vol. ED-32, no. 6, pp. 1103-1113, Jun. 1985.
[15] B. Y. Chen and J. B. Kuo, "An accurate knee current model considering quasi-saturation for BJTs operating at high current density", *Solid State Electron*, vol. 38, no. 6, pp. 1282-1284, Jun. 1995.
[16] J. S. Yuan, "Effect of base profile on the base transit time of bipolar transistors for all levels of injection", *IEEE Trans. Electron Devices*, vol. 41, no. 2, pp. 212-216, Feb. 1994.
[17] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., New York: Wiley, 1986.
[18] W. M. Webster, "On the variation of junction-transistor current-amplifier factor with emitter current", *Proc. IRE*, vol. 42, pp. 914-921, 1954.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Analytical MOSFET Model Including Gate Voltage Dependence of Channel Length Modulation Parameter for 20nm CMOS

Akira Hiroki, Akihiro Yamate, and Masayoshi Yamada

Graduate School of Science and Technology, Department of Electronics, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto, 606-8585, Japan
E-mail: hiroki@kit.ac.jp

Abstract – This paper describes an analytical MOSFET model for 20nm CMOS. The model includes the gate voltage dependence of the channel length modulation parameter. It is found that the channel length modulation parameter extracted from experimental data has remarkable gate voltage dependence in sub 65nm region. The dependence has been successfully modeled and included in an analytical MOSFET model. This model can predict the current – voltage characteristics with in good accuracy for n-channel and p-channel MOSFETs down to 20nm.

I. Introduction

In circuit designs, analytical MOSFET current models have been used to analyze circuit behaviour. The analytical MOSFET current models as well as circuit simulations are indispensable tools in modern VLSI design. The alpha power-law MOS model [1], [2], one of the analytical MOS current models, proposed for quarter micrometer MOSFETs. The model is a simple yet practical drain current model because of analytical equations and few model parameters. The model has been widely used for analytical treatments of circuit behaviour. The model has been applied to MOSFETs down to 100 nanometer region due to its accuracy and simplicity. Also, the model applied to the digital circuit analysis of the power consumption and the optimization of the output characteristics for 70nm CMOS design [3]. Recently, the gate length of MOSFETs in VLSI's has scaled down to 65nm and below. Analog circuits have been embedded in a VLSI with digital circuit systems. It is needed to investigate the accuracy of the alpha power-law MOS model in such areas of 10s of nanometers.

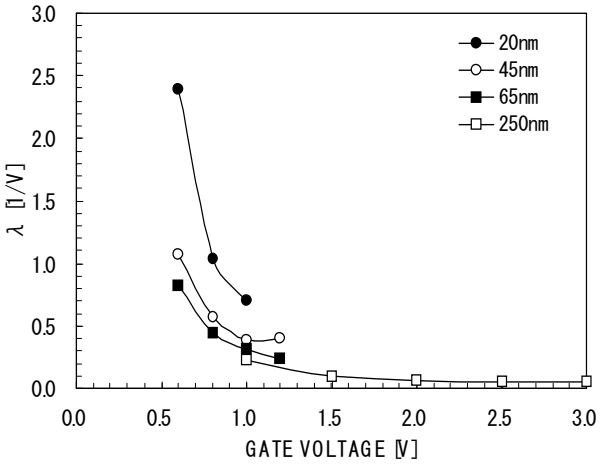
This paper investigates modelling of the drain current characteristics of sub 65nm MOSFETs. It is found that the gate voltage dependence of the channel length modulation parameter becomes remarkable as the gate length is scaled down to 20nm for both n-channel and p-channel MOSFETs. The channel length modulation parameter, which characterizes the slop of the saturation drain currents and is associated with the drain conductance, is an essential parameter in both digital and analog circuit designs. The gate voltage dependence of the channel length modulation parameter is successfully modelled and incorporated into an analytical MOSFET current model. The present MOSFET model is compared the conventional alpha power-law MOS model in order to investigate its accuracy in sub 65nm region. It is found that the present model can predict the current – voltage characteristics with in good accuracy for n-channel and p-channel MOSFETs down to 20nm region.

II. Gate voltage dependence of the channel length modulation parameter

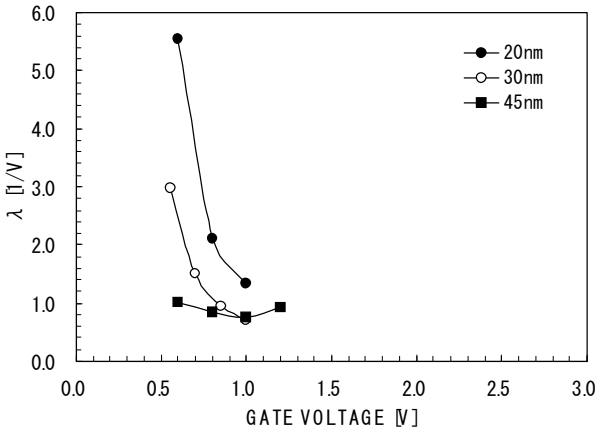
To investigate the gate voltage dependence of the channel length modulation parameter, the drain conductance is decomposed into two parts. One is a parameter associated with the magnitude of the drain currents. The other is a parameter associated with the slop of the drain current.

$$g_D = \frac{\partial I_D}{\partial V_{DS}} = I_{DSAT}\,\lambda \qquad (1)$$

where , g_D is the drain conductance, I_D is the drain current and V_{DS} is the drain voltage. The parameter I_{DSAT} is a parameter associated with the magnitude of the drain current and is defined as a drain current extrapolated at the drain bias equals zero. The parameter λ is the channel length modulation parameter associated with the slop of the drain current. The saturation drain current I_{DSAT} and the channel length modulation parameter λ are extracted from the experimental drain currents in saturation region for various kinds of MOSFETs. The test devices used in this paper are n-channel MOSFETs with a 20nm gate length [4], with a 45nm gate length [5], and with a 65nm gate length [6]. For p-channel MOSFETs, the devices are with a 20nm gate length [4], with a 30nm gate length [7], and with a 45nm gate length [5]. For each gate voltage, the parameters I_{DSAT} and λ are extracted and the gate voltage dependence of the parameters is investigated in detail.



(a)



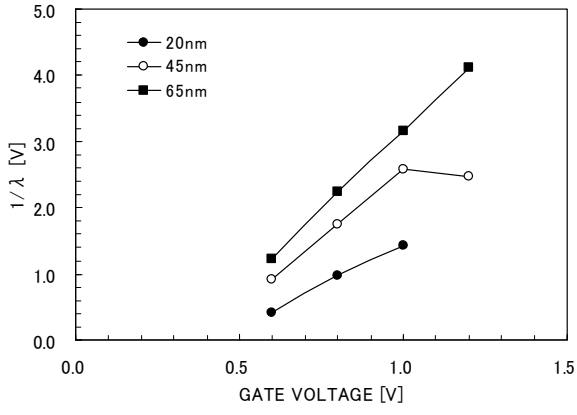
(b)

Fig. 1 Gate voltage dependence of the channel length modulation parameter λ extracted from experimental data of drain currents. (a) n-channel MOSFETs. (b) p-channel MOSFETs.

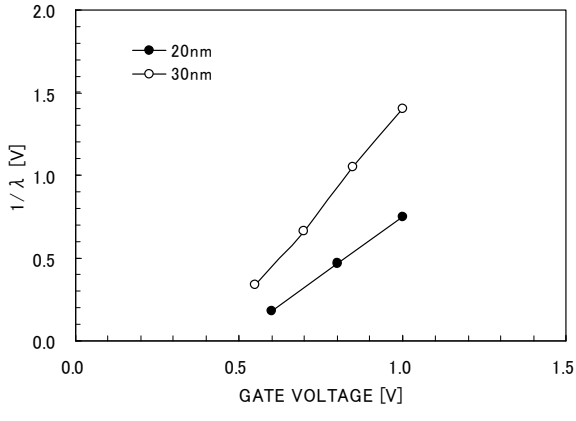
Figure 1 (a) and (b) shows the gate voltage dependence of the channel length modulation parameter λ for n-channel and p-channel MOSFETs, respectively. For comparison the parameter λ for an n-channel MOSFET with a 250nm gate length [2] is also shown in Figure 1 (a). The parameter λ extracted from the 250nm gate-length device has little gate voltage dependence in wide gate voltage region from 1V to 3V. This result suggests that the alpha power – law MOS model has good accuracy in such short channel regions, because the parameter λ in the alpha power – law MOS model has no gate voltage dependence. On the other hand, the parameters λ extracted from the 20, 45, and 65nm n-channel devices indicate significant gate voltage dependence. The larger gate voltage dependence is found as the gate length is scaled down to 20nm. For 20nm device, the parameter λ at a gate voltage of 0.6V is three times larger than that at a gate voltage of 1.0V. For p-channel MOSFETs, the parameters λ of the 20 and 30nm gate length devices show significant gate voltage dependence, while the parameter λ of the 45nm gate length device shows little gate voltage dependence. For 20nm gate length device, the parameter λ at a gate voltage of 0.6V is four times larger than that at a gate voltage of 1.0V. For both n-channel and p-channel devices, the gate voltage dependence becomes significant as the gate lengths are scaled down to 20nm region. This result suggests that the modelling of the parameter λ as a function of the gate voltage is needed for 20nm MOS circuit analysis.

III. Modelling of the gate voltage dependence of the channel length modulation parameter

To model the gate voltage dependence of the channel length modulation parameter λ , the inverse of the parameter λ is calculated form the experimental data. Figure 2 (a) and (b) shows the inverse of the parameter λ extracted from the experimental data of 20, 45, and 65nm gate length devices for n-channel MOSFETs, and 20 and 30nm gate length devices for p-channel MOSFETs, respectively.



(a)



(b)

Fig. 2 Gate voltage dependence of the inverse of the channel length modulation parameter λ extracted from experimental data. (a) n-channel MOSFETs. (b) p-channel MOSFETs.

The calculated values for both n-channel and p-channel MOSFETs indicate the linear dependence of the gate

voltage. This result strongly suggests that the inverse of the channel length modulation parameter λ can be modelled as a linear function of the gate voltage. For 45nm gate length n-channel MOSFET, the data of the inverse of the parameter λ at a gate voltage of 1.2V shows some difference from the linear function. The effect of this difference on the drain current is discussed in the next section.

In this work, the gate voltage dependence of the channel length modulation parameter λ is proposed as follows,

$$\lambda(V_{GS}) = \lambda_0 (1 + \gamma V_{GS})^{-1} \qquad (2)$$

where, V_{GS} is the gate voltage. The parameter λ_0 is a factor which indicates the parameter λ without gate voltage. The parameter γ is a factor which indicates the gate voltage dependence of the parameter λ.

IV. An Analytical MOSFET Current Model for 20nm CMOS

To investigate the accuracy of the present model, the MOSFET drain currents are calculated by using the following MOSFET current model, which is extended from the alpha power – law model [2],

$$I_D = I_{D5} = I_{DSAT}(1 + \lambda(V_{GS})V_{DS}) \qquad (3)$$

$$I_D = I_{D3} = I_{D5}\left(2 - \frac{V_{DS}}{V_{DSAT}}\right)\frac{V_{DS}}{V_{DSAT}} \qquad (4)$$

where, I_{D5} and I_{D3} are drain currents in the saturation region and linear region, respectively. The channel length modulation parameter λ is a function of the gate voltage, as shown in equation (2), instead of a constant parameter in the alpha power-law model. I_{DSAT} is the saturation drain current and V_{DSAT} is the saturation drain voltage defined as following equations,

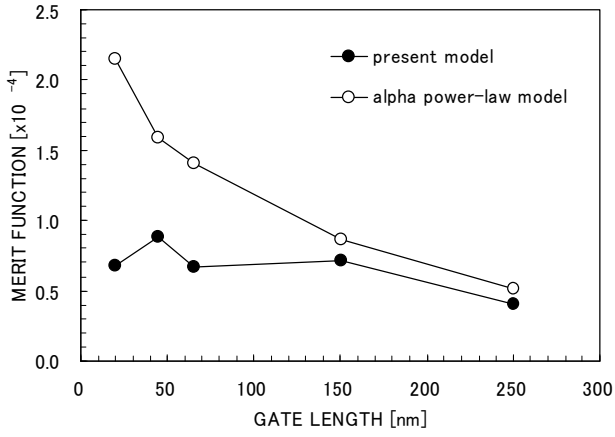
$$I_{DSAT} = \frac{W}{L} B(V_{GS} - V_{TH})^{\alpha} \qquad (5)$$

$$V_{DSAT} = K(V_{GS} - V_{TH})^{m} \qquad (6)$$

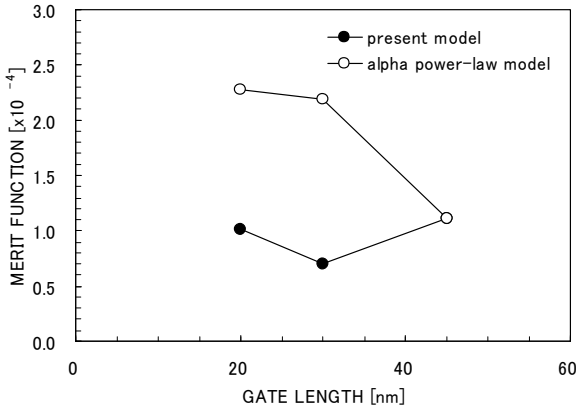
where, L and W are the gate length and the gate width, respectively. V_{TH} is a model parameter associated with the threshold voltage. The parameters B, α, K, and m are model parameters. These model parameters and the parameters λ_0 and γ in equation (2) are determined so that a merit function is minimized by using the Levenberg – Marquardt method [8]. The following normalized χ^2 merit function is used,

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{I_D(V_{DS,i}, V_{GS,i}) - I_{D,i}}{I_{DMAX}}\right)^2 \qquad (7)$$

where, N is a number of the experimental data of the drain currents. I_D ($V_{DS,i}$,$V_{GS,i}$) is the drain current calculated by using the present model (2)-(6) at a drain voltage of $V_{DS,i}$ and a gate voltage of $V_{GS,i}$. $I_{D,i}$ is the experimental drain current. The differences between the present model and the experimental data are normalized by the maximum drain current I_{DMAX} for each test devices in order to compare the merit functions for various test devices.



(a)



(b)

Fig. 3 Optimized merit function of the present model compared with that of the alpha power-law model. (a) n-channel MOSFETs. (b) p-channel MOSFETs.

Figure 3 (a) and (b) shows the optimized merit functions for n-channel and p-channel MOSFETs, respectively. The test devices are n-channel MOSFETs with a 20nm gate length [4], with a 45nm gate length [5], with a 65nm gate length [6], with a 150nm gate length [9], and with a 250nm gate length [2]. For p-channel MOSFETs, the test devices are with a 20nm gate length [4], with a 30nm gate length [7], and with a 45nm gate length [5]. In order to investigate the significance in modelling of the gate voltage dependence of the channel length modulation parameter, the results using the alpha power - law model are also shown. For n-channel MOSFETs, the merit function of the alpha power-law model increases as the gate length decreases. The merit function at 20nm gate length is 4 times larger than that at 250nm gate length. On the other hand, the merit function of the present model maintains its values even if the gate length becomes down to 20nm. There is a small increase at 45nm gate length. This is because the difference of the parameter λ from

the linear function as shown in Figure 2 (a). For p-channel MOSFETs, the merit function using the present model maintains its small values, while the merit function of the alpha power-law model increases as the gate length is scaled down to 20nm. For 45nm gate length p-channel MOSFET, both merit functions of the present model and the alpha power-law model show same and small values. This is consistent with the fact that the parameter λ of the 45nm gate length p-channel MOSFET has no significant gate voltage dependence as shown in Figure 1 (b). This result means that the gate voltage dependence of the channel length modulation parameter is essential in modelling of the MOSFET drain current in 20nm region.

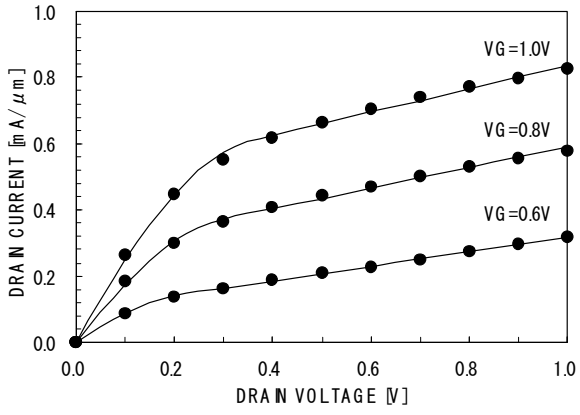
V. Drain Current – Voltage Characteristics of 20nm gate length CMOS

In this section, the present analytical MOSFET model is evaluated by comparing with the experimental drain current – voltage characteristics. For both n-channel and p-channel MOSFETs with a 20nm gate length, drain currents are calculated by using the present analytical MOSFET model. Each model parameter set is optimized so that the merit function is minimized as discussed in the previous section. Table 1 lists the optimized model parameter set for n-channel and p-channel MOSFETs.

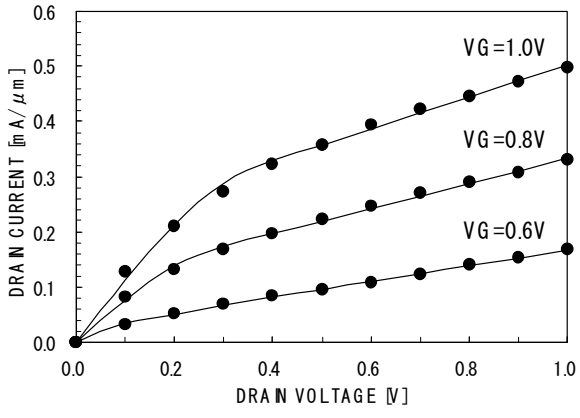
Table 1 Optimized model parameters for 20nm CMOS.

Parameters	n-MOSFET	p-MOSFET
B [$\times 10^{-5}$]	2.025	1.100
V_{TH}	0.468	0.475
α	1.162	1.480
K	0.469	0.555
m	0.416	0.614
λ_0	-0.996	-1.557
γ	-2.383	-2.134

Using these model parameters, the drain currents calculated by using the present analytical model are compared with the experimental data in Figure 4 (a) and (b). In both n-channel and p-channel MOSFETs, the results of the present model are consistent with the experimental data in wide range of drain voltage from the linear to the saturation region. Moreover, the present model predicts the slope of the drain current in the saturation region in the range of the gate voltage from 0.6V to 1.0V. This is because that the present model includes the gate voltage dependence of the channel length modulation parameter. A slight discrepancy at the low drain voltage in the linear region is observed. However, the difference in such low voltage region is negligible for circuit design. For other test devices used in this work, we also calculate the drain current – voltage characteristics using the present model and compared with the experimental data. It is found that the present model can predict the drain voltage characteristics in good accuracy in the wide range of the gate length from



(a)



(b)

Fig. 4 Drain current – voltage characteristics calculated by using the present analytical MOSFET model for 20nm gate length MOSFETs. The present model (lines) is compared with the experimental data (symbols). (a) n-channel MOSFETs. (b) p-channel MOSFETs.

20 to 250nm. This result shows that the present model allows excellent prediction of the current – voltage characteristics for both n-channel and p-channel MOSFETs down to 20nm.

VI. Conclusion

The modelling of the drain current characteristics of 20nm gate length MOSFETs has been investigated. The gate voltage dependence of the channel length modulation parameter is investigated for n-channel and p-channel MOSFETs with sub 65nm gate length. It is found that the channel length modulation parameter has remarkable gate voltage dependence as the gate length is scaled down to 20nm. A simple MOSFET current model is proposed to include the gate voltage dependence of the channel length modulation parameter. The accuracy of the present model has been evaluated by comparing with the experimental data of the drain current – voltage characteristics. The present model allows excellent prediction of the current – voltage characteristics for n-channel and p-channel MOSFETs down to 20nm gate length.

References

[1] T. Sakurai and A.R. Newton, "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas," Journal of Solid-State Circuits, Vol.25, No.2, pp.584-594, April 1990.

[2] T. Sakurai and A.R. Newton, "A Simple MOSFET Model for Circuit Analysis," IEEE Tran. Electron Devices, Vol.38, No.4, pp.887-894, April 1991.

[3] B. Amelifard, F. Fallah, and M. Pedram, "Low-power fanout optimization using multiple threshold voltage inverters," Proc. International Symposium on Low Power Electronics and Design, pp.95-98, August 2005.

[4] H.-Y. Chen, C.-Y. Chang, C.-C. Huang, T.-X. Chung, S. D. Liu, J.-R. Hwang, Y.-H. Liu, Y.-J. Chou, H.-J. Wu, K.-C. Shu, C.-K. Huang, J.-W. You, J.-J. Shin, C.-K. Chen, C.-H. Lin, J.-W. Hsu, B.-C. Perng, P.-Y. Tsai, C.-C. Chen, J.-H. Shieh, H.-J. Tao, S.-C. Chen, T.-S. Gau, and F.-L. Yang, "Novel 20nm Hybrid SOI/bulk CMOS Technology with 0.183 μ m^2 6T-SRAM Cell by Immersion Lithography," 2005 Symposium on VLSI Technology Digest of Technical Papers, pp.16-17, June 2005.

[5] F. Arnaud, B. Duriez, B. Tavel, L. Pain, J. Todeschini, M. Jurdit, Y. Laplanche, F. Boeuf, F. Salvetti, D. Lenoble, J.P. Reynard, F. Wacquant, P. Morin, N. Emonet, D. Barge, M. Bidaud, D. Ceccarelli, P. Vannier, Y. Loquet, H. Leninger, F. Judong, C. Perrot, I. Guilmeau, R. Palla, A. Beverina, V. DeJonghe, M. Broekaart, V. Vachellerie, R.A. Bianchi, B. Borot, T. Devoivre, N. Bicais, D. Roy, M. Denais, K. Rochereau, R. Difrenza, N. Planes, H. Brut, L. Vishnobulta, D. Reber, P. Stolk, and M. Woo, "Low Cost 65nm CMOS Platform for Low Power and General Purpose Applications," 2004 Symposium on VLSI Technology Digest of Technical Papers, pp.10-11, June 2004.

[6] C.H. Chen, C.S. Chang, C.P. Chao, J.F. Kuan, C.L. Chang, S.H. Wang, H.M. Hsu, W.Y. Lien, Y.C. Tsai, H.C. Lin, C.C. Wu, C.F. Huang, S.M. Chen, P.M. Tseng, C.W. Chen, C.C. Ku, T.Y. Lin, C.F. Chang, H.J. Lin, M.R. Tsai, S. Chen, C.F. Chen, M.Y. Wei, Y.J. Wang, J.C.H. Lin, W.M. Chen, C.C. Chang, M.C. King, C.M. Huang, C.T. Lin, J.C. Guo, G.J. Chern, D.D. Tang, and J.Y.C. Sun, "A 90nm CMOS MS/RF based Foundry SOC Technology Comprising Superb 185 GHz f_T RFMOS and Versatile, High-Q Passive Components for Cost/Performance Optimization," International Electron Devices Meeting Digest of Technical papers, pp.39-42, December 2003.

[7] F.-L. Yang, C.-C. Huang, C.-C. Huang, T.-X. Chung, H.-Y. Chen, C.-Y. Chang, H.-W. Chen, C.-H. Lee, S.-D. Liu, K.-H. Chen, C.-K. Wen, S.-M. Cheng, C.-T. Yang, L.-W. Kung, C.-L. Lee, Y.-J. Chou, F.-J. Liang, L.-H. Shiu, J.-W. You, K.-C. Shu, B.-C. Chang, J.-J. Shin, C.-K. Chen, T.-S. Gau, P.-W. Wang, B.-W. Chan, P.-F. Hsu, J.-H. Shieh, S.K.-H. Fung, C.H. Diaz, C.-M.M. Wu, Y.-C. See, B.J. Lin, M.-S. Liang, J.Y.-C. Sun, and C. Hu, "45nm Node Planar-SOI Technology with 0.296 μ m^2 6T-SRAM Cell," 2004 Symposium on VLSI Technology Digest of Technical Papers, pp.8-9, June 2004.

[8] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, *NUMERICAL RECIPES in C*, Cambridge University Press, 1988.

[9] C. Diaz, M.Cox, W. Greene, F. Perlaki, E. Carr, I. Manna, A. Bayoumi, M. Cao, N. Shamma, M. Tavassoli, C. Chi, N. Farrar, D. Lefforge, Y. Chang, B. Langley, and P. Marcoux, "A Novel Low-Temperature Gate Oxynitride for CMOS Technologies," 1997 Symposium on VLSI Technology Digest of Technical Papers, pp.49-50, June 1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A comparison of Quantum Mechanical Corrections in Surface Potential Based MOSFET Compact Models

M.A. Karim[1] and Anisul Haque[2]

[1,2]Depertment of Electrical and Electronic Engineering, East West University, Dhaka 1212, Bangladesh
E-mail: [1]mak@ewubd.edu, [2]ahaque@ewubd.edu

Abstract - Compact models of MOSFETs are important for simulation of electronic circuits. Surface potential based compact models have become popular for sub 100 nm MOSFETs. However, these models are based on semi classical analysis. Quantum mechanical effects, important in nano-MOS devices, are added to the surface potential based models separately as corrections. In this study we have investigated the accuracy of existing quantum mechanical corrections to surface potential based MOSFET compact models. Results show that the existing quantum mechanical corrections are not very accurate, particularly for the derivative of the surface potential with respect to terminal voltages. The error in the derivative of the surface potential with respect to the gate voltage is found to increase with increasing substrate doping density. This makes the existing corrections more prone to error with device scaling.

I. Introduction

Compact MOSFET models ties up the process of fabrication and circuit design. The development of compact models are particularly challenging due to the incompatible requirement of accuracy, generality and computational efficiency. In recent compact model formulations, a considerable attention has been focused on developing the surface potential (φ_s) based models [1-4]. PSP [1], a surface potential based model developed by Phillips and Penn State, has been accepted by the Compact Modelling Council as a new industry standard.

On the other hand, the continuous scaling of MOS transistors into nanometer regime greatly increases the quantum mechanical effects (QME). The use of thinner gate oxide and higher substrate doping results in a very high normal field at the semiconductor-oxide interface. As a result the energy spectrum consists of a set of discrete energy levels, where the first allowed energy level does not coincide with the bottom (top) of the conduction (valance) band [5]. The semi classical surface potential models do not consider the QME. So, extra correction terms are added to the surface potential based compact models to improve accuracy [2, 6-8]. In this work, a comparison of existing QM correction models to the MOS surface potential is shown and their accuracy is investigated with reference to the surface potential calculated by self-consistent numerical solution of coupled Poisson's and Schrödinger's equations.

II. Semiclassical Surface Potential Model

Rigorous theory for MOSFET terminal characteristics has been developed in the 60's in terms of the well-known Pao-Sah equation [9], which requires numerical solution of the double integral for the terminal currents. With the potential/charge balance and Poisson's solution, the semiclassical φ_s is related to the terminal gate voltage (V_g) and channel voltage (V_{ch}) by

$$V_g - V_{fb} - \varphi_s = \gamma \left\{ \begin{aligned} &\varphi_s + \varphi_t \cdot \left[\exp\left(-\frac{\varphi_s}{\varphi_t} \right) - 1 \right] \\ &+ \varphi_t \cdot \exp\left(\frac{-V_{ch} - \varphi_B}{\varphi_t} \right) \\ &\cdot \left[\exp\left(\frac{\varphi_s}{\varphi_t} \right) - 1 \right] \end{aligned} \right\}^{1/2} \quad (1)$$

Where V_{fb} is the flat band voltage, φ_t is the thermal voltage defined by kT/q, φ_f is the intrinsic Fermi potential defined by $\varphi_t \ln(N_A/n_i)$, $\varphi_B = 2\varphi_f$ and γ is the body factor defined by $(2q\varepsilon_{si}N_A/C_{ox})^{1/2}$. As equation (1) is an implicit equation of φ_s, numerical method is required to solve it. Consequently to solve equation (1) explicitly, several techniques were proposed [8, 10, 11].

Fig. 1 shows the semi classical φ_s as a function of gate voltage calculated using the approximate forms proposed by Prégaldiny et al. [8], Rios et al. [10] and Chen and Gildenblat [11]. Actual solution of φ_s found by solving equation (1) exactly is also shown in Fig. 1. Results show that the approximate forms of [10] and [11] accurately reproduce the solution of equation (1). These techniques are also computationally efficient.

III. Self Consistent Quantum Mechanical Model

Stern [5] proposed a self-consistent solution of coupled Schrödinger's and Poisson's equations to analyze quantum mechanical effects in MOSFETs. Results obtained from the self-consistent solution can be made more accurate if we consider wave function penetration into the gate

dielectric by using open boundary condition at silicon-oxide interface.

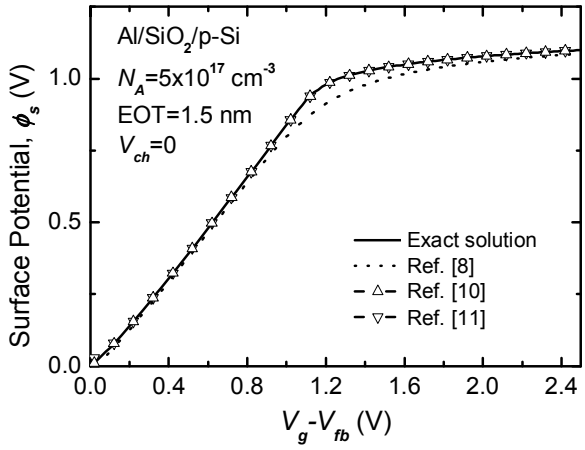


Fig. 1 Comparison of approximate solutions of equation (1) with the exact solution.

We have solved the coupled Schrödinger's and Poisson's equation numerically using logarithmic derivative technique [12] for the solution of Schrödinger's equation. Poisson's equation is solved by applying the finite difference technique with non-uniform mesh size. Wave function penetration is naturally included in the model by applying the boundary conditions that the electric field is zero deep inside the bulk silicon substrate and inside the gate electrode.

For closed boundary condition that neglects wave function penetration, wave functions go to zero at the silicon-oxide interface. Open boundary condition that considers wave function penetration causes a shift of the probability densities, hence the charge distribution, towards the silicon-oxide interface by a fraction of a nm. Such apparently small shift of the charge density has significant effects on the electrostatic properties of MOSFETs.

Fig. 2 shows the average distance (Z_{avg}) of inversion electrons from the silicon-oxide interface determined from self-consistent calculation using open boundary condition as well as closed boundary condition. Reduction in Z_{avg} is clearly visible when wave function penetration is taken into account.

Fig. 3 shows that φ_s, calculated by considering QME, is higher than the semiclassical solution obtained from equation (1). However, relative to the more accurate QM solution with open boundary condition, φ_s is over-estimated when wave function penetration is neglected.

IV. Existing Quantum Mechanical Corrections to the Surface Potential Based Models

Rios et al. [2] proposed the first quantum mechanical correction to the surface potential. Gildenblat et al. proposed a correction in [6]. PSP includes quantum mechanical correction in the form described in [7]. Prégaldiny et al. [8] suggested an explicit solution of surface potential considering quantum mechanical effects. By nature of compact models, these correction terms can be evaluated in computationally efficient manners. But none of these works incorporates the wave function penetration effect.

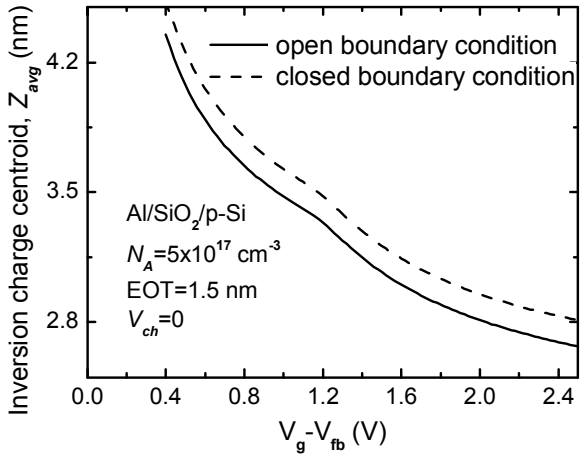


Fig. 2 Average distance (Z_{avg}) of the inversion electrons from the silicon-oxide interface as a function of gate bias voltage (V_g).

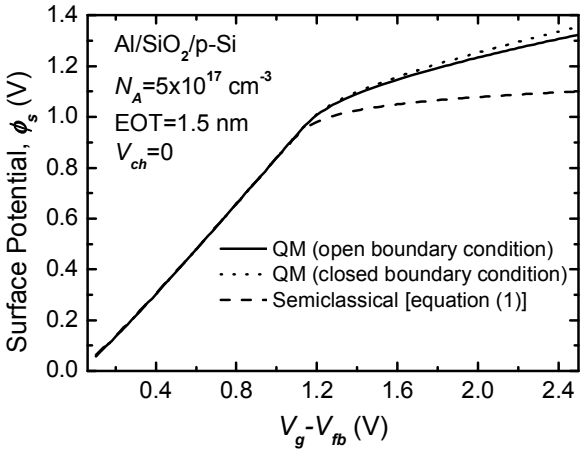


Fig. 3 Comparison of φ_s-V_g relationship calculated using different techniques.

Due to the requirement of computational efficiency, QM correction models are based on approximate analytical solutions. Triangular potential approximation has been used in [2, 6, 7] and the variational method has been used in [8]. However, it is known that these approximations are not accurate, particularly in strong inversion.

In this work, we have investigated the accuracy of existing quantum mechanical corrections to the compact surface potential based models by comparing with our self-consistent numerical solution with open boundary condition. In this verification, we focus particularly on the derivation of φ_s with respect to the terminal voltages. $d\varphi_s/dV_g$ is an important quantity required in many circuit simulation. But until now, there has been no systematic study on the accuracy of this quantity.

Figs. 4, 5 and 6 show the self-consistent quantum mechanical φ_s and its derivative as a function of gate

voltage (V_g) calculated with open boundary condition. The approximate solutions of [7] and [8], including QME are also shown and the semiclassical φ_s is shown as well.

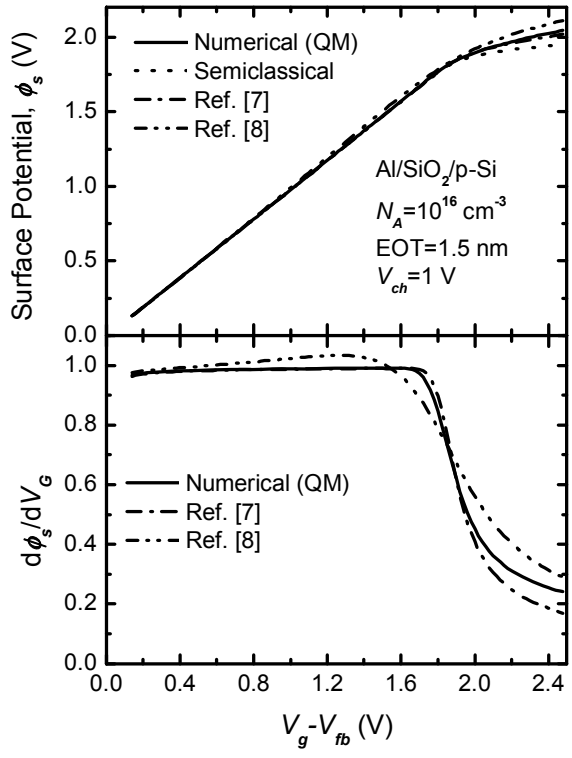


Fig. 4 Comparison of φ_s and $d\varphi_s/dV_g$ as a function of gate bias voltage (V_g) calculated using different models. Here N_A=10^{16} cm^{-3} and V_{ch}=1V.

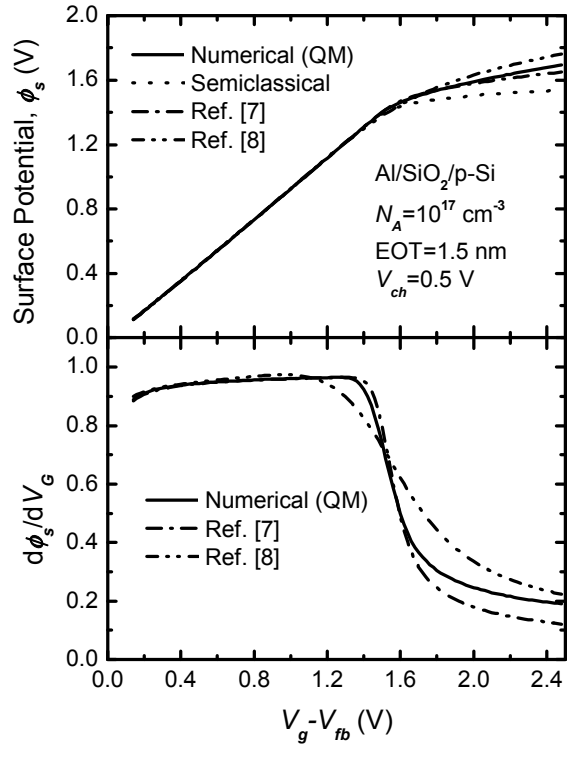


Fig. 5 Comparison of φ_s and $d\varphi_s/dV_g$ as a function of gate bias voltage (V_g) calculated using different models. Here N_A=10^{17} cm^{-3} and V_{ch}=0.5V.

It is observed that while the approximate analytical QM corrections to φ_s as proposed in [7] and [8] reproduces the φ_s-V_g characteristics reasonably well, the derivative of φ_s with respect to V_g is rather inaccurate. For all combinations of N_A and V_{ch} considered here, [7] underestimates $d\varphi_s/dV_g$ in strong inversion and [8] overestimates this quantity in strong inversion. [8] also provides inaccurate $d\varphi_s/dV_g$ in weak inversion region. The differences between the QM numerical results and the analytical corrections increase as the substrate doping density is increased. As the scaling of the MOSFETs require higher substrate doping density with reduced dimensions, it is anticipated that the existing QM corrections to the surface potential based compact models will become less accurate with device scaling. It is therefore necessary to develop a more accurate compact correction term for QME in φ_s. Such a work is underway and the results will be reported elsewhere.

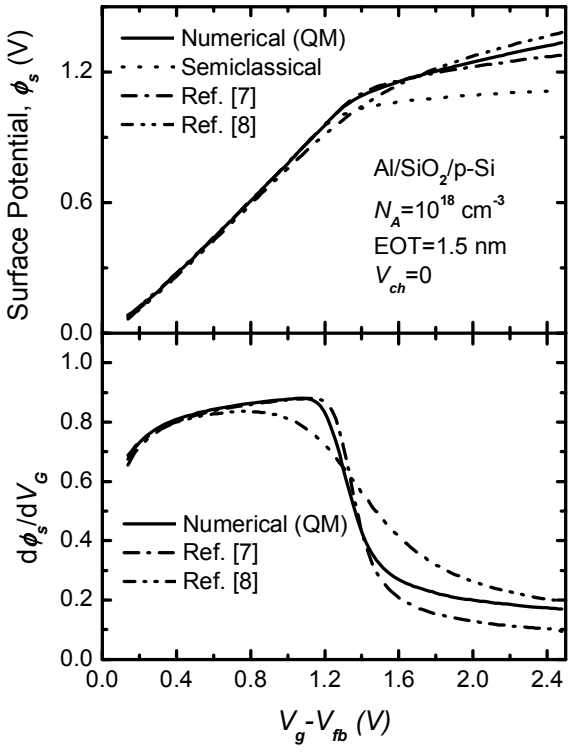


Fig. 6 Comparison of φ_s and $d\varphi_s/dV_g$ as a function of gate bias voltage (V_g) calculated using different models. Here N_A=10^{18} cm^{-3} and V_{ch}=0V.

V. Conclusions

Surface potential based compact models have become popular in recent years for simulation of sub 100 nm MOSFETs. Quantum mechanical effects are added as corrections to the semiclassical surface potential based models. We have compared two existing QM corrections to numerical results obtained from self-consistent solution of Schrödinger's and Poisson's equations including wave function penetration effect. Results show that the derivative of the surface potential with respect to the gate voltage is not accurate in strong inversion when calculated from the existing QM corrections. The error in the derivative increases with increasing substrate doping

density, making the existing corrections susceptible to increasing error with device scaling.

References

[1] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. Langevelde, G. Smit, A. Scholten, and D. Klaassen, "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation," *IEEE Trans. Electron Devices*, Vol. 53, No. 9, pp. 1979 - 1993, Sep. 2006.

[2] R. Rios, N.D. Arora, C. Huang, N. Khalil, J. Faricelli, L.Gruber, "A physical compact MOSFET model, including quantum mechanical effects, for statistical circuit design applications," *IEDM Tech. Digest.*, pp. 937-940, Dec. 1995.

[3] M. Miura-Mattausch, H. Ueno, M. Tanaka, H. J. Mattausch, S. Kumashiro, T. Yamaguchi, K. Yamashita, and N. Nakayama, "HiSIM: A MOSFET model for circuit simulation connecting circuit performance with technology," *IEDM Tech. Dig.*, pp. 109–112, Dec. 2002.

[4] A. R. Boothroyd, S. W. Tarasewicz, and C. Slaby, "MISNAN—A physically based continuous MOSFET model for CAD applications," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 10, no. 12, pp. 1512–1529, Dec. 1991.

[5] F. Stern, "Self-consistent results for n-type Si inversion layers." *Phys. Rev. B*, Vol. 5, No.12, pp. 4891-4899, Jun. 1972.

[6] G. Gildenblat, T.L. Chen, P. Bendix "Closed-form approximation for the perturbation of MOSFET surface potential by quantum-mechanical-effects," *Electron. Lett.*, Vol. 36, No.12, pp. 1072-1073, Jun. 2000.

[7] R. van Langevelde, A.J. Scholten and D.B.M. Klaassen, "Physical Background of MOS Model-11 Level 1101," *Nat.Lab.Unclassified Report 2003/00239*, April 2003.

[8] F. Prégaldiny, C. Lallement , R. van Langevelde ,D. Mathiot , "An advanced explicit surface potential model physically accounting for the quantization effects in deep-submicron MOSFETs." *Solid State Electron.*, Vol. 48, No. 3, pp. 427-435, Mar. 2004.

[9] H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal–oxide (insulator)–semiconductor transistors," *Solid State Electron.*, vol. 9, no. 10, pp. 927–937, Oct. 1966.

[10] R. Rios, S. Mudanai, S. Wei-Kai, P. Packan, "An efficient surface potential solution algorithm for compact MOSFET models," *IEDM Techn. Dig.*, pp. 555–557, Dec. 2004.

[11] T. L. Chen and G. Gildenblat, "Analytical approximation for the MOSFET surface potential," *Solid State Electron.*, vol. 45, no. 2, pp. 335–339, Feb. 2001.

[12] A. Haque, A. N. Khondker, "An efficient technique to calculate the normalized wave functions in arbitrary one-dimensional quantum well structures," *J. Appl. Phys.*, Vol. 84, No.10, pp. 5802-5804, Nov. 1998.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

EFFECT OF FRINGING FIELD IN MODELING OF SUBTHRESHOLD SURFACE POTENTIAL IN DUAL MATERIAL GATE(DMG) MOSFETS

[1,] Swapnadip De*, [2]Angsuman Sarkar and [3]N. Mohankumar, Member, IEEE, [4]Chandan Kumar Sarkar, Senior Member, IEEE

[1] Dept. of ECE, Meghnad Saha Institute of Technology, Nazirabad, Kolkata 700017, India,
E-mail: swapnadipde26@yahoo.co.in
[2]ECE Dept., Jalpaiguri Govt. Engg. College, Jalpaiguri-735102, India,
E-mail: angsumansarkar@hotmail.com
[3]Senior Research Fellow, ETCE Dept., Jadavpur University, Kolkata-700032, India,
E-mail: vinamokushi@yahoo.com
[4]ETCE Dept., Jadavpur University, Kolkata-700032, India.
E-mail: phyhod@yahoo.co.in

*Corresponding author.

Abstract **An analytical subthreshold surface potential model for Dual Material Gate MOSFET including the effect of inner fringing field is presented, considering surface potential variation with the depth of the channel depletion layer. A pseudo two dimensional method is adopted and a more accurate prediction of surface potential including the fringing field effect is reported.**

Index Terms—Pseudo 2-D analysis, Surface Potential, Inner Fringing Field, DMG MOSFET.

I. INTRODUCTION

An analytical model for the subthreshold surface potential in a DMG MOS transistor is developed by solving a pseudo-2D Poisson's equation,formulated by applying the Gauss's law to a rectangular box in the channel covering the whole depletion region.The non-uniformity of the channel depletion layer depth is taken care in determining the surface potential.It is functionally one-dimensional but provides two-dimensional accuracy[1].The horizontal component of the electric field is assumed to be uniform.So application of Gauss's law gives a 1-D second order differential equation which can be solved analytically.

Use of DMG devices provides an improved short channel effect such as the reduced DIBL,the increased hot carrier immunity, the gate capacitance reduction, the enhanced early voltage and the driving current etc., making these devices suitable for mixed—signal (analog and digital) applications[2,3].

Fringing fields arising in short channel devices have a significant effect on its performances.For accurate analysis of the devices,these effects should be included in determining the surface potential.A simple expression for the parasitic inner fringing capacitance from the bottom edge of the gate electrode is considered and the charges induced in the source and the drain regions due to this capacitance is included. An accurate model of subthreshold surface potential is developed for DMG MOSFET taking into account this effect.

II. MODEL DESCRIPTION

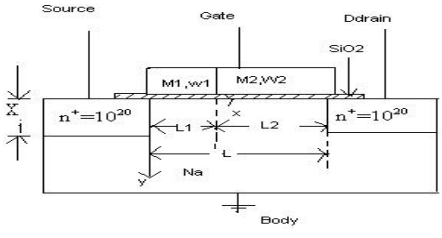


Fig. 1 The structure of the n-channel DMG-MOS transistor.

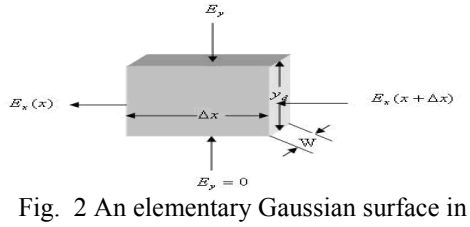


Fig. 2 An elementary Gaussian surface in the channel covering the entire depletion region.

The application of the Gauss's law to a rectangular box and neglecting the inversion layer charge,a pseudo-2D

Poisson's equation in the channel depletion region can be obtained as below:

$$\varepsilon_{si}\frac{d^2\psi_s}{dx^2}-\frac{C_{ox}}{Y_d}\psi_s=qN_a-\frac{C_{ox}}{Y_d}V_{GS} \quad \ldots\ldots\ldots\ldots (1)$$

where the symbols have their usual significances[1].

An accurate model must take into account the fringing effect for perfect estimation of surface potential.We give a brief account of the Fringing capacitances first and then introduce the fringing potential due to the inner fringing capacitances in our model.

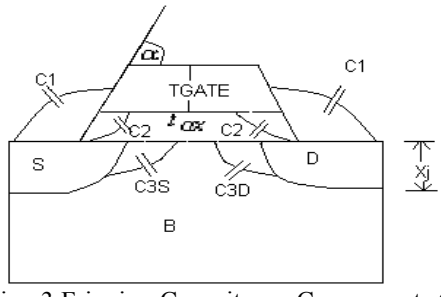


Fig. 3 Fringing Capacitance Components C1, C2, and C3.

The Capacitances are as follows: C1 is the outer-fringing-field capacitance between the gate and the source or the drain electrode. C2 is the direct overlap capacitance between the gate and the source or the drain junction. C3 is the inner-fringing-field capacitance between the gate and the side wall of the source or the drain junction.

The expressions of C1, C2, and C3 are given in details in [4].Here α is the slanting angle of the gate electrode in radians, T_{GATE} is the thickness of the gate electrode as shown in Fig. 3.Further, C_F is the maximum value of the inner-fringing capacitance component C3. The capacitance component C3 is bias-dependent and it is modeled as a charge based form. Hence,

$$Q_{D.F}=-C_F.\frac{V_{DS}-VDS}{1+\exp\left(-\frac{V_{GB}-V_{FB}}{30\Phi_t}\right)}=-C_F.V_{fd}$$

$$Q_{S.F}=-C_F.\frac{V_{GST}-V_{GS}+V_{FB}+2\phi_F+\gamma\sqrt{2\phi_F-V_{BS}}}{1+\exp\left(-\frac{V_{GB}-V_{FB}}{30\Phi_t}\right)}=-C_F.V_{fs}$$

Here V_{DS} = drain to source voltage, V_{GS} = gate to source voltage, V_{GB} = gate bias and V_{FB} = flat-band voltage. VDS is obtained from[4]. ϕ_F = Fermi potential of Silicon. V_{GST} = V_{GS} -V_{TH} , where V_{TH} is the threshold voltage.

Φ_t =Thermal voltage. V_{fs} = the inner fringing potential in the source end and V_{fd} = inner fringing potential in the drain end. $Q_{D.F}$ =Charge induced in the drain due to inner fringing capacitance and $Q_{S.F}$ = Charge induced in the source due to inner fringing capacitance.

In terms of the work function,if W_m and W_{si} are the work functions of the gate and the silicon substrate respectively.The flat-band voltage is given by

$$V_{FB}=(W_m - W_{si})/q.$$

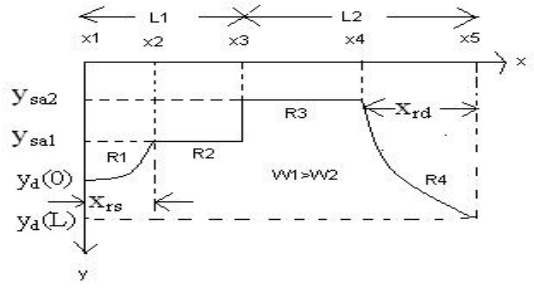


Fig. 4 Typical variation of the depletion layer depth $Y_d(x)$.

An empirical model for the depletion layer width is proposed for this purpose as $Y_d(x)=(ax+b)^2$ with the source and the drain end values given by

$$Y_d(0)=X_j+\sqrt{2\varepsilon_{si}(V_{SB}+V_{bi}+V_{fs})/(qNa)}=X_j+X_{rs}$$

and

$$Y_d(L)=X_j+\sqrt{2\varepsilon_{si}(V_{DB}+V_{bi}+V_{fd})/(qNa)}=X_j+X_{rd}$$

respectively where, V_{fs} = the inner fringing potential in the source end , V_{fd} = the inner fringing potential in the drain end , X_j = the junction depth, V_{bi} =the built-in potential of the substrate. V_{SB} and V_{DB} are the source and the drain bias respectively. Na =the acceptor ion concentration.

$$X_{rs}=\sqrt{2\varepsilon_{si}(V_{SB}+V_{bi}+V_{fs})/(qNa)} \text{ and}$$

$$X_{rd}=\sqrt{2\varepsilon_{si}(V_{DB}+V_{bi}+V_{fd})/(qNa)}$$

are the depth of penetrations of the depletion layers into the channel / substrate due to the built-in potential V_{bi} (between the n^+ -source/drain and the p-type channel/substrate) and the reverse bias V_{SB} and V_{DB} at the source and the drain ends . The channel may be divided into four regions R_1, R_2, R_3, R_4 with known values at the two ends as shown in Fig-4.

The best fit of the model surface potential profile with ISE TCAD is found for the bias dependent fitting parameter $\zeta_s = 2(V_{SB} + V_{bi} + V_{fs}) / V_{bi}$ for the source side and $\zeta_d = 2(V_{DB} + V_{bi} + V_{fd}) / V_{bi}$ for the drain side. In other words, while computing *a* and *b* we use $Y_d(o) / \zeta_s$ and $Y_d(L) / \zeta_d$, instead of, $Y_d(0)$ and $Y_d(L)$ respectively. Thus fitting parameters take into account other fringing effects and leakage capacitances at the source and the drain ends. The surface potential $\Psi_s(x)$ can be determined as in

[5],where the surface potential is calculated for DMG MOSFET without taking the inner fringing field into account. The surface potential is calculated for all the four regions of the channel .The fitting parameter is reduced considerably in this model by including the inner fringing field.

III. RESULTS

The developed structure shown in Fig 1 is used to verify the model against the 2-D numerical device simulator DESSIS.Two different metals M_1 and M_2 are used.The typical parameters for the oxide thickness,the junction depth and the channel length are t_{ox}=3.5 nm,X_j =40 nm and L=100 nm which are representative for a typical 130-nm device,along with V_{SB}=0V are used.Similarly equal values for L_1 and L_2 with typical work functionsW_1=4.25 eV and W_2=4.1 eV are used in this study.Solid lines are used for the model predictions,while the circular symbols are used for the corresponding predictions by DESSIS.

Fig 5 and 6 shows the comparison of the subthreshold surface potential profiles generated by our model and the DESSIS,against the variation of the substrate doping,the channel length and the drain voltage.

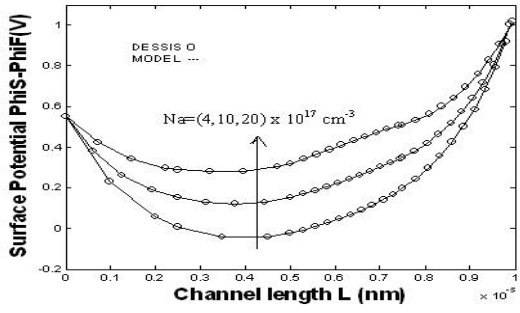


Fig. 5 The plots for three different values of the substrate doping N_a=4×10^{17} ,10^{18} and 2×10^{18} cm^{-3} against the applied voltages V_{GS}=0 V and V_{DS}=0.5 V.

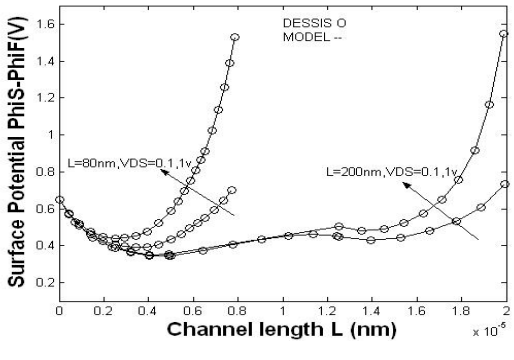


Fig. 6 The variation of the potential profiles in a device with two different channel lengths L=80 and 200 nm for N_a=6×10^{17} cm^{-3} against two different drain voltages V_{DS}=0.1 V and 1 V along with V_{SB}=0.1 V and V_{GB}=0.2 V.

All the plots show a convincing agreement of the model with DESSIS,for a wide variation of the device length,the substrate doping and the drain voltage,in addition to nonzero voltages on the source and the gate terminals.

The influence of varying the work function and the length of the two gate materials on the potential profiles are then studied.The corresponding potential profiles in a device with N_a=6×10^{17} cm^{-3} , V_{DS}=0.5 V and the grounded gate terminal are plotted.

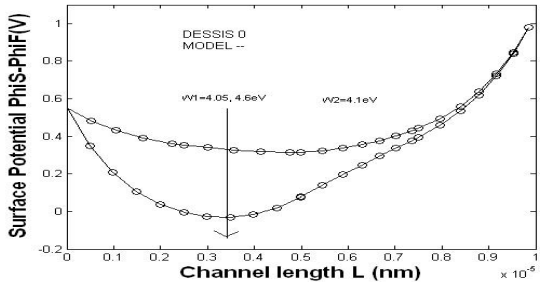


Fig. 7 The potential profiles for a fixed value of W_2=4.1 eV,with two different values of W_1=4.05 and 4.6 eV,respectively.

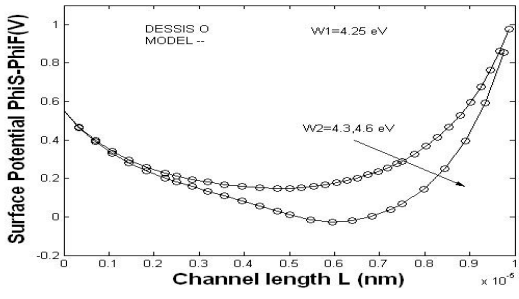


Fig. 8 The potential profiles for a fixed value of W_1=4.25 eV with two different values of W_2=4.3 and 4.6 eV,respectively.

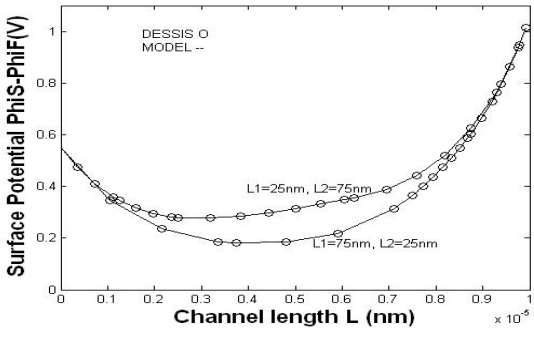


Fig. 9 The potential profiles for two sets of gate materials L_1=25 nm and L_2=75 nm andL_1=75 nm and L_2=25 nm.

When the length of the gate materials are changed for fixed work function values,we get the plots shown in Fig 9 for two sets of material length.In one of them,L_1=25

nm and L_2=75 nm,while just the reverse values i.e., L_1=75 nm and L_2=25 nm are used for the other set.

A good agreement of the model calculation with DESSIS proves that the model can be applied for any combinations of the gate material work functions and/or lengths.

IV. CONCLUSION

An improved analytical subthreshold surface potential model for DMG MOSFET is proposed in this paper that accounts for its dependence on varying depth of the channel depletion layer due to the source and the drain junctions and also the effect of inner fringing field is considered in the present model. Our model can predict the surface potential profile fairly accurately for a wide variation of the device parameters such as the substrate concentration,the channel length,the gate oxide thickness and also for the different biasing conditions. The dependence of the surface potential on the junction depth is accommodated as per the scaling guide lines of ITRS roadmap. This model is superior to the model in [5] since the fitting parameter is reduced significantly by including the inner fringing field in the reference model.

ACKNOWLEDGMENT

C.K.Sarkar wishes to thank AICTE for their financial support under RPS scheme. N.Mohankumar also wishes to thank CSIR for providing fellowship for his research activities.

REFERENCES

[1] S.Baishya,A.Mallik and C.K.Sarkar,"A subthreshold surface potential model for short-channel MOSFET taking into account the varying depth of channel depletion layer due to source and drain junctions", IEEE Trans. Electron Devices, vol. 53,pp. 507-514, Mar. 2006.

[2] S.Baishya,A.Mallik and C.K.Sarkar,"A subthreshold surface potential and drain current model for lateral asymmetric channel(LAC) MOSFETs", IETE Journal of Research , vol. 52,pp. 379-390, Sept-Oct. 2006.

[3] S.Baishya,A.Mallik and C.K.Sarkar," Subthreshold surface potential and drain current models for short-channel pocket implantedMOSFETs,"Microelectronics Engineering.Available online.

[4] Hong-June Park,Ping Keung Ko,Chenming Hu,"A Charge Sheet Capacitance Model of Short Channel MOSFET's for SPICE,"IEEE TRANSACTIONS ON COMPUTER AIDED DESIGN,vol-10,NO-3,Mar 1991.

[5] S.Baishya,A.Mallik and C.K.Sarkar,"A Pseudo Two-Dimensional Subthreshold surface potential model for Dual-Material Gate MOSFETs", IEEE Trans. Electron Devices, vol. 54,pp. 2520-2525, Sept. 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Comparative Analysis of Subthreshold Swing Models for Different Double Gate MOSFETs

Mehdi Zahid Sadi, Nittaranjan Karmakar, Mohammed Khorshed Alam and M. S.Islam

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
E-mail: islams@eee.buet.ac.bd

***Abstract*- we have analyzed the different structures of the DG MOSFETs and their potentials in suppressing short channel effects (SCEs). In particular, we have developed compact physics-based model of the subthreshold swing. Asymmetric DG MOSFET shows superior performance in the nanometer region. The new scale length for DG MOSFETs has been derived from the subthreshold swing model (S model). The new scale length is compared with that of the reported values. The developed physics-based model shows better results.**

I. Introduction

In accordance with Moore's law, the packing density of transistors per unit chip area is ever increasing in the VLSI microelectronic industries. This has been possible due to comprehensive scaling of MOSFET. The inherent benefits of MOSFET scaling are the speed improvements, increased packing density and energy reduction associated with binary logic transition. However the aggressive reduction in channel length leads to SCEs. The SCEs manifest themselves in deteriorating two important properties of the MOSFET: threshold voltage roll off and subthreshold swing roll up. In consequence of SCEs the ratio of drive (on) current to the leakage (off) current is substantially reduced imposing a severe tradeoff between circuit speed and stand by power. The methods and the consequences of suppressing SCEs in conventional bulk MOSFET have been extensively studied [1, 2, 3].

An alternative structure the DG MOSFET may push MOSFET scaling beyond the limits of conventional planar device structures as projected by International Technology Roadmap for semiconductors [4]. In DG MOSFETs, the silicon film is preferred to be undoped to avoid the effects of heavy channel doping [5,6]. Most of the research works on SCEs of DG MOSFETs are based on numerical simulations, compact physics-based models are highly desired in order to i) gain physical insight into operating principles of DG MOSFETs, ii) facilitate device design, iii) identify key technological challenges to their fabrication, and iv) comprehensively project their ultimate scaling capability. In this paper we have analyzed the different structures of the DG MOSFETs and their potentials in suppressing SCEs. In particular we have developed a compact physics-based model of the subthreshold swing characteristics of different DG MOSFET structures. Finally, the derived new scale length has been compared with that of the reported values.

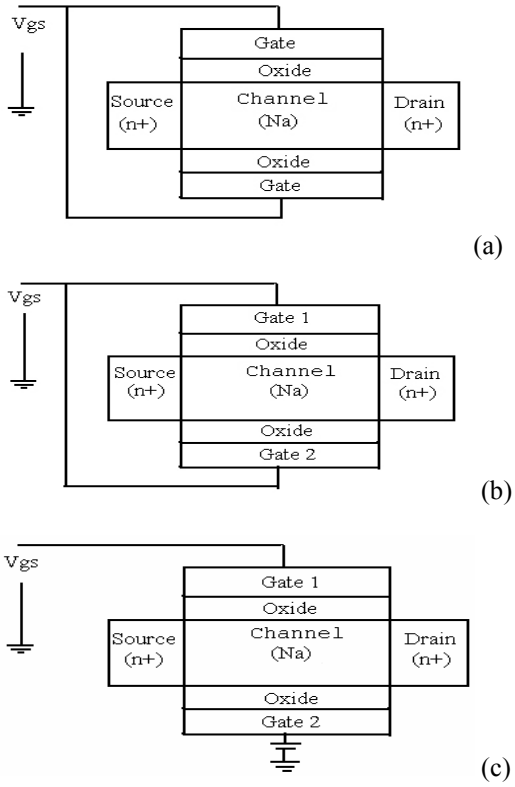


Fig. 1. Structure of DG MOSFETs: (a) Symmetric (two gates are identical), (b) Asymmetric (two gates are made of different materials), and (c) Ground-plane (one gate is at constant bias).

II. Different Structure of DG MOSFETs

Depending upon the way the gate voltage is applied, the DG MOSFET may be categorized as Symmetric Double Gate (SDG), Asymmetric Double Gate (ADG) and Ground Plane (GP) MOSFET as illustrated in Fig. 1. A symmetric DG MOSFET results when both the gates have the same

work-function and a single input voltage is applied to both gates. An asymmetric DG MOSFET can be formed in two ways i) synchronized but different input voltages to two identical gates. or, ii) the same input voltage is applied to two identical gates of different work-functions. The names 'symmetry' and 'asymmetry' essentially reflect the presence or absence of symmetry of the electric field inside the channel. A ground plane (GP) MOSFET has one of its gates biased to a constant voltage, while the input signal enters the other gate only. The configuration of the channel electric field in GP MOSFET is asymmetric in general.

III. Reported Subthreshold Swing Models

The reported subthreshold swing models (S models) of DG MOSFET are based on the symmetric structure (SDG). The model proposed by Agrawal *et al.* [7] is based on rigorous solution of two-dimensional (2-D) Poisson's equation in the channel region. In this model it was assumed that the subthreshold leakage currents in DG MOSFET flow at the silicon surface analogously to the bulk MOSFET. Therefore, the subthreshold swing was determined by how the surface electrostatic potential changes with the gate voltage. Tosaka *et al.* [8] derived an alternative model based on the observations of numerical simulations that showed center of the silicon film (or mid-depth) to be most leaky. Consequently the center potential and its response to the gate voltage were chosen to derive the subthreshold swing expression. For asymmetric structure Suzuki *et. al.* [9] proposed an S model for p+/n+ poly-si gates. As for GP structures not much research has been performed. Comparative analysis of S models for the three different structures has not been performed extensively.

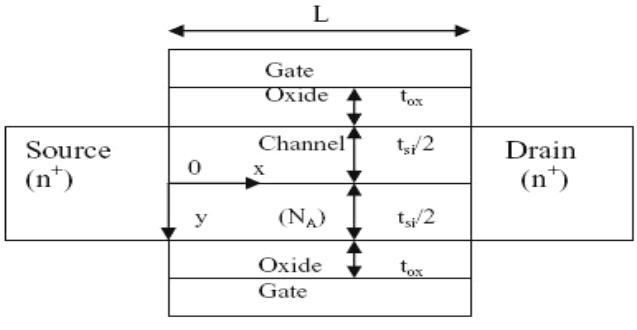


Fig. 2. Schematic of cross-section of a DG MOSFET and the coordinate system used for the solution of Poisson's equation.

IV. Development of Physics-based S Model

For analytical derivation of S model, the channel potential distribution in the DG MOSFET is first obtained by solution of 2-D Poisson's equation with ionized dopant term included only in the channel as the body is lightly doped. Based on the solution an analytical subthreshold swing model is developed. The coordinate system for the solution of 2-D poison equation is given in Fig. 2. The origin of the coordinate system is placed at the center of the source side boundary of the silicon channel. The silicon channel is fully depleted so the dopant atoms in the channel are fully ionized. In the subthreshold region, when the gate voltage is low, so the mobile charges can be ignored in comparison with the ionized dopant charges. The channel electrostatics thus is approximately described by the following 2-D Poisson's equation,

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = \frac{qN_A}{\varepsilon_{si}} . \quad (1)$$

Where, φ is channel potential referenced to the Fermi level in equilibrium MOSFET or equivalently, the Fermi level in the Source [6]. N_A is the doping concentration in lightly doped channel ε_{si} is the permittivity of the silicon channel.

The Boundary conditions for solving equation (1) are (referred to the structure of Fig.2):

$$\varphi(0, y) = V_{bi,i} \quad (2)$$

$$\varphi(L, y) = V_{bi,i} + V_{DS} \quad (3)$$

$$\varepsilon_I \frac{V_{GS,F} - \phi_{FM,i} - \varphi(x, -t_{si}/2)}{t_I} = -\varepsilon_{SI} \frac{\partial \varphi(x,y)}{\partial y}\Big|_{y=\frac{-t_{SI}}{2}} \quad (4)$$

$$\varepsilon_I \frac{V_{GS,B} - \phi_{BM,i} - \varphi(x, t_{SI}/2)}{t_I} = \varepsilon_{SI} \frac{\partial \varphi(x,y)}{\partial y}\Big|_{y=\frac{t_{SI}}{2}} \quad (5)$$

The parameter $V_{bi,i}$ is the build-in voltage of the junction between source and intrinsic silicon

$$V_{bi,i} = \frac{KT}{q}\left(\frac{N_{D/S}}{n_i}\right) \quad (6)$$

$N_{D/S}$ is doping concentration of source and drain. $\phi_{FM,i}$ and $\phi_{BM,i}$ are work function differences between front/back gates and intrinsic silicon, respectively. Thus

$$\phi_{FM,i} = \phi_{FM} - \phi_i \quad (7)$$

$$\phi_{BM,i} = \phi_{BM} - \phi_i \quad (8)$$

For simplification purpose the effective front and back gate voltages can be introduced as,

$$V_{F,eff} = V_{GS,F} - \phi_{FM,i} \quad (9)$$

$$V_{B,eff} = V_{GS,B} - \phi_{BM,i} \quad (10)$$

The 2-D Poisson equation (with boundary conditions (2)-(5)) is solved through Laplace reduction followed by separation of variables as performed in [7]. The channel potential distribution $\varphi(x,y)$ is obtained as,

$$\varphi(x,y) = V_{1D}(y) + \varphi_{2D}(x,y) \quad (11)$$

where the 1-D function $V_{1D}(y)$ describes the channel potential in a long channel device that is completely defined by ionized dopant charges, and the 2-D function $\varphi_{2D}(x,y)$ characterizes the impact or disturbance, caused by the source and drain. Function $V_{1D}(y)$ is given as,

$$V_{1D}(y) = \frac{V_A}{2}\left(\frac{y}{t_{SI}}\right)^2 + \frac{r}{r+2}\left(V_{B,eff} - V_{F,eff}\right)\left(\frac{y}{t_{SI}}\right) + \frac{(V_{B,eff} + V_{F,eff})}{2} - V_A\left(\frac{1}{2r} + \frac{1}{8}\right) \quad (12)$$

where, r is an insulator-to-silicon capacitance ratio,

$$r = \frac{\varepsilon_I t_{SI}}{\varepsilon_{SI} t_I} \tag{13}$$

V_A is a shorthand notation for,

$$V_A = \frac{qN_A t^2_{SI}}{\varepsilon_{SI}} \tag{14}$$

Function $\varphi_{2D}(x,y)$ is given as,

$$\varphi_{2D}(x,y) \approx \Gamma \left(V_1 \frac{\cosh\frac{x - L/2}{\lambda}}{\cosh\frac{L}{2\lambda}} + V_{DS} \frac{\sinh\frac{x}{\lambda}}{\sinh\frac{L}{\lambda}} \right) \cos\frac{y}{\lambda} \tag{15}$$

Where V_1 is shorthand notation for,

$$V_1 = V_{bi,i} - \frac{V_{B,eff} + V_{F,eff}}{2} + \frac{\lambda^2}{t_{SI}^2} V_A \tag{16}$$

and λ is the lowest order eigenvalue determined as the lowest order solution to the following transcendental equation,

$$\tan\left(\frac{t_{SI}}{2\lambda}\right) = \frac{r\lambda}{t_{SI}} \tag{17}$$

Parameter Γ is expressed as,

$$\Gamma = \frac{2\frac{\lambda}{t_{SI}}\sqrt{1 + \frac{t_{SI}^2}{r^2\lambda^2}}}{\frac{1}{r} + \frac{1}{2} + \frac{t_{SI}^2}{2r^2\lambda^2}} \tag{18}$$

Among the potential distribution all over the channel region, the minimum potential point in the channel length direction, known as the "virtual cathode," is of the most interest for the device modeling purpose. At this point an energy barrier is formed, over which free electrons diffuse from the source and then are swept into the drain forming the subthreshold drain current. By setting,

$$\frac{\partial\varphi(x,y)}{\partial x}\Big|_{x=x_{min}} = 0 \tag{19}$$

where, X_{min} is the location of the virtual cathode, the electrostatic potential at the virtual cathode can be found as,

$$\varphi_{min}(y) \approx V_{1D}(y) + 2\Gamma\cos\frac{y}{\lambda}\sqrt{V_1(V_1 + V_{DS})}\exp(\frac{-L}{2\lambda}) \tag{20}$$

Assuming that the subthreshold drain current, I_D, is proportional to the total amount of free electrons diffusing over the virtual cathode, that is,

$$I_D \alpha \int_{y=\frac{-tsi}{2}}^{y=\frac{tsi}{2}} W n_i \exp(\frac{KT}{q}(\varphi_{min} - \phi_F))dy \tag{21}$$

where W is the device width and ϕ_F is the difference between the equilibrium Fermi level and non-equilibrium quasi-Fermi level caused by the current flow in the channel length direction.

The subthreshold swing, S, can be expressed as,

$$S = \frac{\partial V_{GS}}{\partial \log I_D} = \left[\frac{\int_{y=\frac{-t_{SI}}{2}}^{y=\frac{t_{SI}}{2}} n_m(y)\frac{\partial\varphi_{min}(y)}{\partial V_{GS}}dy}{\int_{y=\frac{-t_{SI}}{2}}^{\frac{t_{SI}}{2}} n_m(y)dy} \right]^{-1} \frac{KT}{q}\ln 10 \tag{22}$$

where, $n_m(y)$ is introduced to denote,

$$n_m(y) = \exp(\frac{q}{KT}\varphi_{min}(y)) \tag{23}$$

S models for different DG structure can be obtained by exploiting equation (22) with proper parameters.

V. S Model for SDG MOSFET

In a symmetric DG MOSFET as two gates have the same work-function and the input gate voltages are identical, $\Phi_{MS,\ I}$ and V_{GS} are introduced to denote the gate work-function and gate voltage, respectively. The electrostatic potential at the virtual cathode can then be obtained by modifying (20) as,

$$\varphi_{min}(y) = V_{GS} - \phi_{MS,i} + \frac{V_A}{2}\left[\left(\frac{y}{t_{SI}}\right)^2 - \frac{1}{r} - \frac{1}{4}\right] + 2\Gamma\cos\frac{y}{\lambda}\sqrt{V_1(V_1 + V_{DS})}\exp(\frac{-L}{2\lambda}) \tag{24}$$

where, $$V_1 = V_{bi,i} - V_{GS} + \phi_{MS,i} + \frac{\lambda^2}{t_{SI}^2}V_A \tag{25}$$

Taking the derivative of $\varphi_{min}(y)$ with respect to V_{GS} and putting the result into (22) yields,

$$S = \left[1 - 2\Gamma\frac{(V_1 + \frac{V_{DS}}{2})}{\sqrt{V_1(V_1 + V_{DS})}}\cos\frac{d_{eff}}{\lambda}\exp(\frac{-L}{2\lambda})\right]^{-1}\frac{KT}{q}\ln 10 \tag{26}$$

where the parameter d_{eff} is defined such that,

$$\cos\frac{d_{eff}}{\lambda} = \frac{\int_{y=\frac{-t_{SI}}{2}}^{\frac{t_{SI}}{2}} \cos\frac{y}{\lambda} n_m(y)dy}{\int_{y=\frac{-t_{SI}}{2}}^{\frac{t_{SI}}{2}} n_m(y)dy} \tag{27}$$

Fig. 3 is obtained using MATLAB to numerically solve the equation (27) and plotting normalized d_{eff} against doping concentration.

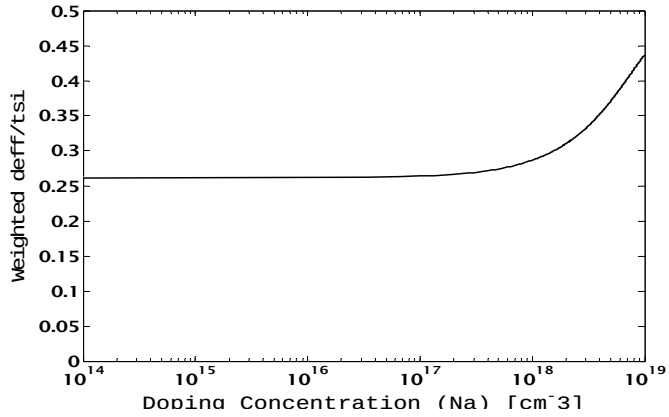


Fig. 3: Normalized effective conduction path against channel doping concentration.

From Fig.3 it is observed that for moderate doping levels the effective conduction path is almost located around $t_{SI}/4$. So using $d_{eff}= t_{SI}/4$ in equation (26) an approximate model can be obtained which will be very similar to the general model. Figure 4 shows the subthreshold swing as a function of channel length for the general model and approximate model (undoped and center models).

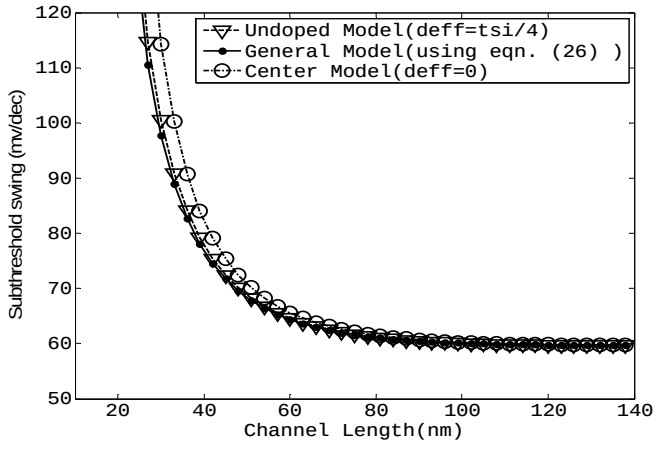


Fig. 4 Plots of general and approximation S models.

VI. S Model for ADG MOSFET

In asymmetric DG MOSFET, generally the bottom and front gate have different work-functions. So the effective gate voltages will be different as follows,

$$V_{F,eff} = V_{GS,F} - \phi_{FM,i} \quad (28)$$

$$V_{B,eff} = V_{GS,B} - \phi_{BM,i} \quad (29)$$

Generally front and back gates are n+ and p+ polysilicon, respectively. From [10] $\phi_{FM,i}$ = -1.1V and $\phi_{BM,i}$ = 0.3 V. Using the above work function values and the corresponding data from Table:1 S model for ADG MOSFET can be plotted (Fig. 5). Due to asymmetric potential profile in the channel depth direction electrons are asymmetrically distributed in the channel depth. The gate with higher effective voltage or smaller work function experiences stronger surface inversion than the other gate. Consequently, the effective conduction path should be located close to this surface. So as an approximation we can assume d_{eff} to be equal to $t_{si}/2$ or $-t_{si}/2$ depending on the value of the effective gate voltage. The equation for S is similar to that of SDG with V_1 obtained from equation (16).

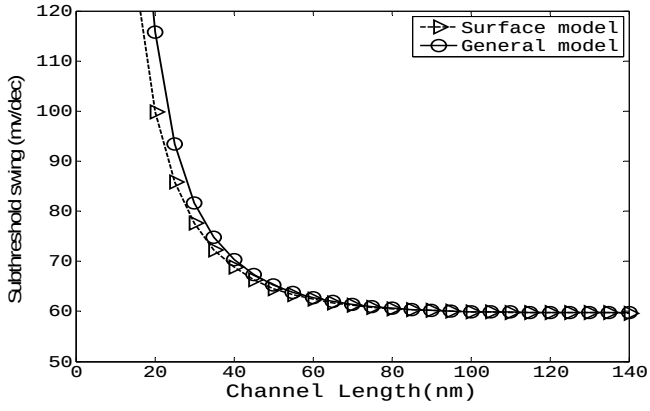


Fig. 5. Subthreshold Swing Models of ADG MOSFET using data from Table 1.

Table 1: Data used in plotting

N_A	t_{ox}	t_{si}	V_{DS}	V_{GS}
10^{16} cm^{-3}	1.5nm	20nm	0.1V	0.1V

VII. S Model for GP DG MOSFET

In GP MOSFET, generally the bottom gate is biased to a constant voltage whereas the top gate experiences voltage sweep. To obtain the S model minimum potential $\varphi_{min}(y)$ from equation (20) is differentiated with respect to sweeping front gate voltage and the result substituted in the general S model equation (22) resulting the following S model,

$$S = \frac{KT}{q}\ln 10\left[\frac{1}{2} - \frac{r}{r+2}\frac{d_{eff.linear}}{t_{SI}} - \Gamma\exp\left(\frac{-L}{2\lambda}\right)\frac{V_1 + \frac{V_{DS}}{2}}{\sqrt{V_1(V_1+V_{DS})}}\cos\frac{d_{eff}}{\lambda}\right]^{-1} \quad (30)$$

where $d_{eff.linear}$ is simplified notation for,

$$d_{eff.liner} = \frac{\int_{y=\frac{-t_{SI}}{2}}^{\frac{t_{SI}}{2}} y n_m(y)dy}{\int_{y=\frac{-t_{SI}}{2}}^{\frac{t_{SI}}{2}} n_m(y)dy} \quad (31)$$

Both the d_{eff} and $d_{eff.linear}$ parameters average the free electron concentration along the silicon channel thickness and are similar to one another. Similar to the ADG structure

the S model can be approximated as having a effective conducting path along the surface that is $d_{eff}=t_{SI}/2$ or $d_{eff}=-t_{SI}/2$. Figure 6 shows the subthreshold swing as a function of channel length for the GP DG MOSFET.

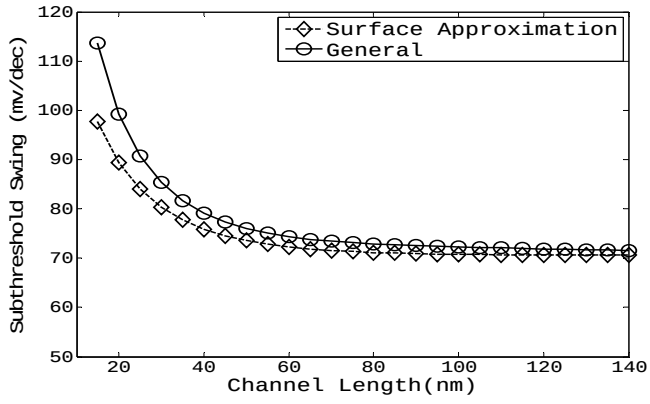


Fig. 6 Plots of S models for GP structure.

The general S models of the three DG MOSFET structures are plotted in Fig. 7.

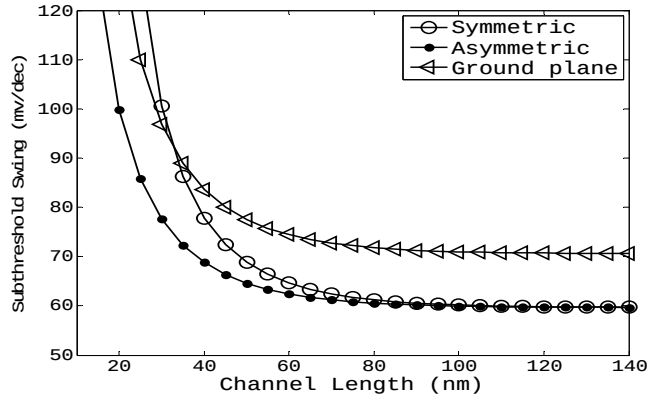


Fig. 7. Comparison of subthreshold swing of different DG MOSFET structures.

From Fig. 7 it is apparent that both SDG and ADG MOSFET possess an ideal subthreshold swing of 60mv/dec at large channel lengths. This can be explained by the electrostatic coupling between gate and channel enabled by fully depleted silicon film with negligible amount of free electrons and the synchronizations of input voltage at both gates. In case of GP structure as only one gate controls the electrostatics of the channel a much weaker gate control and a larger subthreshold swing results in long channel devices. As the channel length decreases the subthreshold swing difference between SDG and GP MOSFET reduces. This is due to the fact that at smaller channel lengths the effective conduction path of the GP structure shifts closer to the controlling gate than that of SDG structure resulting in a less susceptibility to source-drain impact. The ADG MOSFET reveals the best performance in terms of subthreshold swing considerations (Fig.7). This can be explained by the superior electrostatic control of the two gates with different work functions.

VIII. Scale Lengths

Scaling limits of DG MOSFETs have been investigated using various criteria including 1) subthreshold swing, 2) threshold voltage roll-off [7], 3) process tolerance [11, 15], 4) system-level requirements [13], and 5) source-to-drain tunneling limits [12]. A scale length is a parameter that is derived as a function of the vertical dimensions of a DG MOSFET, namely, the gate oxide thickness and the silicon film thickness, in such a way that the channel length of the device must be a few times as large as the scale length in order to suppress SCEs.

Suzuki *et al.* [15] proposed an alternative scale length using parabolic approximation of potential profile and the assumption of center conduction showing an improved correlation to the subthreshold swing in SDG MOSFETs,

$$\lambda_{Sukuki}=\sqrt{\frac{\varepsilon_{SI}t_{SI}t_I}{2\varepsilon_I}\left(1+\frac{\varepsilon_I t_{SI}}{4\varepsilon_{SI}t_I}\right)}. \qquad (32)$$

Challenging the accuracy of the parabolic approximation on potential profile, Monroe, *et al.* [16], Frank, *et al.* [17] proposed other scale lengths using the evanescent-mode analysis. Monroe *et al.* derived a scale length for fully-depleted SOI MOSFETs [16] that can be adapted for DG MOSFETs as

$$\frac{\left(\frac{\varepsilon_{SI}}{\varepsilon_I}-1\right)\tan\frac{t_I}{\lambda_{Monroe}}}{\frac{\varepsilon_{SI}}{\varepsilon_I}\tan^2\frac{t_I}{\lambda_{Morone}}+1}=\frac{\frac{\varepsilon_{SI}}{\varepsilon_I}\tan\frac{t_I}{\lambda_{Morone}}+\tan\frac{t_I+t_{SI}}{\lambda_{Monroe}}}{\frac{\varepsilon_{SI}}{\varepsilon_I}\tan\frac{t_I}{\lambda_{Monroe}}\tan\frac{t_I+t_{SI}}{\lambda_{Monroe}}-1} \qquad (33)$$

The scale length derived by Frank. *et al.* [17] is given as

$$1=\frac{\varepsilon_{SI}}{\varepsilon_I}\tan\left(\frac{t_I}{\lambda_{Frank}}\right)\tan\left(\frac{t_{SI}}{2\lambda_{Frank}}\right) \qquad (34)$$

Approximation and simplification of (34) and (35) were possible only for $t_I << t_{Si}$ [16],

$$\lambda_{Monroe/Frank}\approx\frac{t_{SI}+2\frac{\varepsilon_{SI}}{\varepsilon_I}t_I}{\pi} \qquad (35)$$

The ratio of the minimum channel length to a scale length was suggested to be between 5 and 10 [16], more than 6 [15], or between 4.7 and 6.3 [17].

IX. New Scale Length

From the inspection of S models we find the exponential function of $(-L/2\lambda)$ plays an important role. For new scale length derivation, equation (17) can be adopted to formulate for the lowest-order eigenvalue λ as the lowest-order solution. Since λ is a solution to a transcendental equation, its general solution can not be found explicitly. Approximate expressions are derived using trigonometric

identities and valid assumptions. Final solution of equation (17) is,

For $r \le \pi/2$,

$$\lambda = \frac{1+\frac{1}{r}}{1+\frac{\pi}{2}} t_{SI} = \frac{t_{SI} + \frac{\varepsilon_{SI} t_{OX}}{\varepsilon_{ox}}}{1+\frac{\pi}{2}} \qquad (36)$$

and for $r>\pi/2$,

$$\lambda = \frac{1+\frac{\sqrt{2}}{r}}{\sqrt{2}+\frac{\pi}{2}} t_{SI} = \frac{t_{SI} + \frac{\sqrt{2}\varepsilon_{SI} t_{OX}}{\varepsilon_{ox}}}{\sqrt{2}+\frac{\pi}{2}} \qquad (37)$$

The relative errors of (36) and (37) are less than 3% for *r*-values from 0.8 to 20 (corresponding to t_{Si} from 3.6 nm to 91 nm for t_{ox}=1.5 nm), which seems acceptable for device design and scaling study purposes. Expressions (36) and (37) show the needs of reducing vertical dimensions for SCE suppression. When t_{Si} is much larger than t_{ox}, λ is primarily determined by t_{Si}. In this regime, reducing t_{Si} is more effective than reducing t_{ox} for the same percentage of reduction. As t_{Si} becomes comparable to t_{ox}, reduction of t_{ox} becomes equally critical. The derived new scale length is compared with the reported results and is shown in Fig. 8. The derived new scale length shows better performance (Fig. 8).

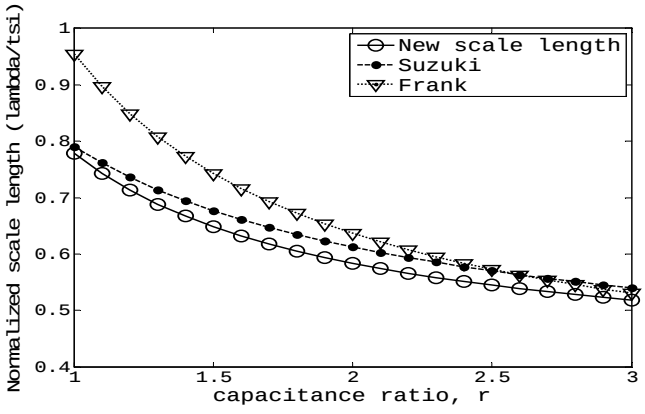


Fig. 8. Comparison of scale lengths for DG MOSFETs.

X. Conclusions

Compact physics-based subthreshold swing models for different DG MOSFETs have been developed in this paper. Asymmetric DG MOSFET shows superior performance in the nanometer region. At long channel length, device subthreshold swing of ADG and SDG may not differ much. But at channel lengths less than 60 nm there is significant difference, with ADG offering better performance. The derived new scale length is better when compared with that of the reported values. Approximate channel length can be found for specific application from equation (36) and (37) for calculated r values.

References

[1] X. Tang, V. K. De, and James D. Meindl, "Intrinsic MOSFET parameter fluctuations due to random dopant placement," *IEEE T-VLSI Systems*, vol. 5, pp. 369-376, Dec.1997.

[2] Y. Taur, C. H. Wann, and D. J. Frank, "25nm CMOS design considerations," *IEEE IEDM Tech. Dig.*, 1998, pp.789-792.

[3] H.-S. P. Wong, Y. Taur, and D. Frank, "Discrete random dopant distribution effects in nanometer-scal MOSFET's," *Microelectronic Reliability*, vol. 38, no. 9, pp.1447-1456, 1998.

[4] The International Technology Roadmap for Semiconductors, *http://public.itrs.net*, 2001,

[5] X. Tang, V. K. De, L. Wang, and J. D. Meindl, "SOI MOSFET fluctuation limits on gigascale integration (GSI)," *IEEE Int. SOI Conf.*, 1999, pp. 42-43.

[6] Y. Taur, "An analytical solution to a double-gate MOSFET with undoped body," *IEEE EDL,* vol. 21, no. 5, pp. 245-247, May 2000.

[7] B. Agrawal, "Comparative Scaling Opportunities of MOSFET Structures for Gigascale Integration (GSI)," *Doctoral Thesis, Rensselaer Polytechnic Institute, 1994.*

[8] Y. Tosaka, K. Suzuki, and T. Sugii, "Scaling-parameter-dependent model for Subthreshold swing S in double-gate SOI MOSFET's," *IEEE EDL,* vol. 15, no. 11, pp. 466-468, Nov. 1994.

[9] K. Suzuki and T. Sugii, "Analytical models for n+-p+ double-gate SOI MOSFET's,"*IEEE T-ED,* vol. 42, no. 11, pp. 1940-1948, Nov. 1995.

[10] Werner,W.M. "The Work function Differnce of MOS system with Aluminum Field Plates and Polycrystalline Silicon Field Plates", *Solid State Electronics*(1974)pp. 769-75

[11] D. J. Frank, S. E. Laux, and M. V. Fischetti, "Monte Carlo simulation of a 30 nm dual-gate MOSFET: how short can Si go?" *IEEE IEDM Tech. Dig.*, 1992, pp. 553-556.

[12] Z. Ren, R. Venugopal, S. Datta, M. Lundstrom, D.Jovanovic, and J. Fossum, "The ballistic nanotransistor: A simulation study," *IEEE IEDM Tech. Dig.*, 2000, pp. 715-718

[13] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 221-420, Mar. 2001.

[14] R.-H. Yan, A. Ourmard, and K. F. Lee, "Scaling the Si MOSFET: from bulk to SOI to bulk," *IEEE T-ED*, vol. 39, no. 7, pp. 1704-1710, July 1992.

[15] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie, and Y. Arimoto, "Scaling theory for double-gate SOI MOSFET's," *IEEE T-ED,* vol. 40, no. 12, pp. 2326-2329, Dec. 1993.

[16] D. Monroe and J. M. Hergenrother, "Evanescent-mode analysis of short-channel effects in fully depleted SOI and related MOSFETs," *Proc. IEEE Int. SOI Conf.*, 1998, pp. 157-158.

[17]. D. J. Frank, Y. Taur, and H.-S. P. Wong, "Generalized scale length for twodimensional effects in MOSFET's," IEEE EDL, vol. 19, no. 10, pp. 385-387, Oct. 1998.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Direct Extraction of Interface Trap States from the Low Frequency Gate C-V Characteristics of MOS Devices with Ultrathin High-K Gate Dielectrics

Md. M. Satter [1] *and A. Haque* [2]

[1] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh.
Email: mmsatter@eee.buet.ac.bd
[2] Department of Electrical and Electronic Engineering, East West University, Dhaka-1212, Bangladesh.
Email: ahaque@ewubd.edu

Abstract - A simple but accurate D_{it} extraction technique has been proposed from low frequency C-V characteristics of MOS devices with ultrathin high-K gate dielectrics. The proposed method incorporates quantum mechanical effect with wave function penetration for theoretical calculation of MOS electrostatics. Fermi-Dirac distribution function and the effect of finite temperature have also been included in the proposed technique. The extraction technique has been applied to different simulated devices with different D_{it} profiles. Excellent agreement has been found between extracted and actual D_{it} profiles.

I. Introduction

The generation of interface traps at the semiconductor/oxide interface of metal-oxide-semiconductor (MOS) structures has long been demonstrated as one of the prime factors behind the degradation of MOS device characteristics. It has already been established that low frequency C-V characteristic curves depend on the distribution of D_{it} states within Si bandgap even for the same average value of D_{it} [1]. Proper device modeling requires the knowledge of the density of interface trap states (D_{it}) throughout the bandgap, rather than the average density of states at midgap as often reported. Characterization of the interface traps has been an important job for accurate estimation of device life-time and reliability.

Several techniques have been proposed for the determination of the density of these traps and their energy distribution in the forbidden energy gap of the semiconductor materials [2-6]. Berglund [5] described a method for determining the energy distribution of interface traps by using low-frequency capacitance measurements on MOS structures, while Koukab *et. al.* [7] demonstrated a high-frequency C-V method for the analysis of the interface traps. However, various approximations and assumptions are usually made for the study of interface properties using the C-V measurements, which are not always well justified. For instance, both the low- and high-frequency capacitance methods use a theoretical curve of semiconductor capacitance as a function of surface potential, which comes from classical treatment.

Quantum mechanical (QM) effects are important in deeply scaled MOS devices with high-K gate dielectrics. Standard C-V method can be applied only after proper modification to incorporate appropriate QM effects for reliable extraction of D_{it} in ultrathin high- K gate dielectrics. Pacelli *et. al.* presented an improved version of the low frequency C-V technique [8] considering QM effects without wave function penetration which is prone to error in the context of today's aggressively scaled CMOS devices. Besides, their proposed method doesn't extract D_{it} profile in the entire bandgap range. In this work, we present a simple but robust method for direct extraction of D_{it} in the entire bandgap range from experimental low frequency C-V by comparing it with the simulated one without traps accounting QM effects with wave function penetration.

II. Extraction Technique

In the proposed method, the ideal gate C-V is simulated without interface traps using a physically based, accurate, QM model [1, 9]. Gate capacitance, C_g of MOSFETs with ultrathin high-K gate dielectrics is influenced by QM effects under large gate bias voltage for both accumulation and inversion conditions. Wave function penetration into the gate dielectric causes a shift in the semiconductor charge centroid resulting in an increase in calculated C_g. This effect is more severe in devices with high-K gate dielectrics due to lower values of φ_b. Modeling and characterization of these devices are nontrivial because of the variation of the potential barrier height, φ_b at the dielectric–Si interface and the dielectric constant, ε_{di} from one dielectric to another. It has been found that for the same equivalent dielectric thickness, C_g varies significantly with gate dielectric materials due to

variations in φ_b and ε_{di}. Only a physically based, accurate model can incorporate these QM effects through self-consistent solution of Schrödinger and Poisson equations. Open boundary conditions are used for the solution of Schrödinger's equation to incorporate wave function penetration effect. Parallel capacitance, C_p versus surface potential, φ_s curve is extracted from experimental C-Vs. For depletion and inversion, C_p is the parallel combination of depletion capacitance, C_{depl}, inversion capacitance, C_{inv} and interface trap capacitance, C_{it}. Whereas $C_p = C_{acc} + C_{it}$ for accumulation bias. Since C_g is the series combination of oxide capacitance, C_{ox} and C_p, $C_p = (1/C_g - 1/C_{ox})^{-1}$. Similarly φ_s can be extracted from gate C-V [5] according to the following equation.

$$\phi_S = \int_{V_{FB}}^{V_g} \left(1 - \frac{C(V_g)}{C_{ox}}\right) dV_g \,. \qquad (1)$$

Here V_{FB} is the flat-band voltage. C_{it} is the difference between C_p and the semiconductor capacitance, C_s (= C_{acc} or $C_{depl} + C_{inv}$). Proper C_{it} extraction is dependent upon the accuracy of the theoretically calculated C_s versus φ_S curve. QM modeling neglecting wave function penetration with closed boundary conditions inherently assumes that $\varphi_b \rightarrow \infty$, such models cannot incorporate any effect of variation of φ_b. Fig. 1 shows theoretically calculated C_s versus φ_S curves using open and close boundary conditions respectively. It is observed that neglecting wave function penetration under-estimates C_s in accumulation as well as inversion. If C_s, calculated using closed BC, be used to extract C_{it} from C_p, extracted result would be higher than the actual C_{it}. It is therefore essential to consider wave function penetration using open BC for accurately simulating C_s - φ_S characteristics of MOS structures with different high-K dielectric materials.

By definition, $C_{it} = \dfrac{dQ_{it}}{d\varphi_S}$. So from C_{it}, we can get Q_{it} using the following equation

$$Q_{it} = \int_0^{\varphi} C_{it}\, d\varphi_S \; + \mathrm{I} \qquad (2)$$

where I is a constant of integration. Q_{it} is actually the D_{it} profile, weighted by the Fermi-Dirac occupation probability and integrated within the bandgap.

$$Q_{it} = -q\int_{E_i}^{E_C} F(E) D_{it}\, dE \; + q\int_{E_V}^{E_i} [1 - F(E)] D_{it}\, dE \,. \qquad (3)$$

It is difficult to derive a closed form expression of D_{it} from this equation. Instead, if step-like Fermi-Dirac function is assumed, the D_{it} profile can be extracted from the Q_{it} profile.

$$D_{it}(E) \approx \frac{1}{q} \times \left| \frac{dQ_{it}(E)}{dE} \right| . \qquad (4)$$

Of course, extraction done using Eq. (4) introduces some error, especially near the band edges, but the error can be easily corrected using the anti-symmetric property of Fermi-Dirac distribution. This point will be properly explained in section III.

III. Results

The results of D_{it} extraction technique are presented in this section. First low frequency gate C-V characteristics of MOSFETs are simulated assuming certain distribution of interface trap states. All calculations are performed at room temperature. Since C-V characteristics is not sensitive to electron (hole) effective mass, m_{di} in gate dielectric [1], we use a constant value of $m_{di} = 0.5m_0$. Fig. 2 shows the calculated low frequency C_g as a function of V_g for the following MOS structure: Al/SiO_2/Si with T_{di} = 1.0 nm. The substrate doping density is 5×10^{15} cm^{-3} for p-Si. The work function of Al is 4.1 eV, whereas the potential barrier height, Φ_b at dielectric-Si interface is 3.1 eV for electrons and 4.68 eV for holes. It is observed that the plateau of C_g is raised when a uniform D_{it} profile of 5×10^{12} eV^{-1} cm^{-2} is considered. Due to the contribution of interface trap capacitance C_{it} and the modification of the surface electric field by Q_{it}, C_g is decreased both under inversion and accumulation bias voltages. In Fig. 2, both donor and acceptor type trap states are considered.

Fig. 3 shows the extracted C_p versus φ_s plot from the C-V shown in Fig. 2. The simulated C_s for the ideal MOS structure without any interface traps is also presented in Fig. 3. The difference between the two curves represents C_{it} as shown in Fig. 4. Here it should be noted that both the curves merge together once they enter strong inversion or accumulation. This is due to the fact that the D_{it} profile exists only within the bandgap. Fig. 5 shows the extracted D_{it} profile and the corrected profile as well as the actual profile with which gate C-V is simulated in Fig. 2.

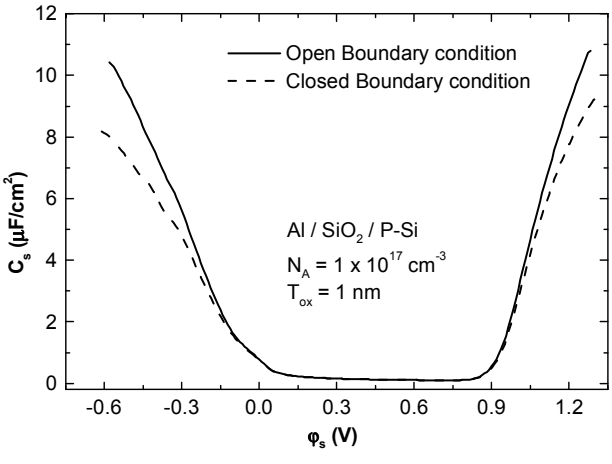


Fig. 1 Theoretically calculated C_s versus φ_S curves using open and close boundary conditions respectively. Effects of wave function penetration are clearly visible in both accumulation and inversion.

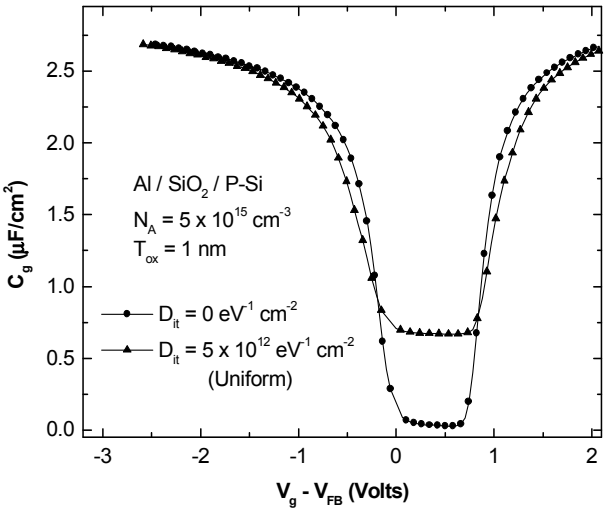


Fig. 2 C_g versus ($V_g - V_{FB}$) curve for Al/SiO_2/Si MOS structure with & without uniform D_{it} profile.

The raw extracted D_{it} profile comes from Fig. 4 when it undergoes successive integration and differentiation according to Eqs. (2) and (4) respectively.

Location of the Fermi-level within the bandgap at the surface can easily be determined since both φ_s and the Fermi potential are known. When the raw result is compared with the actual one, a closer inspection reveals that the extracted profile is smeared near both the band-edges. It is expected and the reason is cited below. Actual Fermi-Dirac distribution is not a step curve at room temperature but rather has smearing symmetry. When the Fermi level just crosses the valence (conduction) band edge, trap states immediately above E_V (below E_C) are not completely empty (filled), rather they have a small probability to remain occupied (unoccupied) because of the tail of the Fermi-Dirac function at room temperature.

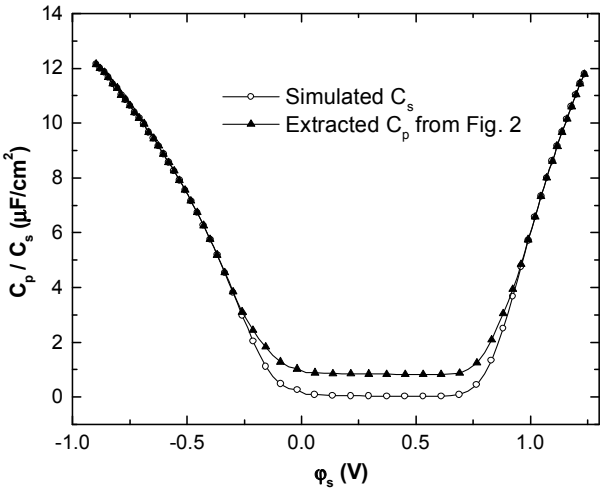


Fig. 3 Extracted C_p and simulated C_s versus φ_S curves for the C-V shown in Fig. 2.

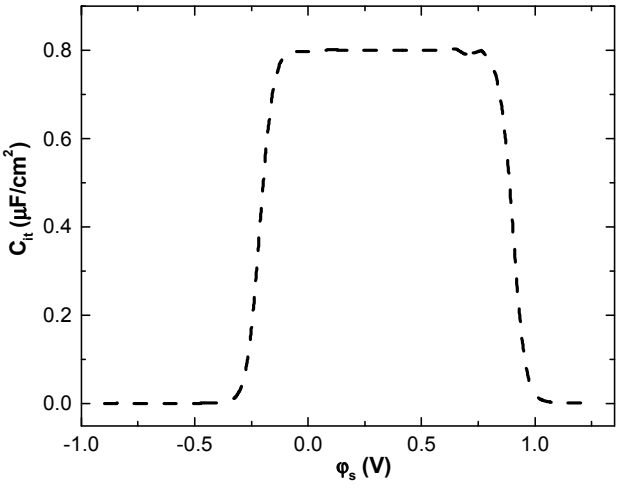


Fig. 4 C_{it} versus φ_S plot extracted from the difference between C_p and C_s shown in Fig. 3.

As a result the extracted trap charge profile (not shown here) outside the bandgap continues to change for a while before coming to a constant level. This is equivalent to having trap states just outside the band-edges when step-like Fermi-Dirac function is used for Q_{it} calculation. In order to get the corrected D_{it} profile, the raw extracted D_{it} profile outside the band-edges are reflected and then added to the D_{it} profile just inside the respective band-edges so that it approaches the actual profile. Fig. 5 clearly shows that the extracted profile after correction closely resembles the actual profile.

To show the robustness of our extraction technique, the proposed technique of D_{it} extraction has been applied to simulated gate C-Vs with more realistic U-shaped D_{it} profiles and excellent agreement has been found. The results are shown in Fig. 6 and Fig. 7 respectively.

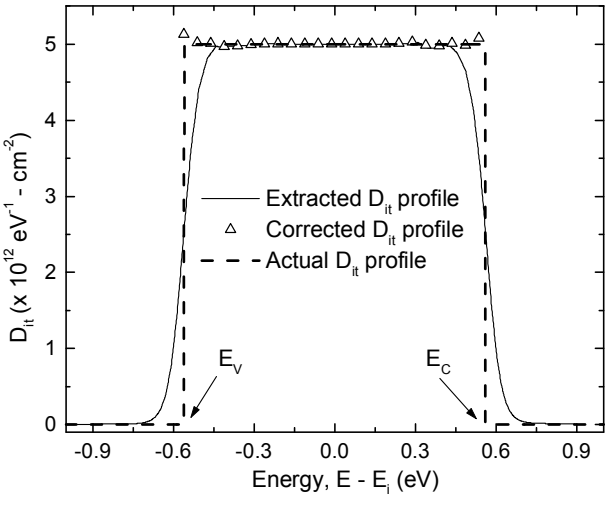


Fig. 5 Comparison of extracted D_{it} profile (before and after correction) with the actual one. Correction is based on the smearing symmetry of the Fermi-Dirac distribution.

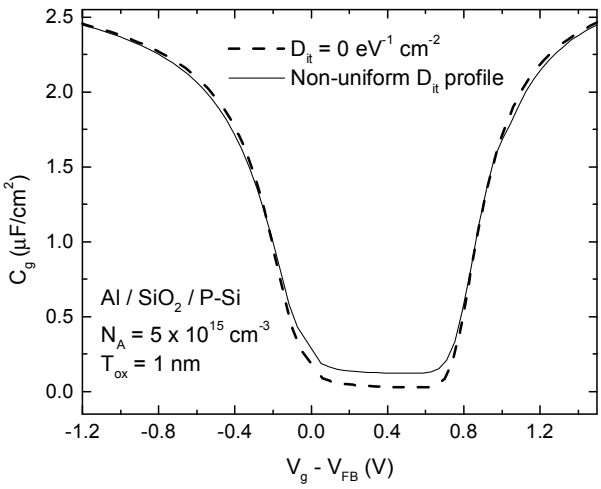


Fig. 6 C_g versus ($V_g - V_{FB}$) curve for Al/SiO_2/Si MOS structure with a U-shape D_{it} profile. The C-V of the ideal structure without any D_{it} is also shown for comparison.

Here it should be mentioned that the accuracy of the extraction technique is dependent on the number bias points in the C-V curve. If the D_{it} profile exhibits sharp change, then more bias points in the C-V curve would yield a better accuracy after D_{it} extraction and subsequent correction.

From the results reported here, it is clearly evident that if trap states contribute capacitance towards C_g, its signature will be present in the gate C-V characteristics from which D_{it} distribution can be extracted with reasonable accuracy in a rather straight forward way.

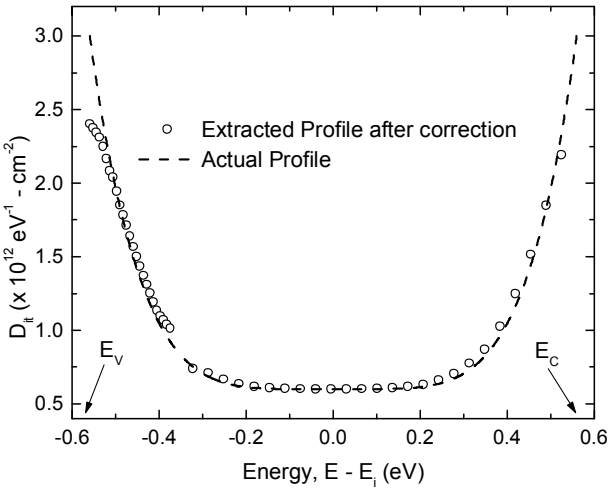


Fig. 7 Comparison of extracted D_{it} profile from Fig. 6 (after correction) with the actual one. The accuracy of extraction can be improved if gate C-V contains more bias points.

Unlike many other D_{it} extraction techniques, the proposed method is also suitable for MOS devices with high-K gate dielectrics. The proposed technique provides a simple way for characterizing the interface trap density profiles of various high-K gate dielectric materials within the entire Si bandgap region.

IV. Conclusion

We have presented a simple, but accurate D_{it} extraction technique from low frequency C-V characteristics of MOS devices with ultrathin high-K gate dielectrics. Existing low frequency C-V method has been modified and improved to incorporate QM effect with wave function penetration for theoretical calculation of MOS electrostatics. Excellent agreement has been found between extracted and actual D_{it} profiles of simulated gate C-V characteristics. It is expected that the proposed method can be effectively used for characterizing interface trap density distributions of high-K gate dielectric materials.

References

[1] M. M. Satter and A. Haque, "Modeling Effects of Interface Trap States on the Gate C-V Characteristics of MOS Devices with Ultrathin High-K Gate Dielectrics," in Proc. EDSSC 2007, Tainan, Taiwan, pp. 157-159, 20-22 December, 2007.

[2] E. H. Nicollian and A. Goetzberger, "MOS Conductance Technique for Measuring Surface State Parameters," Appl. Phys. Lett. vol. 07, no. 08, pp. 216-219, October 1965.

[3] P. V. Grey and D. M. Brown, "Density of SiO_2-Si Interface States," Appl. Phys. Lett. vol. 08, no. 02, pp. 31-33, January 1966.

[4] L. M. Terman, "An Investigation of Surface States at a Silicon/Silicon Oxide Interface Employing Metal-Oxide-Silicon Diodes," Solid-State Electron., vol. 05, no. 05, pp. 285-299, September 1962.

[5] C. N. Berglund, "Surface States at Steam-Grown Silicon-Silicon Dioxide Interfaces," IEEE Trans. Electron Devices, vol. 13, no. 10, pp. 701-705, October 1966.

[6] M. Kuhn, "A Quasi-Static Technique for MOS C-V and Surface State Measurements," Solid-State Electron., vol. 13, no. 06, pp. 873-885, June 1970.

[7] A. Koukab, A. Bath, and E. Losson, "An Improved High Frequency C-V Method for Interface State Analysis on MIS Structures," Solid-State Electron., vol. 41, no. 04, pp. 635-641, April 1997.

[8] A. Pacelli, A. L. Lacaita, S. Villa, and L. Perron, "Reliable Extraction of MOS Interface Traps from Low-Frequency CV Measurements," IEEE Electron Device Lett., vol. 19, no. 05, pp. 148–150, May 1998.

[9] A. E. Islam and A. Haque, "Accumulation gate capacitance of MOS devices with ultrathin high-κ gate dielectrics: modeling and characterization," IEEE Trans. Electron Devices, vol. 53, no. 6, pp. 1364–1372, June 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Adaptive Neuro - Fuzzy Inference Systems into Squirrel Cage Induction Motor Drive: Modeling, Control and Estimation

G. Pandian
Research Scholar
Electrical and Electronics Engg Dept
Sathyabama University
Chennai, India
E-mail: pandian1960@yahoo.co.in

S. Rama Reddy
Professor
Electrical and Electronics Engg Dept
Jerusalem College of Engineering
Chennai, India
E-mail: srr_victory@yahoo.com

Abstract

This paper presents application of Adaptive Neuro-fuzzy inference system (ANFIS) into a squirrel cage induction machine towards modeling, control and estimation. This paper contributes (i) Development of a simple and more realistic model of the induction motor using ANFIS. Using ANFIS, the parameter sets of the motor model are estimated. The simplified model contains eleven estimated parameters. In this paper, a new estimation technique for modeling of induction motor is presented. The identified model can be utilized for electric drives. (ii) Speed, torque and flux control using direct torque control (DTC) algorithm with ANFIS (iii) Design of Estimator through ANFIS which estimates the stator resistance with reference to the temperature when the DTC algorithm is involved. Better estimation of stator resistance results in the improvements in induction motor performance using DTC thereby facilitating torque ripple minimization. The values of stator voltage (V_s), stator current (I_s) and rotor angular velocity (ω_r) are taken from the free acceleration test data of 10 HP motor for simulation.

Key words **-** Parameters, direct torque control, ANFIS, induction motor, modeling, Stator resistance, estimation.

1. INTRODUCTION

In most of the work on induction machine modeling, the parameters of the induction machine are derived from the steady-state data, such as the no load and blocked rotor tests, or the small signal test, such as the standstill frequency response test. The non-linear properties of the mutual and the leakage inductances of the machine are considered. The stator and rotor saturations are represented by specially designed saturation functions. Induction motors can be modeled by both the linear and non-linear equation and the parameters for the specific operating condition are estimated. By using linear equations, time domain parameters are estimated using advanced techniques such as extended Kalman filter and the output error method and these techniques are not suitable to identify the non linear model [2][3]. In this paper, a new approach to identify the non-linear model of induction machine from the free acceleration test data is presented. The measurements used for the modeling are obtained by applying the three-phase ac power to an induction machine while it is in stand still condition. The proposed technique presented in this paper is adaptive neuro-fuzzy inference system for parameter estimation and a system model has been developed. The identified model can be utilized for the on-line computer controlled electric drive system.

2. GENERALIZED INDUCTION MOTOR MODEL

The mathematical model of an induction motor in space vector notation, established in d-q co-ordinate system rotating at speed ω_s is given by the following equation

$$V_s = R_s I_s + \frac{d\psi_s}{dt} + j\omega_s \psi_s \quad (1)$$

$$0 = R_r I_r + \frac{d\psi_r}{dt} + j(\omega_s - \omega_r)\psi_r \quad (2)$$

The above equations are also formulated in synchronously rotating frame as

$$V_{qs} = R_s I_{qs} + \frac{d\psi_{qs}}{dt} + \omega_s \psi_{ds} \quad (3)$$

$$V_{ds} = R_s I_{ds} + \frac{d\psi_{ds}}{dt} + \omega_s \psi_{qs} \quad (4)$$

$$0 = R_r I_{qr} + \frac{d\psi_{qr}}{dt} + (\omega_s - \omega_r)\psi_{dr} \quad (5)$$

$$0 = R_r I_{dr} + \frac{d\psi_{dr}}{dt} + (\omega_s - \omega_r)\psi_{qr} \quad (6)$$

where ω_r and ω_s represent the rotor and synchronously rotating reference frame angular velocities, respectively.
The stator and rotor winding flux linkages are expressed as

$$\psi_{qs} = L_m.\ (I_{qs} + I_{qr}) + L_{ls}.\ I_{qs} \quad (7)$$
$$\psi_{ds} = L_m.\ (I_{ds} + I_{dr}) + L_{ls}.\ I_{d} \quad (8)$$
$$\psi_{qr} = L_m.\ (I_{qs} + I_{qr}) + L_{lr}.\ I_{qr} \quad (9)$$
$$\psi_{dr} = L_m.\ (I_{ds} + I_{dr}) + L_{lr}.\ I_{dr} \quad (10)$$

The parameters such as the mutual inductance, leakage inductance and flux linkages are considered as functions of stator and the rotor currents are given as

$$L_m = L_m\ (I_{qs}, I_{qr}, I_{ds}, I_{dr}) \quad (11)$$
$$L_{ls} = L_{ls}\ (I_{qs}, I_{ds}) \quad (12)$$
$$L_{ls} = Lls\ (I_{qs}, I_{ds}) \quad (13)$$

The mutual inductance is affected by all the stator and rotor currents and it appears in the equations given in (7) – (13). The flux linkage equations should be considered as functions of all the currents expressed as

$$\psi_{qs} = \psi_{qs}\ (I_{qs,}\ I_{ds,}\ I_{qr,}\ I_{dr}) \quad (14)$$
$$\psi_{ds} = \psi_{ds}\ (I_{qs,}\ I_{ds,}\ I_{qr,}\ I_{dr}) \quad (15)$$
$$\psi_{qr} = \psi_{qr}\ (I_{qs,}\ I_{ds,}\ I_{qr,}\ I_{dr}) \quad (16)$$
$$\psi_{dr} = \psi_{dr}\ (I_{qs,}\ I_{ds,}\ I_{qr,}\ I_{dr}) \quad (17)$$

A simplified model is derived by assuming that the d- and q-axes incremental inductances are affected by only their own axes currents.

The simplified model contains eleven parameters to be estimated which are R_s, R_r, L_{ls}, L_{lr}, L_m, L_{lqs}, L_{lqr} L_{mq}, L_{lds}, L_{ldr} and L_{md}. Here the simplified model is used to represent the dynamic behavior of the induction machine.

3. PROPOSED ANFIS MODEL OF INDUCTION MOTOR

Fig. 1 shows the complete estimation block diagram of parameter estimator. The structure of this model is used for representing the parameters. The input values of this model is composed of the magnitudes of the d- and q-axis currents and the rotor angular velocity given as

$$p_s = \{|I_{ds}|, |I_{qs}|, \omega_r\} \quad (18)$$

The magnitudes of d and q-axes stator currents are directly related to the saturation of the stator and ω_r is associated with the rotor saturation and the rotor winding resistances. In previous studies [3], neural network alone is used to identify the parameters of the induction motor. In this paper, an efficient ANFIS estimator is introduced to estimate the parameters of the induction machine from the d- and q-axis stator current values. The system is modeled using MATLAB/SIMULINK [4] combining both fuzzy logic and artificial neural networks taking all of the advantages. It uses the training data set to build the fuzzy system, whose membership functions are adjusted using the back propagation algorithm, allowing that the estimator learns with the data that it is modeling [10][11]. Neuro-fuzzy network maps the inputs by the membership functions and their associated parameters and so through the output membership functions and corresponding associated parameters [5][13]. These will be the synaptic weights and bias, and are associated to the membership functions that are adjusted during the learning process. The computational work to obtain the parameters is helped by the gradient descendent technique.

The operations of ANFIS system at the MATLAB [4] are as detailed below:

1. A set of membership functions has to be chosen.

2. The input output training data is used by ANFIS. It starts making a clustering study of the data to obtain a concise and significant representation of the system's behavior. It is important to note that the system has a good modeling if the training set has enough representative, i.e., it has a good data distribution to make possible to interpolate all necessary values of the system. The clustering technique used was the fuzzy c-means. After setting the number of clusters that are estimated to compose the data, the cluster's centers are searched in an iterative way based on minimizing an objective function.

The proposed ANFIS induction motor model is shown in Fig.1. Three phase stator currents of induction motor are converted into two phase d- axis and q- axis currents and along with the rotor speed given to ANFIS estimator. Various parameters are taken as output from the estimator.

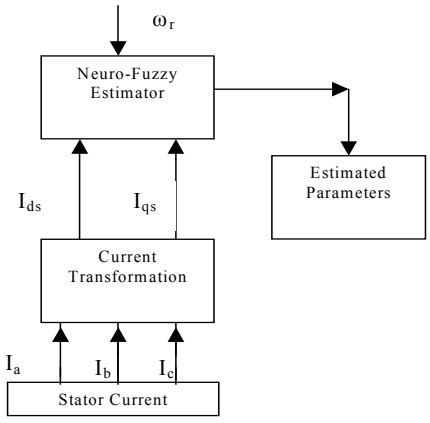


Fig. 1. ANFIS model for induction motor

4. DIRECT TORQUE NEURO FUZZY CONTROL

Direct torque control (DTC) is a relatively novel induction motor control method, that is relatively easy to implement and that enables high performance to be achieved. Direct torque control of induction motor is based on the theory of field oriented control (FOC) and direct self control (DSC) [5]. The core of DTC consists of hysteresis controllers of torque and flux, optimal switching logic, precise motor model. The motor model calculates the torque, stator flux and shaft speed based on the measurement of two-phase currents and voltages. Torque and flux references are compared with these values, and control signals are produced using a two-level flux hysteresis and a three level torque hysteresis. The optimal switching logic defines the best voltage vector based on torque and flux references [5]. The expression for the developed torque of an induction motor is given by equation (19).

$$T = \frac{3}{2}\frac{P}{2}\left(\psi_{ds} i_{qs} - \psi_{qs} i_{ds}\right) \qquad (19)$$

Stator flux is a computational quantity, which is obtained using the stator-measured current *' I_s'* and voltage *'V_s'* as given in equation (20)

$$\psi_s = \int_0^t \left(V_s - I_s R_s\right) dt \qquad (20)$$

Stator resistance is assumed as constant in ideal condition. However, practically, the stator flux and electromagnetic torque developed are dependent on stator resistance, which has a wide variation of 0.75 to 1.7 times of its nominal value due to temperature change and stator frequency variation [18]. The variation of stator resistance deteriorates the drive performance by introducing errors in the estimated flux linkages magnitude and its position and hence in the electromagnetic torque. The error between the stator current phasor reference and its measured value is a measure of stator resistance variation from its set value [4]. In this paper, application of ANFIS into DTC is done in the following ways:

(i) Direct Torque Neuro Fuzzy Control without stator resistance tuning.

(ii) Direct Torque Neuro Fuzzy Control with stator resistance tuning.

Both the systems are explained and results are analyzed in detail.

Fig.2 shows the block diagram of adaptive direct torque neuro fuzzy control of induction motor. Torque and stator flux are estimated mathematically from the motor signals. ANFIS is used as controller to which the torque and flux errors along with position of stator flux are given as inputs and from which inverter-switching states are estimated. The neuro-fuzzy structure used in proposed control strategy has five network layers as given below:

- Layer 1: Every node in this layer contains membership functions.
- Layer 2: This layer chooses the minimum value of two input weights.
- Layer 3: Every node of these layers calculates the weight, which is normalized.
- Layer 4: This layer includes linear functions, which are functions of the input signals.
- Layer 5: This layer sums all the incoming signals.

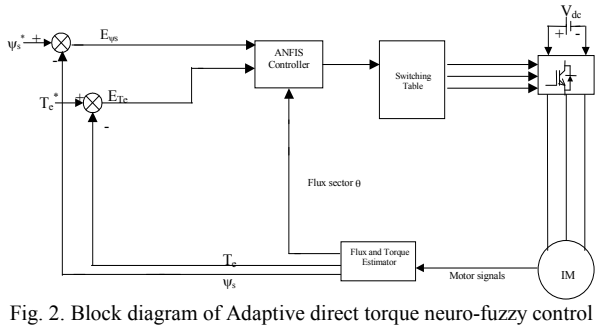


Fig. 2. Block diagram of Adaptive direct torque neuro-fuzzy control

By using Neuro fuzzy system, efficient estimation of switching state is possible.

5. ESTIMATION OF STATOR RESISTANCE

The DTC is based on the evaluation of two quantities that are stator flux and torque. Exact evaluation of *'ψ_s'* requires accurate measurements and good evaluation of 'R_s'. The value of 'R_s', which varies with temperature, needs either an accurate thermal model or an evaluation and estimation method. A comprehensive thermal model for the induction motor, capable of estimating 'R_s', requires detailed information about the motor that is usually unavailable. Estimation is the only method used recently. The sensitivity of the DTC to temperature variations, leading to stator resistance changes, is eliminated by online estimation of stator resistance. This paper proposes an adaptive Neuro-fuzzy method of stator-resistance estimation of an Induction motor. In this paper, an estimator, which is designed through ANFIS for stator resistance estimation with reference to the temperature, has also been discussed. The stator winding resistance primarily varies with winding temperature, which is given by the following equation [20].

$$R_t = R_{to} + \alpha R_{to} (T_t - 25°C) \quad (21)$$

Where R_t is the resistance at T°C, R_{to} is the initial resistance at room temperature, T_t is the stator winding temperature and 'α' is the temperature co-efficient of copper, which is equal to 11.21 X 10^{-3}/ °C. If a temperature- sensing thermistor is inserted in a distributed manner in the stator winding, the stator winding temperature can be monitored and correspondingly, stator resistance can be estimated accurately by using equation (21). However, the use of such temperature sensors in direct torque control drive is not acceptable [16] [18]. The stator winding temperature with reasonable accuracy can be predicted by using adaptive systems. Basically, the losses in the machine contribute to stator winding temperature rise, and those losses classified as stator copper loss, rotor copper loss, stator iron loss, rotor iron loss and some amount of stray loss. The heat generated by the losses flow through distributed parameter thermal equivalent circuit of the motor and cause temperature rise at different parts [17]. The stator copper and iron losses will contribute to stator winding temperature rise. In this paper, stator resistance is derived from adaptive neuro fuzzy inference system based estimation of stator winding temperature.

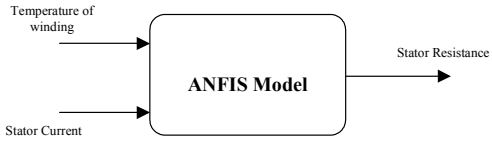


Fig.3. ANFIS model for stator resistance estimation

Fig. 3 shows the proposed estimator for stator resistance using ANFIS. The behavior of the estimator model is studied for varying temperature. Stator resistance of the motor is 0.5814Ω is at rest. The critical values of stator resistance are chosen from the simulation results. The surface plots of the model with varying stator current and temperature at constant speeds 918, 725 and 357 rpm are shown in the Fig. 2, 3 and 4 respectively. The surface plot of the critical values of the stator resistance with varying time is also shown in the Fig. 5, 6 & 7.

6. SIMULATION PROCEDURE

A 10 H.P, three phase induction machine is considered for simulation. The values of stator voltages, stator current and rotor angular velocity are collected from the free acceleration test data for the simulation purpose. The complete model was developed and simulated using ANFIS, which is a function available in MATLAB. To represent the machine operating conditions for each subset of data, average values of I_{ds}, I_{qs} and ω_r are presented. These variables will be utilized for the input of the model and shown in Table 1. Variables are entered in matrix form in m-file. From the Table 1, the input parameter variables are used for training in ANFIS. The membership functions are formed and from that the output parameter variables are estimated accurately. Output parameter variables such as R_s, R_r, L_{ls}, L_{lr}, L_m, L_{lqs}, L_{lqr} L_{mq}, L_{lds}, L_{ldr} and L_{md} are estimated accurately. Estimated parameter values are compared with the actual values and errors found out. Similarly, ANFIS model has been developed for DTNFC with and without stator resistance estimation or tuning. For direct torque control strategy, IGBT based PWM inverter and the same 10 H.P induction machine are considered for simulation.

7. SIMULATION RESULTS AND DISCUSSIONS

Simulation results are obtained for ANFIS model of squirrel cage induction motor, direct torque neuro fuzzy control with and without stator resistance tuning systems. They are discussed in the following sections.

7.1 MODELING OF INDUCTION MOTOR USING ANFIS

The behavior of the estimator model is studied for various values of stator current and rotor angular velocity. The critical values of stator resistance, Rotor resistance, and other parameter variables are chosen from the simulation results. The surface plots of the model for varying stator current and rotor angular velocity are shown in Fig. 4 – 14. From these figures, it is clear that the output variables are estimated from the input variables such as stator current in d-axis, stator current in q-axis and per- unit rotor angular velocity. Fig. 15, 16 and 17 showing the stator resistance, rotor resistance and mutual inductance variations under comparative study of training data and ANFIS output respectively.

7.2 DIRECT TORQUE NEURO FUZZY CONTROL (DTNFC) OF INDUCTION MOTOR WITH AND WITHOUT ESTIMATING OR TUNING STATOR RESISTANCE.

The simulation of DTNFC scheme is also done using MATLAB / SIMULINK. The stator resistance compensator computes the stator current phasor error by comparing the stator current command and the measured value of stator current. The simulation result of DTNFC without stator resistance tuning is shown in Fig. 18 and with stator resistance tuning is shown in Fig. 19. From Fig. 18, it can be seen that the DTNFC scheme reaches the reference value of torque 50 Nm, in less than 2.5ms thereby providing fast dynamic response. Fig. 19 shows the simulation result of DTNFC scheme with stator resistance compensation. The comparison of the DTNFC scheme with and without tuning can be done based on the Speed of response and Number of switching of inverter.

7.2.1 TORQUE RESPONSE

From Fig. 18 and Fig. 19, it can be seen that the torque response of the proposed scheme without stator resistance tuning and with tuning resembles each other. Hence there is no need for stator resistance tuning the applications, which consider only the response time.

7.2.2 SWITCHING OF INVERTER

The DTNFC scheme without stator resistance tuning requires more number of inverter switching when compared with the results of DTNFC scheme with tuning. This is because, the output voltage of inverter (V_{ab}) for the former changes more frequently than in the latter case. This can be seen from the inverter output voltage V_{ab} shown in Fig. 18 and Fig. 19. Hence, the DTNFC scheme with adaptive stator resistance compensator provides good torque response with reduced number of inverter switching.

The proposed DTNFC scheme has the features and advantages of simple tuning procedure, Constant switching frequency, Fast torque and flux response, Accurate flux and torque estimation with stator resistance tuning.

Table 1
Input Parameter Values Of The Proposed Model

Input variables			
No.	I_{ds} (A)	I_{qs} (A)	ω_r (p.u)
1	88.4	112.8	0.108
2	88.8	110.8	0.144
3	89.1	108.7	0.180
4	89.5	106.6	0.218
5	89.7	104.0	0.255
6	89.8	101.4	0.295
7	89.8	98.5	0.333
8	89.8	95.3	0.376
9	89.6	91.8	0.416
10	89.0	88.0	0.461
11	88.3	83.9	0.505
12	87.1	78.5	0.551
13	85.2	73.0	0.598
14	82.7	66.6	0.645
15	79.0	59.6	0.694
16	74.2	51.8	0.743
17	67.5	43.7	0.790
18	59.6	35.4	0.835
19	50.3	27.7	0.876
20	40.0	21.2	0.911

Fig. 20 & 21 show the surface plots of the stator resistance estimator model along with variation of stator current and temperature as input parameters and stator resistance as output parameter of the model at speeds of 725 rpm and 918 rpm respectively.

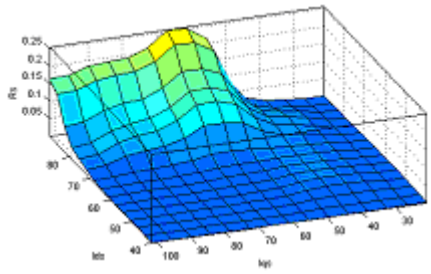


Fig. 4 Variation of stator resistance with varying stator current.

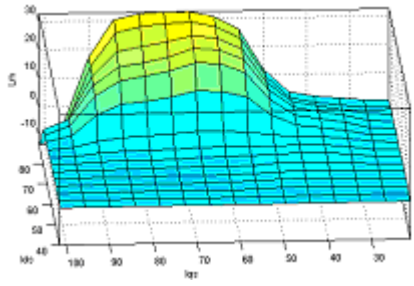


Fig. 8 Variation of mutual inductance with varying stator current.

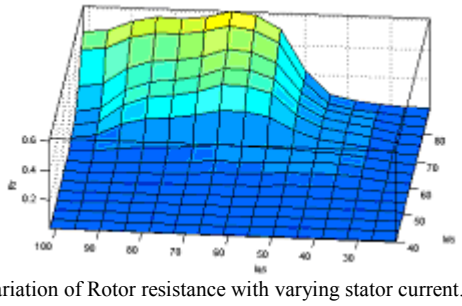


Fig. 5 Variation of Rotor resistance with varying stator current.

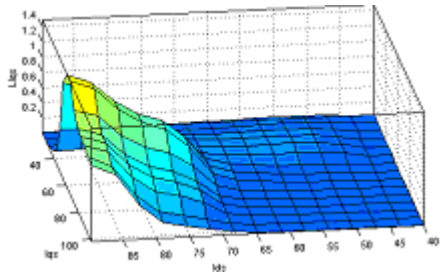


Fig. 9 Variation of stator leakage inductance in q-axis with varying stator current.

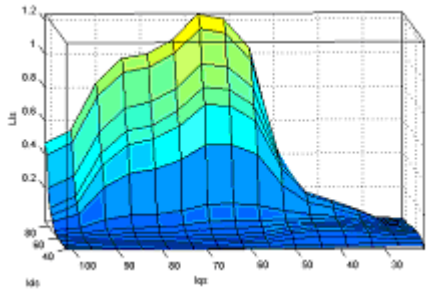


Fig. 6 Variation of stator leakage inductance with varying stator current.

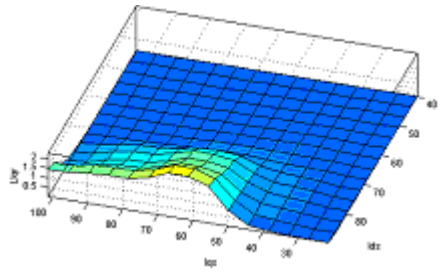


Fig. 10 Variation of rotor leakage inductance with varying stator current in q-axis.

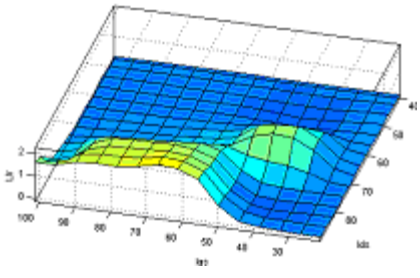


Fig. 7 Variation of Rotor leakage inductance with varying stator current.

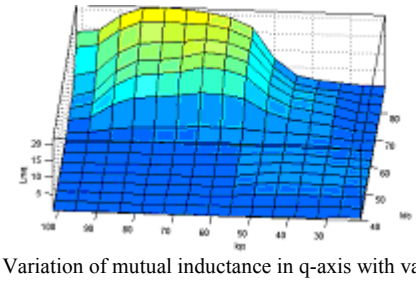


Fig. 11 Variation of mutual inductance in q-axis with varying stator current.

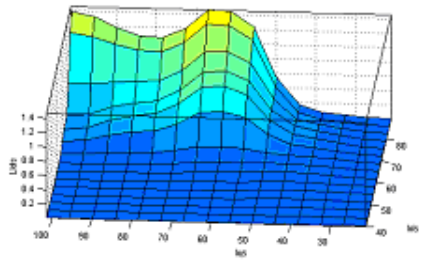


Fig. 12 Variation of stator leakage inductance in d-axis.

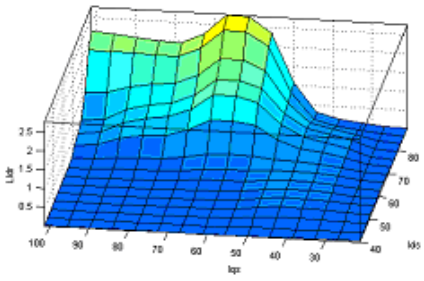


Fig. 13 Variation of rotor leakage inductance in d-axis.

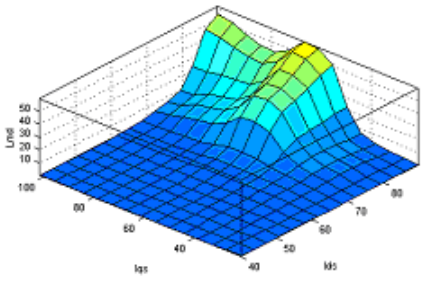


Fig. 14 Variation of mutual inductance in d-axis

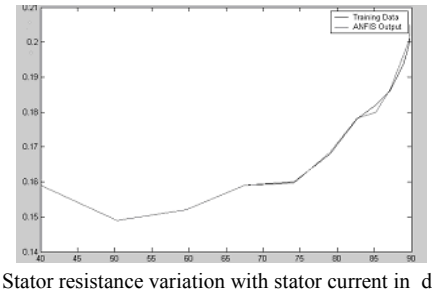


Fig. 15 Stator resistance variation with stator current in d-axis.

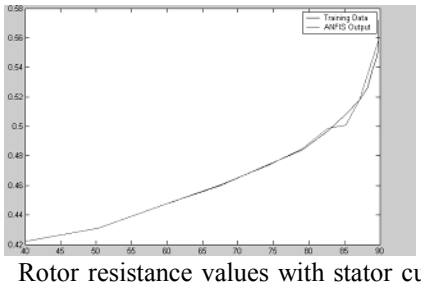


Fig. 16 Rotor resistance values with stator current in d-axis.

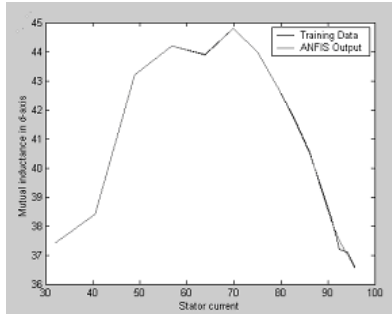


Fig. 17 Estimated values of L_{md} with variation of stator current.

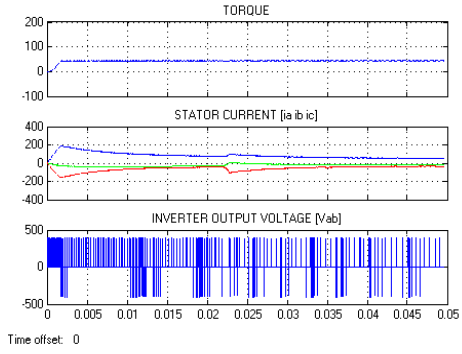


Fig.18 Torque response, stator phase currents and inverter output line voltage [V_{ab}] of DTNFC scheme with out stator resistance tuning

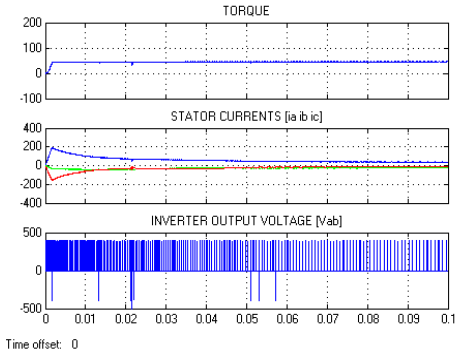


Fig. 19. Torque response , stator phase currents and inverter output line voltage [V_{ab}] of DTNFC with stator resistance tuning

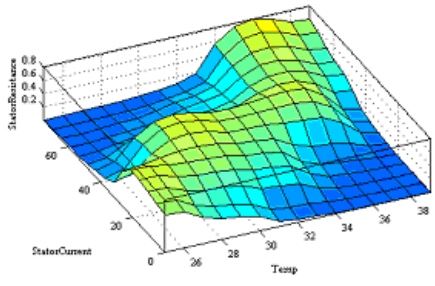


Fig. 20 Surface plot of the model with varying stator current at a constant speed of 725 rpm.

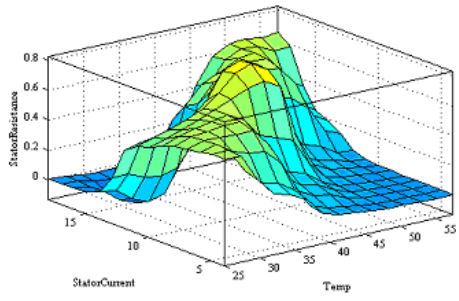


Fig. 21 Surface plot of the model with varying stator current at a constant speed of 918 rpm.

8. CONCLUSION

The paper proposes three contributions such as (i) Modeling of Induction motor (ii) Direct Torque Control of Induction Motor (iii) Estimation of stator Resistance of induction motor. All these three works have been carried out using adaptive neuro fuzzy inference systems abbreviated as ANFIS. In first contribution, the neuro-fuzzy algorithm estimates the parameters of a squirrel cage induction motor rated at 5 H.P. with operating condition changes. In second, ANFIS is applied to an advanced control methodology of induction motor named Direct torque Control and results have been obtained and in third part of the paper, stator resistance estimator has been added to direct torque control methodology with ANFIS. Proposed systems have been simulated using MATLAB/SIMULINK. The simulation results show that the estimation of parameters and control are more accurate. The results demonstrate the effectiveness of the models.

9. REFERENCES

[1] T.A. Lipo and A. Consoli, "Modeling and simulation of induction motors with saturable leakage reactances," IEEE Trans. Ind. Applications, Vol.-20, pp. 180-198,jan./Feb.1984.

[2] S.I. Moon and A.Keyhani, "Estimation of induction machine parameters from standstill domain data," IEEE Trans.Ind. Applications, Vol.30 pp. 1609-1615. Nov/Dec. 1994.

[3] Seung-Iii Moon, Ali Keyhani, Srinivas Pillutla, "Nonlinear Neural Network Modeling of an Induction Machine", IEEE Trans. On Control systems technology, Vol.7,no2, pp. 203-211. March1999.

[4] "The MATLAB compilers user's guide" in Math works hand book Math works 1994.

[5] Pawel z. Grabowski, Marian.P. Kazmierkowski, B.K.Bose and Frede Blaabjerg'" A simple Direct Torque Neuro-Fuzzy control of PWM inverter Fed Induction Motor Drive.

[6] G.J. Rogers and D.Shirmohammdi, "Induction machine modeling for electromagnetic transient program," IEEE transactions. Energy conversion, Vol, EC-2, pp. 662-668, Dec. 1987.

[7] A. Keyhani and H.Tsai, "IGSPICE simulation of induction machines with saturable inductances," IEEE transactions. Energy conversion, Vol-4, pp. 118-125, Mar. 1989.

[8] Pierre Biden, Thierry Lebey, Gerard Montrery, Clandice Nealsu and Jacques saint- Michel "Transient voltage distribution in inverter fed motor windings: Experimental study and Modeling." IEEE transactions on Power Electronics, vol.16, No.1., Jan. 2000.

[9] Robert T. Novotnak, John Chiasson, and Marc Bodson, "High – Performance Motion control of an Induction motor with Magnetic saturation" IEEE transactions on Control systems Technology, Vol.7, No.3., May. 1999.

[10] B.J.A. Krose and P.Patrick van der smagt, An Introduction to Neural Networks, The University of Amsterdam, The Netherlands, Sept. 1991.

[11] T. Fukuda and T.Shibibata, "Theory and applications of Neural networks for industrial control systems," IEEE transactions on Industrial Electronics, Vol.39, pp.472-489., Dec. 1992.

[12] K.S. Narendra and K.Parthasarthy, "Identification and control of dynamic systems using Neural networks," IEEE Trans. Neural Networks, vol. 1,pp. 4-27,Mar-1990.

[13] B.Kosko, "Neural Networks and Fuzzy systems. Englewood Cliffs, NJ: Prentice- Hall, 1992.

[14] J.R. Willis, G.J. Brock, and J.S. Edmonds, "Derivation of induction motor models from standstill frequency response test," IEEE transactions. Energy conversion, Vol, 4, pp. 608-615, Dec. 1989.

[15] J.A..A. Melkebeek and D.W.Novotny, "The influence of saturation on induction machine drive dynamics," IEEE transactions. Industrial applications, Vol, IA-19, pp. 671-681, Sep/Oct. 1983.

[16] J. Stephan, M. Bodson, and J.Chiasson, "Real time estimation of induction motor parameters," IEEE transactions. Industrial applications, Vol, 30, pp. 746-759, May/June. 1994.

[17] C.R. Sullivan and S.R. Sanders, "Modeling the effects of magnetic saturation on electrical machine control systems," in Proc. IFAC Nonlinear Contr.Syst,Design Symp., Bordeaux, France, June 1992.

[18] Pawel Z. Grawbowski, Marian P. Kazmierkowski, Bimal K.Bose, and Frede Blaabjerg, " A Simple Direct-Torque Neuro Fuzzy Control of PWM- Inverter- Fed Induction motor Drive" IEEE transactions on Industrial Electronics, Vol.47, pp.863-870, Aug 2000.

[19] B.K.Bose, "Expert system, fuzzy logic, and Neural network applications in power electronics and motion control," Proc. IEEE, Vol.82, pp. 1303-1323, Aug. 1994.

[20] Bimal K.Bose, and Nitin R. Patel, "Quasi – Fuzzy Estimation of Stator Resistance of Induction motor" IEEE transactions on Power Electronics, Vol.13, No.13 pp.401-409., May 1998.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

3-PHASE 3-LEVEL SINGLE-STAGE AC-TO-DC SERIES RESONANT CONVERTER

[1]*M.M.A. Rahman, Senior Member IEEE and* [2]*M.M.Atiqur Rahman*

[1]Associate Professor
Grand Valley State University
301 W Fulton Street, KEN 327
Grand Rapids, MI 49504
E-mail: *rahmana@gvsu.edu*

[2]Lecturer
Ahsanullah University of Science & Tech
Dhaka, Bangladesh

ABSTRACT

This paper presents a new three-phase, 3-level, pulse width modulated (PWM), boost-integrated ac-to-dc series resonant converter (SRC) with high frequency transformer isolation. This converter uses a new gating scheme suitable for integrating a boost converter with a PWM 3-level dc-to-dc SRC converter with above resonance operation. Principle of operation of this converter along with its different operating modes is presented with analysis. Based on the analysis design example of a 1 kW, 48 V, 50 kHz ac-to-dc series resonant converter is presented to explain design procedure. PSPICE simulation results of the designed converter are presented for different loads. Results show soft-switching operation, low switch voltage and low total harmonic distortion of this converter for a reasonably wide range of loads.

I. INTRODUCTION

IN new century, research on ac-to-dc converters [1,3,5-7,9,10] with power factor correction and low total harmonic distortion (THD) is consistently enjoying increasing interests. As the industry interests and applications of these converters increase, enforcing agencies are coming up with more stringent harmonic standards such as IEC 61000-3-2, ANSI/IEEE-519, etc. Implementation of these standards calls for front-end power factor correctors (PFC) requiring additional components, cost, reduced efficiency and increased size. Single-stage ac-to-dc converters [1,3,5-7] address many of these challenges. All of these converters are operating in discontinuous current mode (*DCM*) with very high peak device current and deserve further study to reduce under-utilization of its current capacity. In [1] a single-stage series resonant converter was presented. This work showed that by using resonance phenomenon the device peak current were reduced significantly but the bus voltage increased at light load and maximum input voltage. In [6-8] 3-level converters are proposed to reduce switch voltage resulting from high bus voltage. Therefore, the motivation of this study is to economise switch voltage and current simultaneously at different load conditions while power factor correction is obtained by integrating a front-end PFC with the dc-to-dc resonant stage of the converter.

This paper proposes a new three-phase, 3-level, single-stage ac-to-dc series resonant converter and investigates its behaviour for a wide load range. The objectives of this paper are: (a) to explain the operating principle of this converter with DCM operation of the integrated boost PFC, (b) to identify different modes and intervals of operation at above resonance frequency, (c) to present a design example to illustrate the design procedure, and (d) to verify converter operation and performance using PSPICE simulation results for different load conditions.

These objectives are outlined in this paper as follows: Section II presents principle of operation. Different modes and intervals are identified and analysed in Section III. Section IV presents a design example to illustrate the design steps. PSPICE simulation results are presented in Section V. Section VI concludes this paper.

II. OPERATING PRINCIPLE

The proposed three-phase single-stage 3-level ac-to-dc series resonant converter is shown in Figure 1. The integrated boost PFC stage consists of boost inductors, L_a-L_c, full-bridge rectifier, boost switches S3, S4, diodes D1, D2 and capacitors C_{b1}, C_{b2}. This boost PFC converter operates in *DCM* mode to ensure natural power factor correction. The cascaded dc-to-dc series resonant converter consists of S1-S4, D1-D4, resonant tank (L_r, C_r), (n : 1) HF transformer, output rectifier, filter capacitor, C_o and load, R_L. Boost PFC and dc-to-dc stages are integrated through S3, S4 and D1, D2 while output voltage regulation is obtained by controlling duty cycle of S1-S4. The converter operation is described as follows.

To simplify the control while ensuring the natural PFC, the boost inductor current is maintained to be in discontinuous current mode (*DCM*). At the peak of the input voltage for full-load condition the PFC is operated in just continuous current mode (*JCCM*).

A) General Solutions

The general solutions for PFC stage and dc-to-dc stage of the proposed 3-level single-phase series resonant ac-to-dc converter are presented as follows:

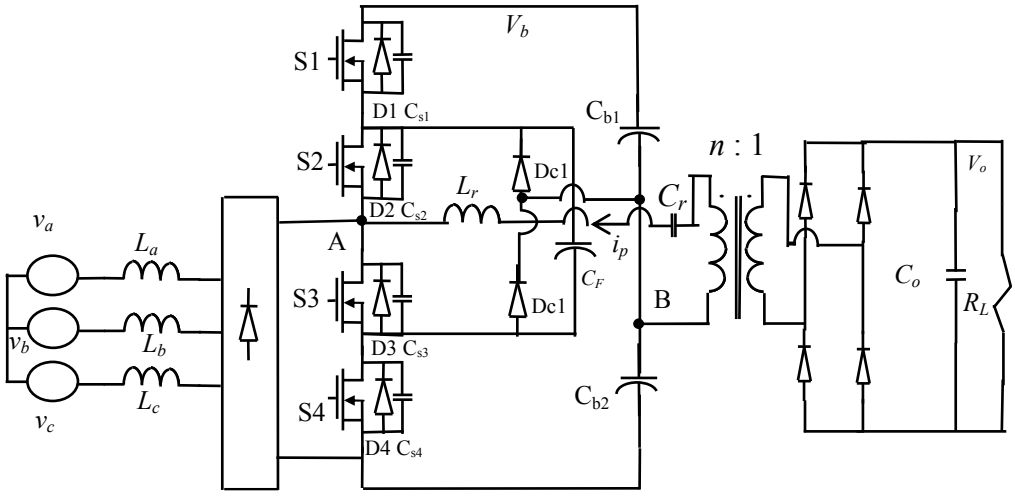


Fig. 1 Proposed three-phase single-stage ac-to-dc series resonant converter with HF isolation.

A new gating scheme [2] is used to switch the dc-to-dc series resonant converter so that voltage v_{BA} becomes a square wave. The resonant tank circuit is operated with above resonance condition (m_s >1) where resonant frequency is higher than the switching frequency to reduce switching stress [4]. At all loads all switches except S3 (at low load) undergo zero-voltage switching (ZVS) as the corresponding diodes conduct prior to the turn-on of the switches. At reduced load a zero-voltage transition (ZVT) circuit will be required to ensure ZVT operation of one of the common switches (S3).

III. CONVERTER ANALYSIS

Typical operating waveforms at different loads of the proposed 3-level single-phase single-stage series resonant converter are shown in Fig. 2. Using these waveforms and operating principle, proposed converter is analysed. This analysis is simplified based on following assumptions: (i) the switching frequency, f_s of the converter is much higher than the line frequency (60 Hz) so that during each HF switching period input voltage can be assumed constant, (ii) all active circuit components are ideal (parasitic elements, internal resistance etc. are neglected), (iii) the effect of HF transformer magnetizing inductance is neglected and the leakage inductance is considered as a part of the resonant tank inductor L_r, and (iv) the dc bus capacitors are large enough to clean the ripple in bus voltage, V_b.

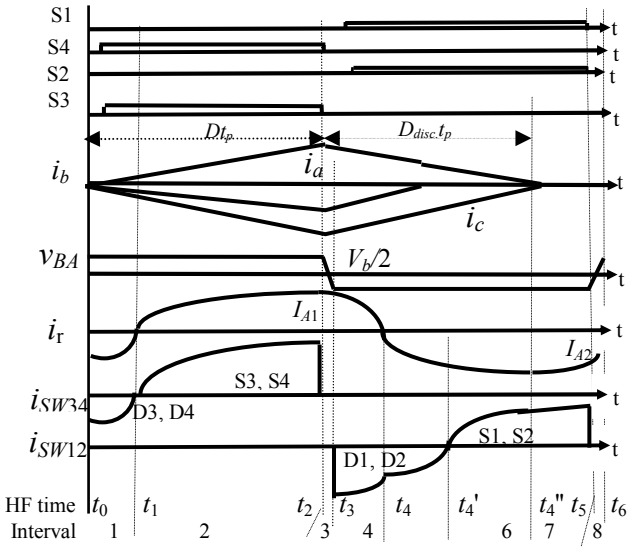


Fig. 2(a) Gating signals and operating waveforms for *RTCCM* at full load with rated input voltage.

A.1 ***Front-end boost converter***

Because of symmetry it will be enough to analyze the boost stage for duration 0 to π/6. When S3, S4 are closed any phase current, i_{phase} can be found by (1):

$$v_{phase} = L_{phase}\, di_{phase}/dt \qquad (1)$$

Solving (1) three phase currents at time, t (i_a, i_b and i_c) are found as shown in (2)-(4),

$$i_a = \frac{V_m \sin \omega t_0}{L_a} t \quad (2)$$

$$i_b = \frac{V_m \sin(\omega t_0 - 120^o)}{L_b} t \quad (3)$$

$$i_c = \frac{V_m \sin(\omega t_0 + 120^o)}{L_c} t \quad (4)$$

where ω is line frequency in rad/sec, V_m is peak line voltage and t_0 is any time within 0 to π/6.

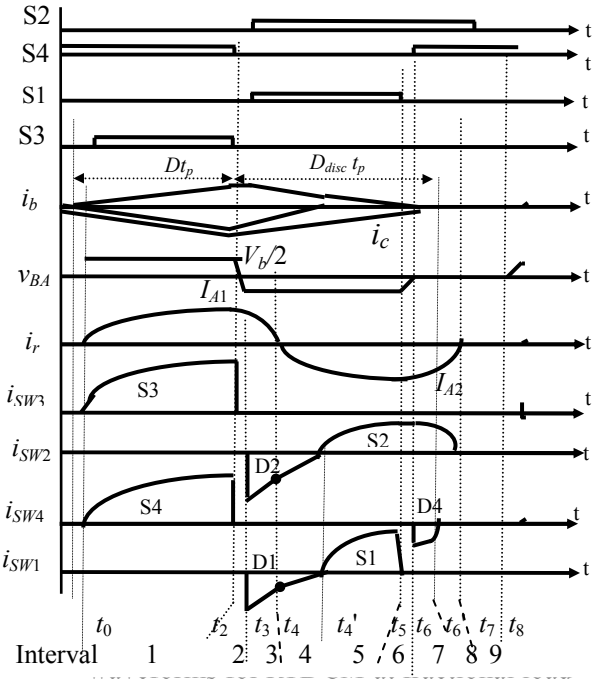


Fig. 2(b) Gating signals and operating waveforms for *RTDCM* at partial load with rated input voltage.

A.2 SRC DC-to-DC Stage

After a careful look at Fig.2, where different modes, devices conduction and intervals are clearly shown, it can be concluded that the resonant tank continuous current mode (*RTCCM*) can be expressed as a special case of the resonant tank discontinuous current mode (*RTDCM*). Hence the analysis will be done for *RTDCM* mode. In this mode the dc-to-dc series resonant stage can be analysed by using equivalent circuit shown in Fig.3. In this circuit V_o' (primary referred v_o) is positive if $i_r > 0$ and negative if $i_r < 0$ and $V_{AB} = \pm V_b/2$ or 0.

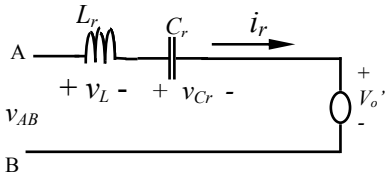


Fig.3 SRC dc-to-dc stage equivalent circuit at terminals A and B

If the *k*-th interval begins at $t = t_{k-1}$ and ends at $t = t_k$ then the generalized differential equation representing the SRC dc-to-dc stage is given by (4).

$$L_l \frac{di_r}{dt} + \frac{1}{C_r} \int_{t_{k-1}}^{t_k} i_r dt + v_{cr}(t_{k-1}) = v_{AB} - V_o' \quad (4)$$

Equation (4) can be solved to find the general solutions for i_r and v_{cr} as shown in (5) and (6).

$$i_r(t) = I_{mk} \sin\{\omega_r (t - t_{k-1}) + \theta_k\} \quad (5)$$

$$v_{cr}(t) = (v_{AB} - V_o') - \omega_r L_l I_{mk} \cos\{\omega_r (t - t_{k-1}) + \theta_k\} \quad (6)$$

where I_{mk} and θ_k are peak current and phase of resonant current at *k*-th interval and ω_r is resonant frequency in rad/s.

B) Steady State Solutions

At the design point (rated voltage and power) the proposed converter operation is expected in *RTCCM*. Therefore, the steady state solution is obtained by using proper boundary conditions to the general solutions in *RTCCM*. Following base values are used to convert these solutions in per unit.

Base voltage, $V_{base} = V_m$ (peak input voltage)
Base power, $P_b = P_{orated}$ (Rated output power)
Base current, $I_b = P_b/V_{base}$
Base inductance, $L_b = V_{base}\, t_b/I_b$
Normalized switching frequency, $m_s = f_s / f_r$
Base time, $t_b = m_s.t_p$ (HF period)

As the base voltage is the peak input voltage, the converter gain is defined as, $M = V_o/(nV_m)$.

B.1 PFC Stage

Output power, P_o and bus voltage, V_b of the front-end boost PFC stage [6] can be represented as in (7-8):

a) Output power, $P_o = \eta \frac{V_m^2 D\delta}{4L_b}$ (7)

b) DC bus voltage, $V_b = \frac{V_m}{1 - D/\delta}$ (8)

where, $D = (t_2\text{-}t_0)f_s$, $\delta = (D + D_{disc.})$, η = efficiency

B.2 SRC DC-to-DC Stage

The steady state solutions for *RTCCM* at design point are as given in (9)-(12). These solutions are identical to those for dc-to-dc stage of 3-phase BISRC [1] converter.

$$I_{m1pu} \sin\theta_1 = -I_{m2}\, pu \sin(\pi / m_s + \theta_1) \quad (9)$$

$$I_{m1pu} \cos\theta_1 + 2V_{bpu} = -I_{m2\,pu} \cos(\pi / m_s + \theta_1) \quad (10)$$

$$I_{m1\,pu} - I_{m2\,pu} = 2M \quad (11)$$

$$t_{10\,pu} = t_{43\,pu} = m_s(\pi - \theta_1)/2\pi \quad (12)$$

IV. DESIGN EXAMPLE

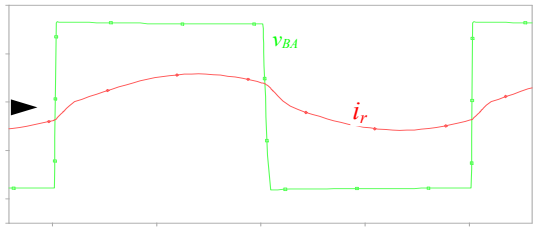
The steady state solutions presented in (7) to (12) are used in MATHCAD for numerical analysis. From this analysis following values are chosen to obtain ZVS operation [1] of all switches at design point: m_s = 1.2, M = 1.8. Using these values a 3-phase ac-to-dc SRC converter is designed with following specifications: V_{in} = 208 V line-to-line, f_s = 50 kHz, P_o = 1 KW, V_o = 48 V, η = 85%. Using these specifications as the design point, we have: n = 0.8, R_L = 2.304 Ω, L_a = L_b = L_c = 122.6 µH and L_r = 219.4 µH, C_r = 66.1 nF. Bus capacitors are selected as C_{b1} = C_{b2} = 1000 µF and flying capacitor is selected C_F = 200 µF, a large enough value to ensure [7] ZVS of S1, S4.

V. SIMULATION RESULTS

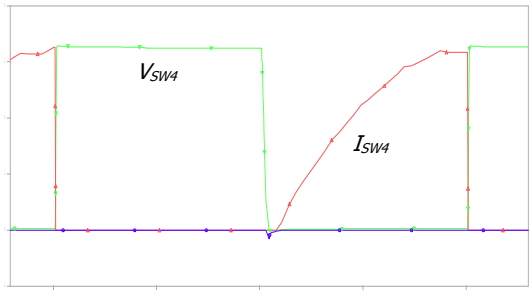
The 3-phase SRC converter designed in Section IV is simulated using PSPICE simulation program. Some typical simulation waveforms for full-load (100%) and 10% load are presented in Fig.4 and Fig.5, respectively. At full-load, as shown in Fig. 4(a), all anti-parallel diodes conduct before the conduction of corresponding switches. This is also obvious from Fig. 4(b)-(e) where switch voltage, V_{SW} and switch current, I_{SW} = I_S + I_D waveforms are shown in HF cycle. Anti-parallel diode currents, I_D are shown as negative portion while I_S are shown as positive portion of switch currents. Hence, all four switches of the dc-to-dc stage undergo ZVS operation at full-load. At 10% load, the common switch S3 loses ZVS and an auxiliary circuit must be employed to ensure their ZVT operation. It is also found that the switch voltage is always half of the bus voltage. Because of segmented sinusoidal shape of resonant tank current peak value is less compared to inductor only configurations [6-7]. Total harmonic distortion is also reasonably small for DCM converter as shown in Fig. 4(f) and 5(f). Increasing the bus voltage can further reduce this *THD*. Summary of simulation results are also provided in Table 1.

Table 1 PSPICE Simulation Results

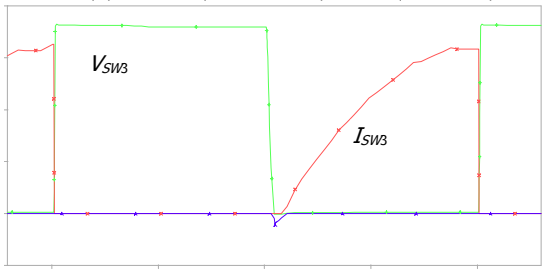
% Load		100%	10%
Duty cycle, D	p.u.	0.49	0.15
Bus voltage, V_b,	V	657	779
Peak tank current, i_r	A	5.81	1.0
Switch voltage, V_s	V	324	388
Peak voltage, v_{cr}	V	325	39
THD	%	6.55	7.57
Operating Mode		*RTCCM*	*RTDCM*



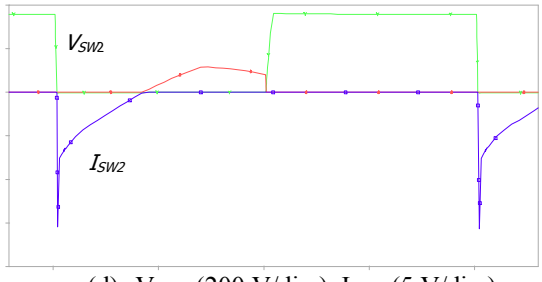
(a) v_{AB} (200 V/div), i_r (10 A/div)



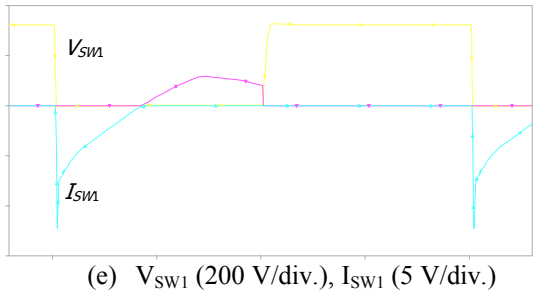
(b) V_{SW4} (100 V/div.), I_{SW4} (5 V/div.)



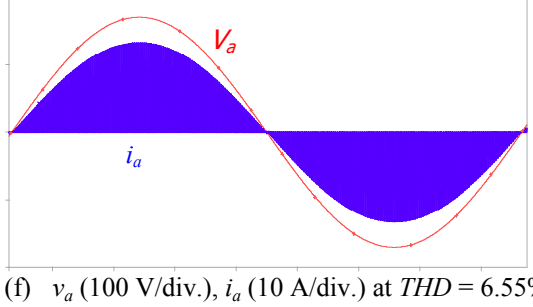
(c) V_{SW3} (100 V/div.), I_{SW3} (5 V/div.)



(d) V_{SW2} (200 V/div.), I_{SW2} (5 V/div.)

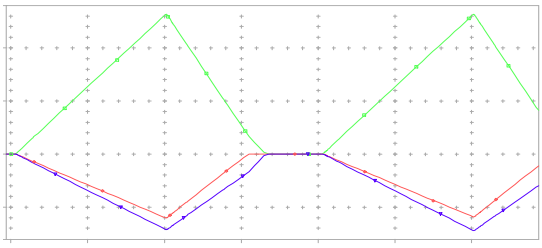


(e) V_{SW1} (200 V/div.), I_{SW1} (5 V/div.)



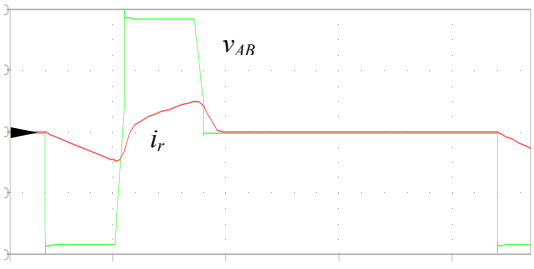
(f) v_a (100 V/div.), i_a (10 A/div.) at *THD* = 6.55%

Fig.4 (a-f) PSPICE Simulation Results Contd…..

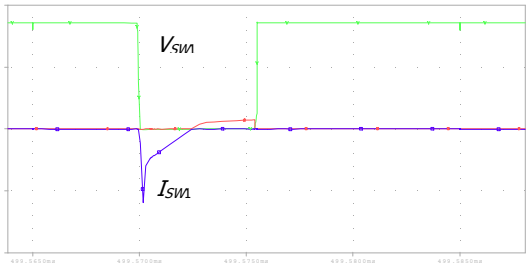


(g) Boost currents, i_a, i_b, i_c (5 A/div)

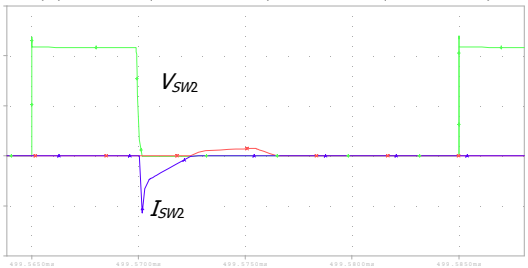
Figure 4 PSPICE simulation results at full-load



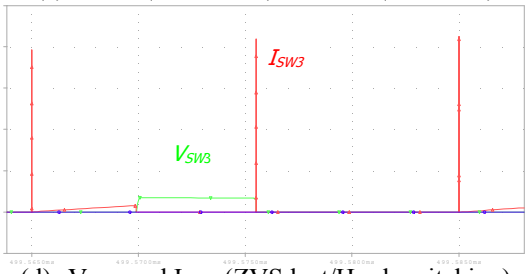
(a) v_{AB} (200 V/div), i_r (2 A/div)



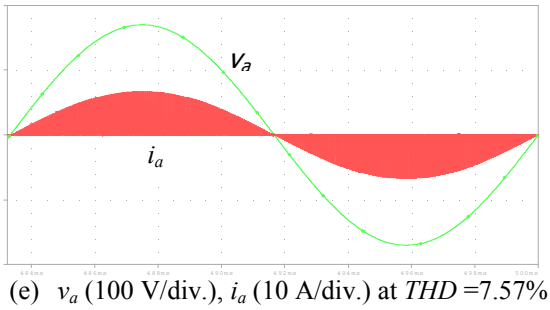
(b) V_{SW1} (200 V/div.) and I_{SW1} (10 A/div.)



(c) V_{SW2} (200 V/div.) and I_{SW2} (10 A/div.)



(d) V_{SW3} and I_{SW3} (ZVS lost/Hard-switching)



(e) v_a (100 V/div.), i_a (10 A/div.) at *THD* =7.57%

Figure 5 PSPICE simulation results at 10% load

VI. CONCLUSIONS

A new gating-scheme based three-phase, 3-level, single-stage ac-to-dc series resonant converter is proposed. Different operating modes and intervals of operation of the proposed converter are identified with analysis. PSPICE simulation results for a designed 48 V, 1.0 kW, 50 kHz converter are presented to verify these analyses and converter performance. These results show that the proposed converter enjoys soft-switching for all switches at full-load while natural power factor correction is maintained with simple control. At 10% load one switch lost ZVS operation and a ZVT circuit must be employed. Because of the segmented sinusoidal shape of resonant tank current, peak value is less compared to inductor only configurations [5-7]. Even though an asymmetrical tank voltage appeared, and a further study on flying capacitor design is under consideration to resolve this issue, this converter reduces the switch voltage to a half of the bus voltage while maintaining low THD. As increasing the bus voltage can further reduce the THD this 3-phase topology is a promising candidate for very low THD, high-voltage and high-power applications.

REFERENCES

[1] F.S. Hamdad, "Soft-switching Single-Stage, 3-phase ac-to-dc converters with high frequency transformer isolation", *Ph.D. Thesis*, University of Victoria, 1999

[2] F.S. Hamdad and A.K.S. Bhat, "A novel pusle width contro scheme for fixed frequency ZVS dc-to-dc PWM bridge converter," *IEEE Power Electronics Specialists Conference, PESC'* 99, pp. 263-268.

[3] M.M.A. Rahman, "Single-stage soft-switched HF transformer isolated ac-to-dc bridge converter and extension to multiphase converter," *Ph.D. Thesis*, University of Victoria, 1999

[4] A. Bhat and M.M. Swamy, "Loss calculations in transistorised PRC above resonance," *IEEE Trans. on Power Electronics*, Vol.4, Oct.1989, pp. 391- 401

[5] A. Rahman and A. Bhat, "A HF isolated 1-phase single-stage soft-switched ac/dc bridge converter," *PESC*'01

[6] R. Venkatraram, "A Soft-switching Single-Stage AC-to-DC Converter with Low Harmonic Distortion", *M.A.Sc. Thesis, University of Victoria, BC, May* 1998

[7] V. Dharmarajan, A. Bhat, "A Three-Level Single-Phase Single-Stage Soft-Switched AC-DC Converter", *IEEE PESC*'06

[8] M.M. A Rahman, "High-Frequency Transformer Isolated 3-Φ 3-Level Single-Stage Soft-Switched AC-to-DC Converter", *IEEE Electro-Information Technology conference, EIT* 2007

[9] F. Canales, P.M. Barbosa and F.C. Lee, "A zero-voltage and zero-current switching three-level DC/DC converter", *IEEE Transaction on Power Electronics*, vol. 17, no. 6, November 2002, pp. 898 -904

[10] M.T. Zhang, Y. Jiang, F. C. Lee and M. Jovanovic, "Single-phase Three-Level Boost Power Factor Correction Converter", *IEEE Applied Power Electronics Conference* 1995, pp. 434- 439

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Diagnostic and Protection of Inverter Faults in IPM Motor Drives Using Wavelet Transform

M. A. S. K. Khan, *Graduate Student Member,* and M. Azizur Rahman, *Life Fellow, IEEE*

Power Research Laboratory
Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, NL, Canada
E-mails: m.a.s.k.khan@mun.ca and rahman@engr.mun.ca

***Abstract*- This paper presents a novel faults diagnostic and protection technique for interior permanent magnet (IPM) motor drives using wavelet transform. The proposed wavelet based diagnostic and protection technique for inverter faults is developed and implemented in real-time for a voltage source inverter fed IPM motor. In the proposed technique, the motor currents of different faulted and unfaulted conditions of an IPM motor drive system are preprocessed by wavelet packet transform. The wavelet packet transformed coefficients of motor currents are used as inputs of a three-layer wavelet neural network. The performances of the proposed diagnostic and protection technique are investigated in simulation and experiments. The proposed technique is experimentally tested on a laboratory 1-hp IPM motor drive using the ds1102 digital signal processor board. The test results showed satisfactory performances of the proposed diagnostic and protection technique in terms of speed, accuracy and reliability.**

I. Introduction

The early and correct diagnosis of a failure in an electrical drive ensures optimum reliability, maximum safety, timely maintenance, and preventive rescue of electric motors in both low and high power applications. Numerous methods have been developed in the last fifteen years to detect faults in electric drives. The types of machines that have been investigated include induction machine [1]–[2], brushless dc machine [3]–[4], permanent magnet dc machine [5]–[7], and permanent magnet ac machine [8]–[12]. The advances in permanent magnet quality, digital electronics, and control theory have increased the use of permanent magnet synchronous motors in ac motor drives recently to fulfill the requirements of high performance industrial drive systems. Among permanent magnet motors, the interior permanent magnet (IPM) motor has magnets buried in the rotor core and shows properties such as robustness, rotor physical non saliency, and small effective air gap. With the development of power electronics and microprocessor, the IPM motors are predominantly fed from pulse width modulation (PWM) inverters for variable speed operation. However, the PWM voltage or current source inverters fed IPM motor drives are sensitive to various faults including faults in the input rectifier, power inverter, control sub-system, and motor phases. When one of these faults occurs, the drive system has to be stopped for maintenance. As a result, the fault diagnosis scheme in a practical motor drive system must perform the following tasks: fault detection, fault identification, and remedial actions. The basic requirement in the development of a fault diagnosis scheme lies on comprehensive understanding of the regular system operation so that its behavior can be compared to those at the onset of faults. The drive system should incorporate suitable diagnostic techniques in identifying and isolating the faulty elements in order to take appropriate remedial actions against faults. Several authors [13]–[15] have proposed fault-tolerant operating strategies for IPM motor drives. These are based on the principle of system redundancy. However, the use of some of these techniques may cause additional problems such as large motor currents, oscillations in electromagnetic torque, and flowing of zero-sequence currents with neutral connected. Research also has been done for faults diagnostics in line-fed IPM motors. Some of these techniques are based on the measurement of negative sequence components of line currents. However, the negative sequence components are strongly attenuated in the closed-loop controlled IPM motor drives.

The motor current signature analysis (MCSA) is used in [16] for the detection of rotor faults in a voltage source inverter fed surface mounted permanent magnet synchronous motor (PMSM). The adaptive network based fuzzy inference system (ANFIS) is used in [6] for the detection of stator winding inter-turn short circuits of a current source inverter (CSI) fed brushless dc motor. The diagnostic indices are extracted from the discrete Fourier transform (DFT) and the short-time Fourier transform (STFT) of motor currents in order to identify the number of shorted turns and location of the fault, respectively. The wavelet transform is used in [11] for detection and classification of intermittent electrical and mechanical faults in permanent magnet ac drives. The pattern of changes of electromagnetic parameters is used in [7] for detecting and isolating faults in a permanent magnet brushless dc motor. However, the demagnetization of magnets under the studied fault conditions has not been considered. The demagnetization effect of a large 1.5 MW PMSM are studied using the two-

dimensional finite element analysis in reference [10] for very limited fault scenario.

Most of the diagnostic techniques mentioned above focused on theoretical investigation of the effects of faults on voltages, currents, and torques of the IPM motor drive system. There have been little efforts in the practical implementation of these techniques. The objective of this paper is to develop and implement a novel hybrid wavelet packet transform (WPT) and wavelet neural network (WNN) based diagnostic and protection technique for inverter faults in a PWM voltage source inverter fed IPM motor drive system. The proposed WPT and WNN based diagnostic and protection algorithm is tested on-line on a laboratory 1-hp IPM motor drive using the ds1102 digital signal processor (DSP) board. Simulation and experimental results are presented to demonstrate the effectiveness of the proposed diagnostic and protection technique.

II. Wavelet Neural Network

A new wavelet based neural network is designed and implemented for the proposed diagnostic and protection technique of the IPM motor drive system. The proposed network has three layers: input layer, wavelet layer, and output layer. The hidden neurons in the wavelet layer have wavelet activation functions of different resolutions. The sigmoid functions are used in the output neurons. The modified Morlet wavelet function is used as the basis function in the wavelet layer. The modified Morlet wavelet function is defined as [17]

$$\psi_{a,b}(t) = \cos(1.75\, t_z)\, e^{-t_z^2/2\theta^2} \tag{1}$$

$$t_z = \left((t-b)/a\right) \tag{2}$$

where θ is the wavelet width, a and b are dilation and translation coefficients of wavelons in the hidden layer, respectively. Figure 1 shows the specific structure of a three-layer WNN for faults diagnostic and protection of the IPM motor drive system. The output of the WNN is calculated as

$$y(t) = \sigma(x) = \sigma\left(\sum_{j=0}^{M} v_j\ \psi_{a,b}\left(\sum_{k=0}^{L} w_{jk}\ x_k(t)\right)\right) \tag{3}$$

$$\sigma(x) = 1/(1+e^{-x}) \tag{4}$$

where y is the output, x_k is the kth component of the input vector, v_j are the connection strengths from hidden (j) to output units, w_{jk} are the connection strengths from input (k) to hidden (j) units, L and M are sum of input and hidden nodes, respectively.

A. WNN Training

The proposed wavelet network is trained using the back propagation algorithm in batches [18]-[20]. The wavelet node parameters (a, b, θ) and the network weights (v_j, w_{jk}) are adjusted to minimize the least square error. The error propagations in layers of the WNN are defined as

$$\delta_{v_j} = -\sum_{p=1}^{P}\left(d^p - y^p\right) y^p\left(1-y^p\right)\psi_{a,b}\left(net_j^p\right) \tag{5}$$

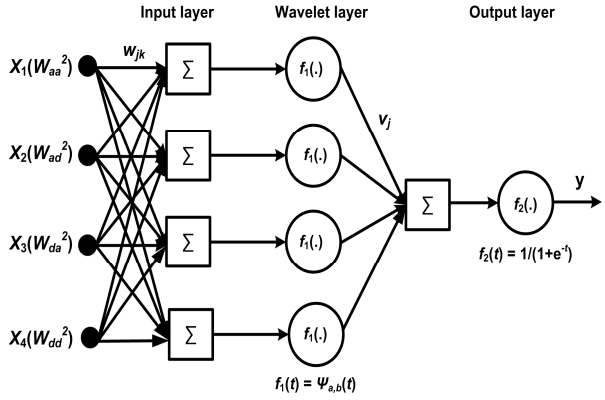


Fig. 1. Specific structure of a three-layer wavelet neural network (WNN) for faults diagnostic and protection of the IPM motor drive system.

$$\delta_{w_{jk}} = -\sum_{p=1}^{P}\left(d^p - y^p\right) y^p\left(1-y^p\right) v_j \times \psi'_{a,b}\left(net_j^p\right) x_k^p / a_j \tag{6}$$

$$\delta_{b_j} = \sum_{p=1}^{P}\left(d^p - y^p\right) y^p\left(1-y^p\right) v_j\ \psi'_{a,b}\left(net_j^p\right)/a_j \tag{7}$$

$$\delta_{a_j} = \sum_{p=1}^{P}\left(d^p - y^p\right) y^p\left(1-y^p\right) v_j \times \psi'_{a,b}\left(net_j^p\right)\left(\left(net_j^p - b_j\right)/a_j^2\right) \tag{8}$$

$$\delta_{\theta_j} = \sum_{p=1}^{P}\left(d^p - y^p\right) y^p\left(1-y^p\right) v_j\ \psi_{a,b}\left(net_j^p\right) \times \left(net_j^p\right)^2/\theta_j^3 \tag{9}$$

$$\psi_{a,b}\left(net_j\right) = \psi\left(\left(net_j - b_j\right)/a_j\right) \tag{10}$$

$$net_j = \sum_{k=0}^{L} w_{jk}\, x_k \tag{11}$$

where d^p is the target output of the pth input pattern. The parameters of the network are updated as

$$w_{jk}(t+1) = w_{jk}(t) - \eta\,\delta_{w_{jk}} \tag{12}$$

$$v_j(t+1) = v_j(t) - \eta\,\delta_{v_j} \tag{13}$$

$$a_j(t+1) = a_j(t) - \eta\,\delta_{a_j} \tag{14}$$

$$b_j(t+1) = b_j(t) - \eta\,\delta_{b_j} \tag{15}$$

$$\theta_j(t+1) = \theta_j(t) - \eta\,\delta_{\theta_j} \tag{16}$$

III. Feature Extraction

The collected data of different faulted and normal conditions of the drive system are decomposed up to the second level of resolution of the wavelet packet transform (WPT) using the selected mother wavelet '*db*3' [21]. The digital data are acquired through the three-channel A/D converters of the ds1102 DSP board. Figures 2(*a*), 2(*b*), and 2(*c*) show normal current, inverter single phasing current, and shot through fault current of an IPM drive system, respectively.

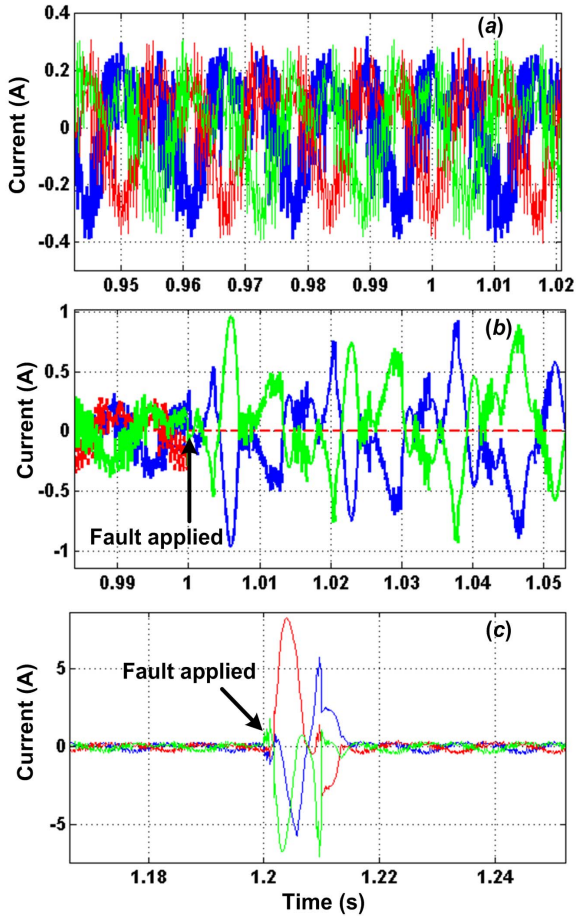


Fig. 2. Responses of the IPM motor drive system: (*a*) normal currents, (*b*) inverter single phasing currents, and (*c*) inverter shot through fault currents.

The second level WPT coefficients of normal and fault currents of a vector controlled IPM motor drive system are shown in Figs. 3(*a*)–3(*d*). The high frequency WPT coefficients of motor current for the case of faulted conditions of Figs. 3(*c*) and 3(*d*) are larger than those of normal (unfaulted) conditions of Figs. 3(*a*) and 3(*b*) after the occurrence of fault. These WPT coefficients are preprocessed by defining a feature vector F in order to get a finite input data vector of feature coefficients for convenient training and validation of the wavelet network. The feature vector F is defined as

$$F = \left[W_{aa^2} \; W_{ad^2} \; W_{da^2} \; W_{dd^2} \right] \tag{17}$$

where,

$$W_{aa^2} = \sqrt{\sum_{n=1}^{N} aa^2(n)/N} \tag{18}$$

$$W_{ad^2} = \sqrt{\sum_{n=1}^{N} ad^2(n)/N} \tag{19}$$

$$W_{da^2} = \sqrt{\sum_{n=1}^{N} da^2(n)/N} \tag{20}$$

$$W_{dd^2} = \sqrt{\sum_{n=1}^{N} dd^2(n)/N} \tag{21}$$

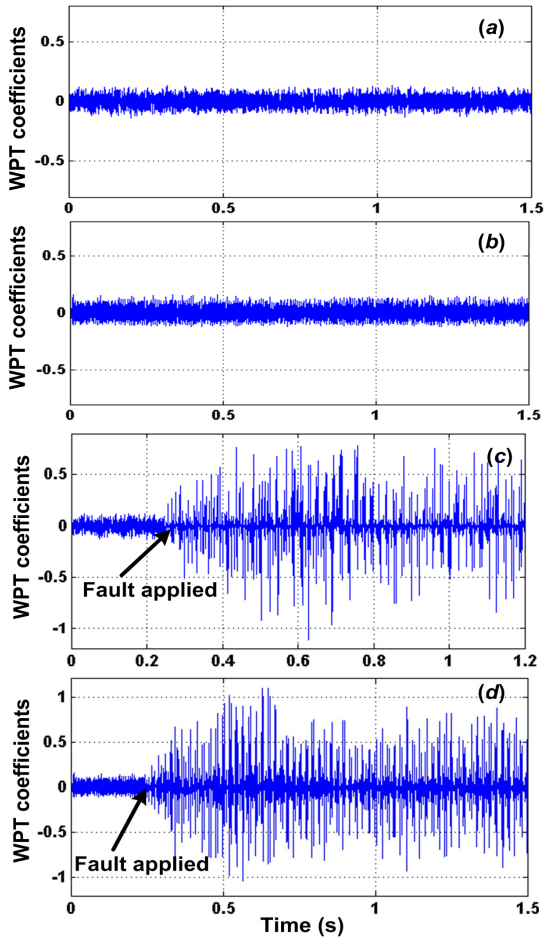


Fig. 3. Second level WPT coefficients of motor current: (*a*) high frequency approximations (da^2) of normal current, (*b*) high frequency details (dd^2) of normal current, (*c*) high frequency approximations (da^2) of fault current, and (*d*) high frequency details (dd^2) of fault current.

N is the total number of coefficients in a certain node of the wavelet packet tree, aa^2 are second level low frequency approximations, ad^2 are low frequency details, da^2 are high frequency approximations, and dd^2 are high frequency details of motor currents. Table-I shows the comparisons of feature vector between faulted and normal conditions using equations (17)–(21). These feature vectors are used to train the wavelet network of the proposed faults diagnostic and protection technique.

TABLE-I
FEATURE VECTORS

Type of faults	$W_{aa}{}^2$	$W_{ad}{}^2$	$W_{da}{}^2$	$W_{dd}{}^2$
Single phasing (phase-*a*)	9.03	3.06	3.07	3.14
Single phasing (phase-*b*)	295.10	17.39	11.13	12.23
Single phasing (phase-*c*)	70.76	4.43	3.36	3.72
Shot through (phase-*a*)	2877.8	101.7	46.8	25.1
Shot through (phase-*b*)	5343.56	111.3	117.4	97.12
Shot through (phase-*c*)	806.36	38.16	23.23	36.27

IV. Proposed Hybrid Diagnostic and Protection Technique

One of the difficult tasks in applying the wavelet network for diagnostic and protection of the motor drive system is to formulate the problem. Finding inputs and outputs is the first step of the problem formulation. In the proposed diagnostic and protection technique, the inputs are feature vectors of second level WPT coefficients of faulted and normal currents. The outputs are binary values of 0 or 1 to indicate whether the measured current is a normal current or a fault current, respectively. A three layer WNN with four inputs and one output is used in the proposed technique. The schematic of the proposed hybrid WPT and WNN based diagnostic and protection technique is shown in Fig. 4. The hybrid WPT and WNN based diagnostic algorithm is implemented through the dSPACE ds1102 controller board. The main processor in this board is the Texas Instrument TMS320C31. The controller board is supplemented by a set of on-board peripherals such as analog to digital (A/D) converter, digital to analog (D/A) converter, and incremental encoder interfaces. The board also includes a DSP microcontroller based digital I/O subsystem with a fixed point Texas Instrument processor TMS320P14. The complete ds1102 board is installed in a PC-AT with a capability of uninterrupted communication with the PC through a dual port memory provision. The PC monitor is used to edit and download the proposed self adjusting WNN based diagnostic algorithm into the DSP controller board for the protection of the IPM motor drive.

The specific procedure to implement the proposed hybrid WPT and WNN based diagnostic and protection algorithm for the IPM motor drive system using the ds1102 DSP board is shown in the flow chart of Fig. 5. In the proposed technique, samples of motor currents are squared and summed into one sample at the beginning for minimizing the computational burden. The hybrid algorithm checks the values of the network output using the trained weights and biases, and determines whether it is greater than the threshold or not in order to generate the tripping action.

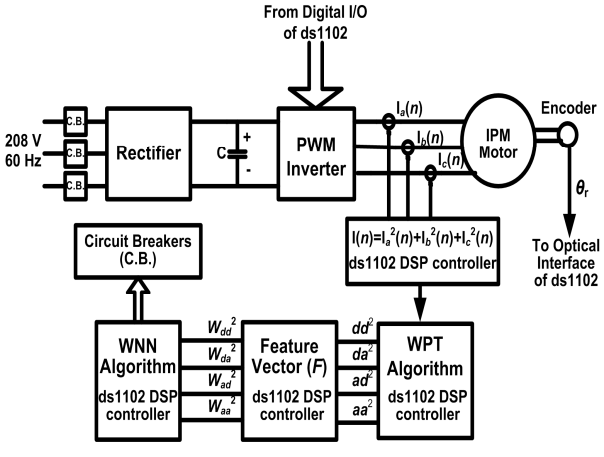


Fig. 4. Hardware schematic of the proposed hybrid WPT and WNN based diagnostic and protection technique for IPM motor drive system using the ds1102 DSP board.

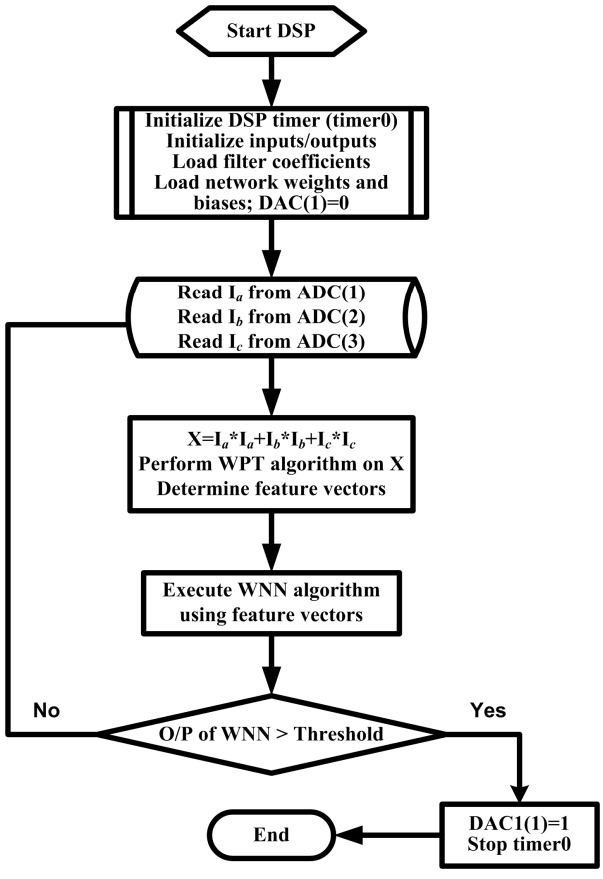


Fig. 5. Flow chart for the real-time implementation of the proposed hybrid WPT and WNN based diagnostic and protection algorithm for the IPM motor drive system.

Three types of faults are studied in this work to test the hybrid WPT and WNN based diagnostic algorithm. These include inverter single phasing, motor terminal line to ground (L-G) short circuit fault, and inverter shot through fault. Faults are initiated by connecting two points involved through a resistance. However, faults resistance are unknown in real-time. So, the scheme is also verified with different fault levels. The test for inverter single phasing is carried out by opening of a single pole single through switch connected between inverter output and motor terminal for few seconds.

V. Implementation and Results

The real-time implementation of the proposed diagnostic and protection technique involved the development of an experimental setup that includes both hardware and software components. The hardware includes the dSPACE DSP controller board model ds1102 with the 32-bit floating point processor TMS320C31. The software loads the values of filter coefficients of the selected mother wavelet '*db*3' and the values of wavelet network parameters for extracting features of fault currents using WPT algorithm and for processing of the tripping decision using WNN algorithm, respectively. The proposed hybrid algorithm is written in the turbo C language. It used a set of initialization and input/output (I/O) functions in order to initialize the TMS320C31's on-chip timers and to access the ds1102's on board A/D and D/A converters. When a timer is started, the A/D converter of the DSP board continuously samples motor

currents at the rate of 4 kHz. The samples of motor currents are sent to the memory of the DSP by the host PC, where they are squared and summed into one sample. This sample is placed into a circular buffer of size six. The six current data are processed using the filter coefficients of the mother wavelet '*db*3', and the weights of the WNN algorithm. The proposed hybrid diagnostic and protection algorithm is tested on-line on a laboratory three-phase, Y-connected, 1-hp IPM motor fed from sinusoidal pulse width modulated voltage source inverter (PWM-VSI). The digital I/O channels of the DSP board provides the firing pulses for the inverter switches using the hysteresis current controller. The proposed protection technique is tested on-line using the experimental setup of Fig. 4. Some sample results of the inverter-fed IPM motor are presented in Figs. 6–7. As can be seen from Figs. 6(*a*)–6(*b*) that the algorithm generated the trip signal within two cycles of the fault occurrence for the case of line to ground fault. This delay is due to the fact that the response time includes the executions of the proposed diagnostic and protection algorithm, the speed control algorithm, and the vector control algorithm for generation of logic signals. Figure 7(*a*) shows the phase-*a* current and the experimental response of no trip signal of the hybrid WPT and WNN based diagnostic algorithm for step increase and step decrease of command speeds in the IPM motor drive system. Figure 7(*b*) shows the phase-*a* current and the experimental response of no trip signal of the hybrid diagnostic algorithm for the sudden change of load torque in the IPM motor drive system. The hybrid algorithm identified these unfaulted conditions of Figs. 7(*a*)–7(*b*) as normal conditions and did not change the status of the trip signal. Thus the proposed hybrid WPT and WNN based diagnostic and protection algorithm correctly and promptly detected the faulted and normal currents of the inverter fed IPM motor.

VI. Conclusions

In this work, a new hybrid wavelet based diagnostic algorithm is developed and implemented in real-time using a DSP board for the protection of inverter faults in the interior permanent magnet (IPM) motor drive system. Sample faults such as inverter shoot through fault and single phasing are tested successfully using this novel hybrid wavelet packet transform (WPT) and wavelet neural network (WNN) based diagnostic and protection technique. The WPT feature coefficients of motor currents are used as inputs to a three-layer WNN. The test results confirm that the proposed technique is able to discriminate quickly between fault and normal currents with high accuracy. The proposed algorithm identified every fault properly and initiated the trip signal within two cycles of the fault occurrence. It is also to be noted that the algorithm did not initiate any trip signal for the presence of harmonics in motor currents during the normal condition and for the sudden change of load torque in the drive system. The proposed technique is quite fast and relatively easy to implement. It also requires less computational memory for on-line implementation.

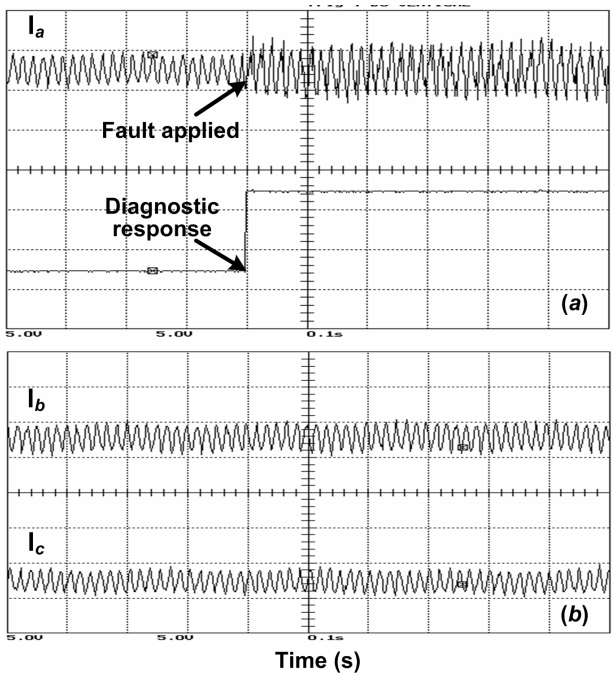


Fig. 6. Experimental responses of the IPM motor drive system for the case of line to ground fault using the proposed hybrid WPT and WNN based diagnostic algorithm: (*a*) phase-*a* motor current and algorithm response, (*b*) phase-*b* and phase-*c* motor currents. (time: 0.1 s/div., trip signal: 5 V/div., I_a: 4.172 A/div., I_b: 4.66 A/div., and I_c: 4.82 A/div.)

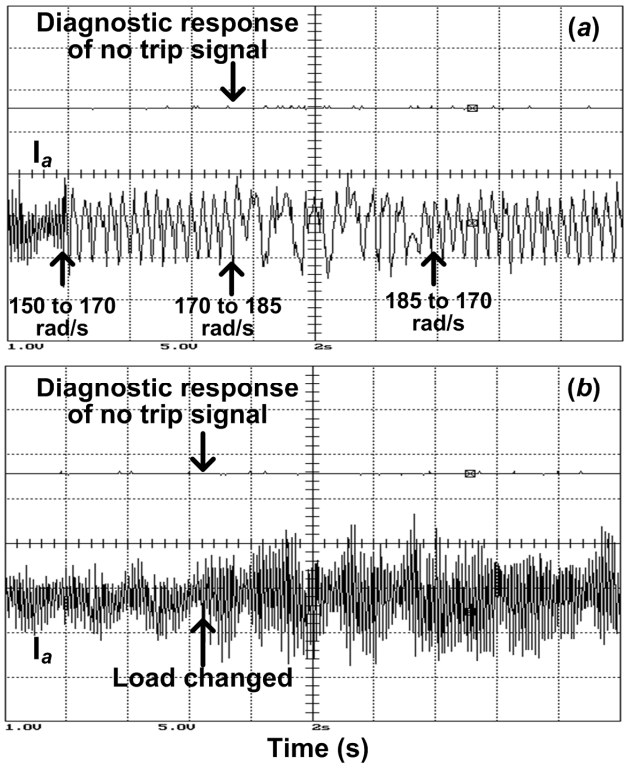


Fig. 7. Experimental responses of the IPM motor drive system using the proposed hybrid WPT and WNN based diagnostic algorithm: (*a*) step changes of command speed (150 to 170 rad/s, 170 to 185 rad/s, and 185 to 170 rad/s) and (*b*) change of load torque. (time: 2 s/div., trip signal: 1 V/div., I_a: 4.172 A/div., I_b: 4.66 A/div., and I_c: 4.82 A/div.)

References

[1] M. A. S. K. Khan, T. S. Radwan, and M. A. Rahman, "Real-time implementation of wavelet packet transform based diagnosis and protection of three-phase induction motors," *IEEE Transactions on Energy Conversion*, vol. 22, no. 3, Sept. 2007, pp. 647-655.

[2] M. A. S. K. Khan, T. S. Radwan, and M. A. Rahman, "Wavelet based diagnosis and protection of three-phase induction motor fed from sinusoidal PWM voltage source inverter," in *Proc. IEE International Power Electronics, Machines and Drives (PEMD) Conference,* Dublin, Ireland, 4-6 Apr. 2006, pp. 226-230.

[3] L. Wu, Z. Wu, W. Jin, and J. Ying, "A novel magnetization failure detection method for 1-phase BLDC motor based on back-emf test," in *Proc. IEEE International Symposium on Industrial Electronics (ISIE)*, Montreal, Canada, 9-13 July 2006, vol. 3, pp. 2200-2204.

[4] O. Moseler and R. Isermann, "Application of model-based fault detection to a brushless DC motor," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, Oct. 2000, pp. 1015–1020.

[5] X. –Q. Liu, H. –Y. Zhang, J. Liu, and J. Yang, "Fault detection and diagnosis of permanent magnet DC motor based on parameter estimation and neural network," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, Oct. 2000, pp. 1021-1030.

[6] M. A. Awadallah, M. M. Morcos, S. Gopalakrishnan, and T. W. Nehl, "A neuro-fuzzy approach to automatic diagnosis and location of stator inter-turn faults in CSI-fed PM brushless DC motors," *IEEE Transactions on Energy Conversion,* vol. 20, no. 2, June 2005, pp. 253-259.

[7] M. Dai, A. Keyhani, and T. Sebastian, "Fault analysis of a PM brushless DC motor using finite element method," *IEEE Transactions on Energy Conversion*, vol. 20, no. 1, March 2005, pp. 1-6.

[8] B. A. Welchko, T. M. Jahns, and S. Hiti, "IPM synchronous machine drive response to a single-phase open circuit fault," *IEEE Transactions on Power Electronics*, vol. 17, no. 5, Sept. 2002, pp. 764-771.

[9] B. A. Welchko, J. Wai, T. M. Jahns, and T. A. Lipo, "Magnet flux nulling control of interior PM machine drives for improved steady-state response to short circuit faults," *IEEE Transactions on Industry Applications*, vol. 42, no. 1, Jan.-Feb. 2006, pp. 113-120.

[10] M. Rosu, J. Saitz, and A. Arkiko, "Hysteresis model for finite element analysis of permanent magnet demagnetization in a large synchronous motor under a fault condition," *IEEE Transactions on Magnetics*, vol. 41, no. 6, June 2005, pp. 2118-2123.

[11] W. G. Zanardelli, E. G. Strangas, and S. Aviyente, "Identification of intermittent electrical and mechanical faults in permanent magnet AC drives based on time-frequency analysis," *IEEE Transactions on Industry Applications*, vol. 43, no. 4, July-Aug. 2007, pp. 971–980.

[12] M. A. S. K. Khan, T. S. Radwan, and M. A. Rahman, "Diagnosis and protection of IPM motors using wavelet packet transform," in *Proc. Conf. Rec. IEEE Industry Applications Society Annual Meeting*, Tampa, FL, Oct. 8-12, 2006, pp. 1970-1977.

[13] O. Wallmark, L. Harnefors, and O. Carlson, "Control algorithms for a fault-tolerant PMSM drive," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 4, Aug. 2007, pp. 1973-1980.

[14] L. Parsa and H. A. Toliyat, "Fault tolerant interior permanent magnet machines for hybrid electric vehicle applications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, July 2007, pp. 1546-1552.

[15] S. Bolognani, M. Zordan, and M. Zigliotto, "Experimental fault-tolerant control of a PMSM drive," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, Oct. 2000, pp. 1134-1141.

[16] W. L. Roux, R. G. Harley, and T. G. Habetler, "Detecting faults in rotors of PM drives," *IEEE Industry Applications Magazine*, vol. 14, no. 2, Mar.-Apr. 2008, pp. 23-31.

[17] C. K. Chui, *Wavelets: Theory, Algorithms, and Application,* California, US: Academic Press, 1994.

[18] Q. –J. Guo, H. –B. Yu, and A. –D. Xu, "Modified morlet wavelet neural networks for fault detection," in *Proc. IEEE International Conference on Control and Automation*, Budapest, Hungary, 27-29 June 2005, pp. 1209-1214.

[19] H. Haykin, *Neural Networks: A Comprehensive Foundation*, NJ, USA: Wiley-IEEE Press, 1994.

[20] Mathworks, *Matlab: Neural Network Tool Box*, 2004, Version 7.0.1.

[21] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, Sept. 1990, pp. 961-1005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Three Phase PWM Cûk AC-AC Converter Employing Minimum Switches

Md. Raju Ahmed[1], and M. J. Alam[2]

[1]Department of Electrical & Electronic Engineering, Dhaka University of Engineering & Technology, Gazipur, Bangladesh
[2]Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology, Dhaka, Bangladesh
E-mail: raju97eee@yahoo.com

Abstract – voltage sags and extended under voltages have the largest negative impact on industrial productivity and could be the most important type of power quality variation for many industrial and commercial customers. This paper proposes a 3-phase PWM ac-ac Cûk converter with minimum switches to maintain constant output voltage irrespective of the variation of input voltage and load. By PWM duty ratio control the ac-ac converter become a "solid state transformer" with a continuously variable turns ratio. The proposed ac-ac converter employ only two switches compared to the existing circuits that use six switches or more. For minimum number of switches the proposed circuit reduces cost, improves reliability. The operating principle, control method and simulation results of the proposed converter circuit are presented in this paper.

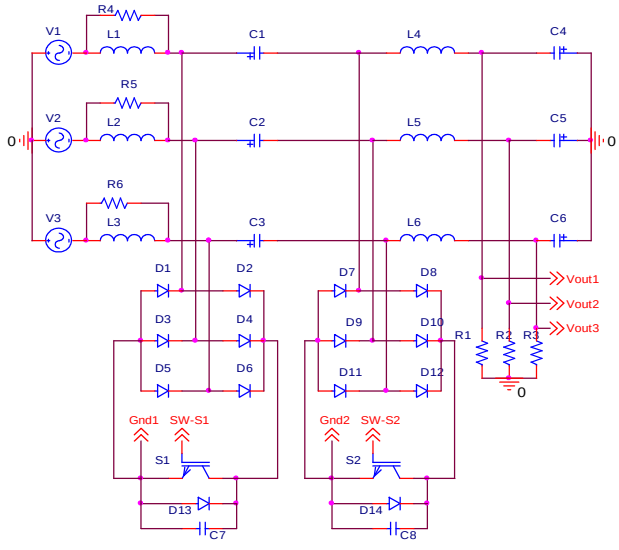
I. Introduction

For three phase ac-ac power conversion that requires variable output voltage and variable frequency, voltage-source inverter with a dc link are generally use. Direct PWM ac-ac converter is used where only voltage regulation is needed due to its smaller size and lower cost. Traditional ac-ac converters employ thyristors to regulate the ac power, such converters, however have slow response speed and need large input output filters to reduce low order harmonics [1]. In [2] a PWM ac-ac converter was reported with a comparison to the older generation thyristor based phase controlled converters. In [2]-[8], each paper proposed a different ac-ac converter and some simulation results were presented to illustrate their performance. A family of ac-ac converter with six switches was proposed in [9].

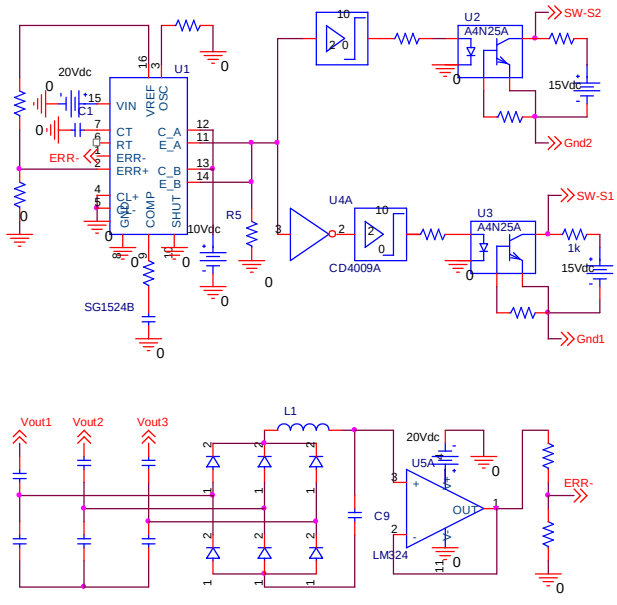
In general the use of fewer switches can reduce cost and improve reliability. In this paper we proposed a three phase PWM ac-ac converter that employs only two switches.

II. Proposed AC-AC Cúk Converter

The Proposed three phase ac-ac converter is shown in Fig. 1. There are mainly two parts in the proposed converter, power circuit and control circuit. The switches S_1 and S_2 of the power circuit are gated on and off in complement in order to provide the transfer of energy. Similar to the traditional Cûk dc-dc converter the operation of the proposed Cûk converter can be describe by two states. Fig. 2 and 3 show the equivalent circuit of power circuit for state I and II.



(a)



(b)

Fig. 1 Three phase Cûk ac-ac converter (a) power circuit (b) control circuit.

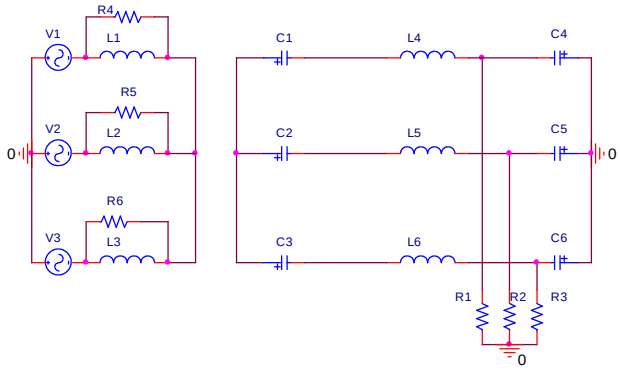


Fig. 2 State I: Switch S_1 is on and S_2 is off.

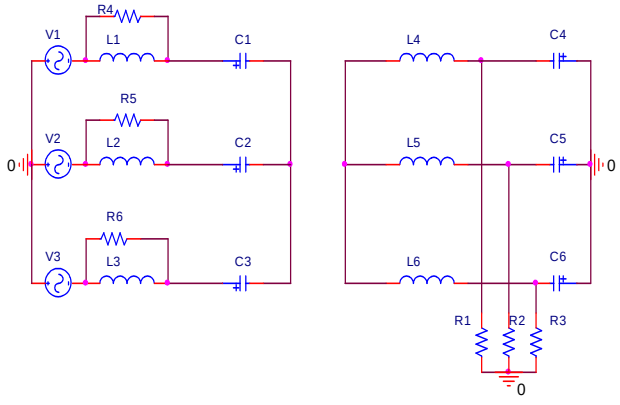


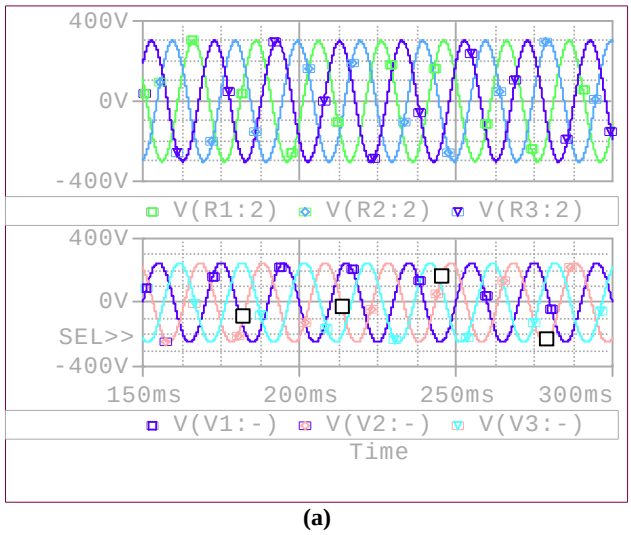
Fig. 3 State II: Switch S_1 is off and S_2 is on.

In state I switch S_1 is turned on and switch S_2 is turned off. The ac source charges the line inductors L_1, L_2 and L_3 shorted by S_1, energy is transfer from source to the inductors L_1, L_2 and L_3 while capacitors C_1, C_2 and C_3 discharge and transfer energy to L_4, L_5 and L_6. In state II, S_1 is turned off and S_2 is turned on. The energy stored in L_1, L_2 and L_3 is released and transfer to capacitors C_1, C_2 and C_3 through switch S_2, whereas energy stored in L_4, L_5 and L_6 is released and transfer to the load. The output voltage is controlled by varying the duty ratio. In this converter, R-C snubber is used for suppressing surge voltage across the switches.

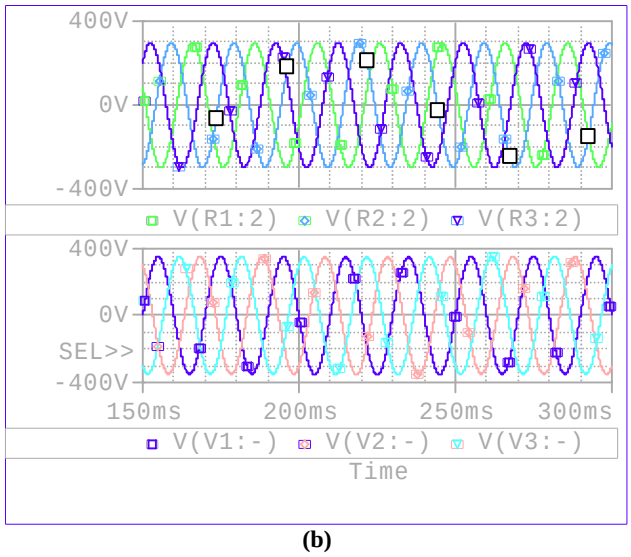
The control circuit of the proposed regulator is shown in Fig. 1(b). For generation of switching voltage, a compact and commercially available IC chip SG1524B is used. The same chip is used to regulate the output voltage. The positive input of the error amplifier of the IC is taken from the reference voltage of the IC after voltage dividing which is fixed. A fraction of the output voltage after capacitor voltage dividing and rectifying and passed through an OPAMP buffer is taken as the negative input of the error amplifier. Buffer is used to remove the loading effect. If the output voltage is equal to the desired voltage the negative input of the error amplifier (ERR-) is equal to its positive input and duty cycle is 0.5 and output remain same. If any change occurs in the output voltage due to variation of input voltage or load, increase or decrease in negative input of the error amplifier will change the duty cycle and hence the output voltage. As a result the regulator will maintain a constant voltage across the load during any change in supply voltage as well as for variation of load. For getting two switching signals with ground isolation two optocouplers are.

III. Simulation Results of Proposed Cûk AC-AC Converter

Simulation results of proposed converter are shown in Fig. 4. All voltage and current in this paper are in peak value. Fig. 4(a) shows the input output voltage waveforms when input source voltage is 250 V and output voltage is 300 V. Fig. 4(b) shows the input output voltage waveforms when input source voltage is 350 V and output voltage is 300 V. By PWM duty ratio control the proposed circuit maintain constant output voltage 300V irrespective of the variation of input voltage and load. In the simulation, the switching frequency is 4 KHz. Because of high frequency switching and filtering effect, input output voltage and current waveforms are sinusoidal



(a)



(b)

Fig. 4 Simulation results of Cûk ac-ac converter (a) V(V1:-), V(V2-) and V(V3:-) are three phase AC sources and V(R1:2), V(R2:2) and V(R3:2) are three phase load voltage (a) input voltage 250 V output voltage 300 V (b) input voltage 350 V output voltage 300 V.

IV. Conclusion

A three phase Cûk ac-ac converter was presented in this paper. The ac-ac converter can be used for ac-ac line conditioning to overcome voltage sags, surges, and load fluctuation. The proposed converter employ only two switches, it can reduce the cost and improve reliability. The control circuit of the proposed converter is very simple. The input output voltage and current of the proposed converter are sinusoidal.

References

[1] D. Vincenti, J. Hua, and P. Ziogas, "Design and implementation of a 25- kVA three-phase PWM ac-ac line conditioner ," IEEE Trans. Power Electron., vol. 9, PP.384-389, July 1994.

[2] P. N. Enjeti and S. Choi, "An approach to realize higher power PWM a-ac controller," in Proc. Record, IEEE APEC Conf., PP. 323-327 San Diego, CA, 1993,.

[2] A. K.S. Bhat and J. Vithayathil, "A simple multiple pulse wide modulated AC-AC chopper," IEEE Trans. Ind. Electron., vol. IE-29, PP.185-189, Aug. 1982.

[3] S. Srinivasan and G. Venkataramanan, "Design of a versatile three-phase ac-ac line converter," in Proc. Record, IEEE IAS Conf., pp. 2492-2499, Lake Buena Vista, FL, 1995.

[4] D. Vincenti, J. Hua, and P. Ziogas, "Design and implementation of a 25- kVA three-phase PWM ac-ac line conditioner ," IEEE Trans. Power Electron., vol. 9, PP.384-389, July 1994.

[5] S.M. Hietpas and R. Pecan, "Simulation of a Three-phase Boost Converter to Compensate for Voltage Sags," in Proc. IEEE 1998 Rural Electric Power Conf, , pp. B4-1-B4-7, April 1998.

[6] M. Kazerani, "A direct ac-ac converter based on current-source converter modules," in Proc. Record, IEEE PESC Conf., pp. 1115-1121, Vancouver, BC, Canada, 2001.

[7] S. M. Hietpas and M. Naden, "Automatic Voltage Regulator Using an AC Voltage- Voltage Converter," IEEE Trans. Ind. Applications, Vol. 36, No. 1. pp.33-38. January/ February 2000.

.[8] P. D. Ziogas, D. Vincenti, and G. Joos, "Practical PWN ac-ac controller topology," in Proc. Record, IEEE IAS Conf., pp. 880-887, Houston, TX, 1992.

[9] S. Srinivasan and G. Venkataramanan, "Comparative evaluation of PWM ac-ac converters," in Proc. Record, IEEE PESC Conf. Rec., pp. 529-535, Atlanta, GA, 1995.

[10] F. Z. Peng, L. Chen and F. Zhang, "Simple topologies of PWM ac-ac converters" IEEE Power Electron. Letters, Vol. 1, No. 1, pp. 10-13, March 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Eight Switch Buck Boost Regulator Topology for High Efficiency in DC Voltage Regulation

Tanwir Zubayer Islam and A. B. M. H. Rashid

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

Email: zubayer@hotmail.com, abmhrashid@eee.buet.ac.bd

Abstract - Improvements in the efficiency and size of DC-DC converters have resulted from advances in components, primarily semiconductors, and improved topologies. One topology, which has shown very high potential in limited applications, is the partial power processing boost technique, wherein a small DC-DC converter output is connected in series with the input bus to provide an output voltage equal to or greater than the input voltage. Since the DC-DC converter switches only a fraction of the power throughput, the overall system efficiency is very high. But this technique is limited to applications where the output is always greater than the input. The Eight Switch Buck Boost Regulator (ESBBR) concept extends partial power processing boost technique to operate when the desired output voltage is higher or lower than the input voltage, and the implementation described can even operate as a conventional buck converter to operate at very low output to input voltage ratios. This paper describes the operation and performance of an ESBBR configured as a bus voltage regulator providing ±50% voltage regulation range, bus switching, and overload limiting, operating above 98% efficiency. The technique does not provide input-output isolation.

I. Introduction

DC-DC converter power loss and size have decreased rapidly due to the improvements in semiconductors [1-4], and to a lesser extent due to improved passive components [1-6], topologies [4, 5], and soft switching schemes [4-6]. This is even true for linear regulators, where higher temperature components have reduced size and weight, and lower dropout voltage semiconductors and drive techniques have reduced the power loss [3, 5]. But these gains are evolutionary, and large step improvements are not likely.

The change from linear to switching regulators was revolutionary, and the power loss and therefore size of regulators made a large step decrease [1-6]. Since then the single most significant parameter determining converter efficiency and size, for a given frequency, has been the kva rating. The size of components to switch, transform, and filter is relatively independent of the topology, as the same amount of power is switched, rectified, and stored in filter components. The technique of partial power processing, wherein only a small fraction of the total output power is required to buck or boost the input to the desired output voltage, can significantly reduce the converter size and power loss.

The boost mode of the partial power processing technique has been used previously [7], wherein the low voltage output of a small DC-DC converter is connected in series with the input voltage. The output voltage can then be adjusted from essentially equal to the input voltage up to the input voltage plus the maximum output voltage of the DC-DC converter. For example, a regulator for a 100 Volt 1 Kw bus could be constructed with a DC-DC converter having a 100 volt nominal input, and a 0 to 10 Volt output and a 100 Watt, 10 Amp, rating. The configuration would allow regulating the output to up to 10% higher than the input, and yet only use switching and filtering components sized for 10% of the total rating, and it therefore offers greatly reduced size and power loss. The technique is very useful in situations where the input/output voltage ratio is relatively small, the output is always greater than the input, and isolation between the input and output is not required.

The eight switch buck boost technique expands the array of applications considerably by also allowing the output to be lower than the input. This allows essentially twice the regulation range for a given size converter. Additionally, a circuit enhancement allows operation of the same converter in the conventional buck mode, so that the output can be regulated down to zero volts, allowing operation as a current limiting voltage regulation remote power controller.

The circuit concept and the performance of the conceptual circuit are described in this paper. Efficiency calculation, voltage regulation and stability simulation is also formulated in this regard.

II. Circuit Operation

The operation of the circuit will be described by separating the modes of operation and describing them

individually, and then showing how they are combined. The simplest mode is the boost mode, which will be described first.

A diagram of the power flow in the boost mode is shown in Fig. 1. In the boost mode, the output of a low voltage converter is added in series with the input voltage to provide an output voltage equal to or greater than the input. Thus, power from the converter is added to the output to increase the voltage. The Eight Switch Buck Boost Regulator (ESBBR) implementation described uses a full bridge input stage and a center tapped transformer – rectifier output stage as shown in Fig. 2. When all the primary side switches are turned off there is no voltage across the transformer, and the input voltage is connected to the output through the input filter, the secondary of the transformer, the rectifier diodes, and output filter. When the switches of the input bridge are conducting, a voltage is impressed on the secondary of the transformer. The instantaneous output at the rectifiers is the input voltage plus the transformer secondary voltage, which is equal to the input voltage divided by the transformer turns ratio. By varying the PWM duty cycle from 0 to 100% the average output can be controlled between a minimum of the input voltage and a maximum of the input voltage plus the boost provided by the transformer's secondary output. The output filter smoothes the output and the average output voltage, ignoring conduction losses in the filters and rectifiers, will be:

$$Vout = Vin + Vin * PWM\ Duty\ Cycle/Turn\ Ratio \quad (1)$$

The turns ratio for the prototype design is 2:1, allowing the output to be boosted up to 150% of the input.

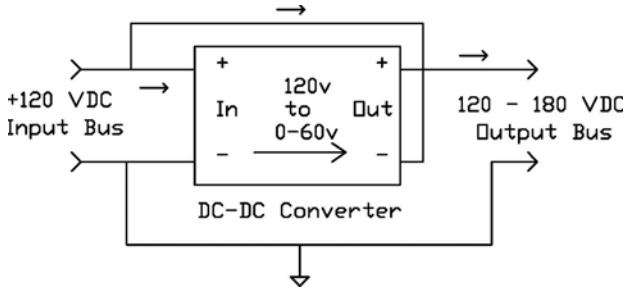


Fig. 1 Boost Mode Power Flow Block Diagram

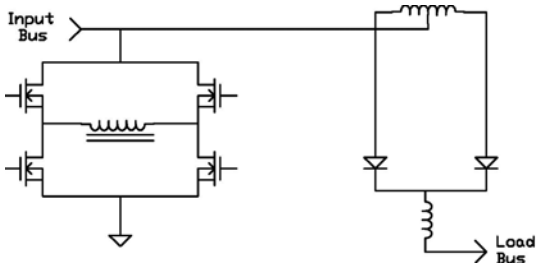


Fig. 2 Boost Mode Schematic Diagram

The buck mode of the ESBBR operates in a very different mode, but uses essentially the same components. The power flow is as shown in Fig. 3. The input voltage is greater than the desired output voltage. The concept of operation is that the input of a DC-DC converter is connected in series, with opposed polarity to the ESBBR input bus, and the output of the DC-DC converter is connected in parallel with the ESBBR input bus. The voltage drop across the input of the DC-DC converter reduces the ESBBR output voltage, and the power associated with this voltage drop is returned to the input bus by the DC-DC converter. The operation is similar to placing a battery in series with the input bus. Connected in one direction it increases (boosts) the output voltage. When it is connected "backwards" it decreases (bucks) the output voltage. In the boost case, the battery is discharged as would be expected, but in the buck case the battery is actually charged, and eventually overcharged. The use of a DC-DC converter allows a continuous process. A block diagram of the buck mode is shown in Fig 3.

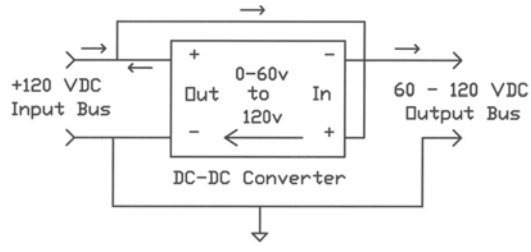


Fig. 3 Buck Mode Power Flow Diagram

It is important to note that the ESBBR input and output voltages are always positive and that the ESBBR input and output currents are always positive. However, in terms of the DC-DC converter within the ESBBR, the voltage on the input bus side is always positive and the current in the output bus side is always positive, but the current into the input bus side of the DC-DC converter itself is positive in the boost mode and negative in the buck mode, and the voltage at the output bus side of the dc-dc converter itself will be positive for the boost mode and negative for the buck mode.

If the transformer voltage was less than 1:1, a negative output voltage could be generated, and the net power flow in the ESBBR would be back to the input. Considering this would add confusion and contribute little to an understanding or appreciation of the capabilities of the ESBBR concept and won't be discussed further.

The ESBBR can operate in a mode where the power flow is from the output to the input if synchronous rectification is used at all points, and it makes some sense as a "regenerative" sort of application. This ability will be discussed later in the paper.

The schematic for buck mode operation is shown in Fig. 4. Many of the same components used for the boost converter are used for the buck converter, although in different roles. The center-tapped side of the transformer, which was the secondary in the boost mode, becomes the primary, and FET switches replace the rectifier diodes of the boost converter. (In the actual application the buck mode FETs are placed in series with the boost mode rectifiers.) The FETs which switched the primary of the transformer in the boost mode now function as a full wave bridge rectifier using the body diodes (or synchronous rectification with the FETs).

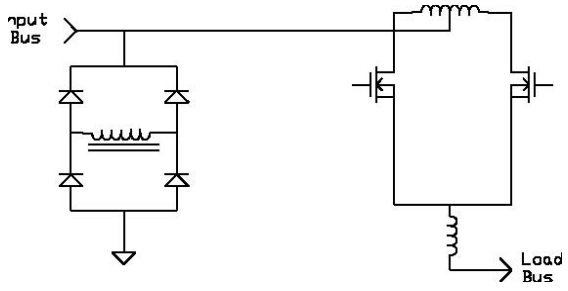


Fig. 4 Buck Mode Schematic Diagram

The output voltage of the DC-DC converter is re-circulated back to the input bus of the ESBBR. Varying the duty cycle of the switches controls the ratio between the input and output voltage of the converter. Since the input bus voltage fixes the output voltage of the DC-DC converter, the effect of varying the duty cycle is to vary the voltage drop between the input bus and the ESBBR output voltage.

The switching action of the two switches in the primary of the buck converter might also be considered unusual in that either one or both switches are always turned on, they are both never off simultaneously, even during the switching cycle. It is a current fed mode of operation. When both switches are conducting there is no voltage drop (except the small conduction losses) across the input of the DC-DC converter, so the output voltage of the ESBBR is equal to its input voltage. When one switch is open, the voltage across the transformer output will be equal to the ESBBR input bus voltage since the bridge rectifiers clamp it to that value. Therefore the input voltage of the converter will be equal to the input bus voltage divided by the transformer turns ratio. Varying the duty cycle controls the average voltage dropped across the DCDC converter, and therefore the ESBBR output voltage.

The ESBBR output voltage in the buck mode can be computed as:

$$\text{Vout} = \text{Vin} - \text{Vin} * \text{PWM Duty Cycle/Turn Ratio} \quad (2)$$

Where a PWM Angle of 0 corresponds to the switches being closed all the time and an angle of 100% corresponds to each switch being closed 50% of the time. Except for the sign in the equation, this equation is identical to the one for the boost mode. If the duty cycle is redefined such that the duty cycle for maximum buck is –100%, the equations become identical and for a 2:1 turns ratio transformer, the output can be varied from 50% to 150% of the input voltage.

The third mode of operation, the Current Limiting (CL) mode, uses many of the components used in the boost and buck modes, as well as some additional components. It is again a completely different mode of switching, and allows operation down to 0 output voltage to allow an increased operating range, and particularly an OFF mode and turn-on/overload limiting function. In the CL mode the primary or full wave bridge side of the converter switches serves no purpose, and they are all tuned off. The FET switches used for the buck mode on the output bus side of the converter are switched simultaneously, effectively as one switch and the transformer has no voltage across it. A diode is added between the FET switches and common, and the FET switches, the diode, and the output inductor function as a conventional buck converter as shown in Fig. 5. The transformer has no function in this mode, and is in fact shorted by an auxiliary FET to reduce voltage transients.

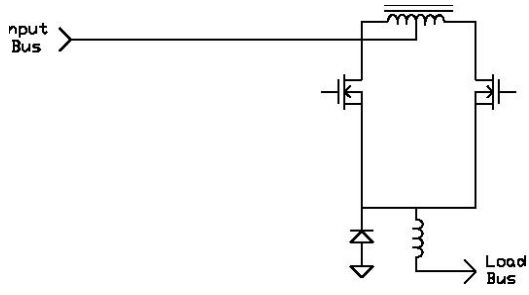


Fig. 5 Current Limiting Mode Schematic Diagram

In this paper this mode will be referred to as the CL mode to avoid confusion with the ESBBR buck mode previously described. In this mode the output voltage can be computed as:

$$\text{Vout} = \text{Vin} * \text{PWM Angle} \quad (3)$$

Although the form of this equation is quite different from the boost and buck mode equations given previously, the output voltage is still defined only by the input voltage and PWM angle and the output can be controlled between 0 volts and the input voltage. Actually the range in this mode overlaps completely the range in the ESBBR buck mode, but with lower efficiency and higher ripple currents in the filters. In the prototype the switching frequency is increased for operation in the CL mode to reduce the current ripple. The current limit mode is used only during turn on and overloads conditions, or if the output voltage must be lower than that that can be obtained with the ESBBR buck mode. Also, the equation is only valid for continuous conduction where the inductor current never goes to 0.

Essentially overlaying the boost, buck, and current limit mode schematics reveals the complete ESBBR schematic, as shown in Fig. 6.

This discussion has assumed that the input and output currents of the ESBBR are always positive. But if synchronous rectification is used on both sides of the

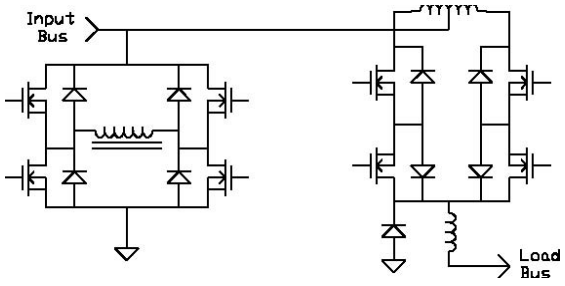


Fig. 6 ESBBR Schematic

DC-DC converter then current flow can be either direction. This is very useful because the magnetizing current of the transformer can be supplied from the input source, otherwise there would be a minimum output current, the current required to magnetize the transformer, below which the ESBBR buck mode would not operate. Switches that can conduct in both directions are already required on two of the switches on the ESBBR output. The logic to control the primary switches is straightforward, and provides the magnetizing current discussed above. By also using active switches in the other two positions of the output the efficiency will increase (synchronous rectification), and the ESBBR will operate in a regenerative mode wherein power is actually returned to the source from the load. The regenerative operation in the CL mode is possible if the free wheeling diode was replaced with a switch, but this has not been done in the prototype.

III. Efficiency

High efficiency was the primary driver for developing the ESBBR technique; therefore the breadboard circuit was optimized for low losses. As such the design uses rather large semiconductors and magnetic components, and a relatively low switching frequency, but it clearly indicates the potential. Figure 7 shows the efficiency as measured for a constant input voltage and load current as the buck or boost ratio is varied. Plots are shown for a 5-amp load, the design full load for the converter; for a 2-amp load where the efficiency peaks; and for a 0.5-amp load, where switching losses dominate. These plots show the high efficiency over a wide range of buck/boost ratios and load currents. The DC resistance of the series connected components (the input filter, the transformer secondary, the secondary switches, and the output filter) is 0.15 ohms, accounting for 0.25% loss at 2 amps, and 0.6% loss at 5 amps load. Two-thirds of the loss at full load is due to switching losses or resistive losses in the primary and magnetic components core losses.

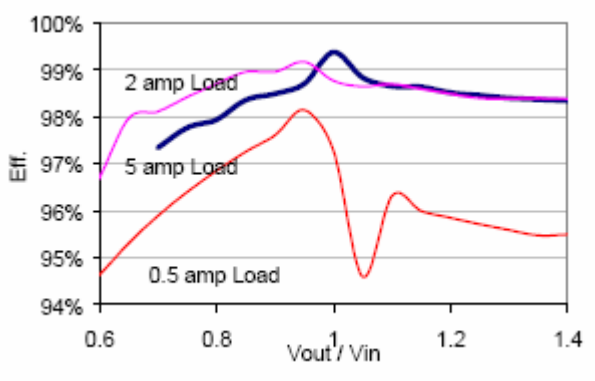


Fig. 7 ESBBR Efficiency

The current development of the ESBBR is for an application as a fuel cell regulator. The prototype ESBBR is being tested with a simulated fuel cell scaled at 50% of the voltage and 10% of the current rating of the intended fuel cell. The target fuel cell is rated at 10kw, 340 volts no load (170 volts for the simulator), 200 volts full load (100 volts for the simulator), at 50 amps (5 amps for the simulator). The intended load is a 270-volt (135 volts for the simulated system) bus. The input voltage variation and ESBBR efficiency is shown in Fig. 8. The efficiency doesn't fall badly until about 10% load. The power loss is a relatively constant 2 to 3 watts below 50% power. These efficiencies are for the power stage only, the control power requirement, including gate drive, of the prototype is 3 to 4 watts.

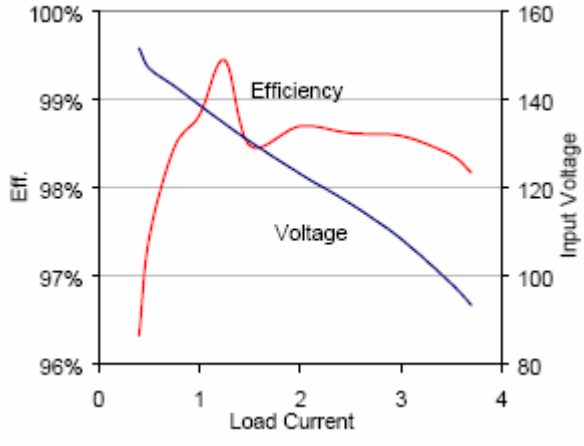


Fig. 8 Efficiency as a Fuel Cell Regulator

IV. Voltage Regulation and Stability

Voltage regulation in a converter is primarily a measure of the performance of the voltage regulation feedback loop. The regulation can be as good as the regulator, independent of the transfer characteristics of the converter. However, the preceding discussions discussed the transfer functions of this converter and the similarity of them throughout the three different modes, and implied that the output voltage could be determined based on the input voltage, the transformer ratio, and the PWM angle. Figure 9 shows the open loop regulation of the ESBBR as

a function of input voltage and output current. The open loop regulation generally continues into the current limit mode also, but not as accurately, and this region is not shown in the figure. A closed loop regulator is also included on the prototype ESBBR. Its only function is to trim out the remaining error shown in the figure, so its control range is limited to altering the output voltage a few percent, and it uses only the integral of the voltage error. Transient response, and damping, is provided by the open loop path, which simplifies the design of the integral controller path.

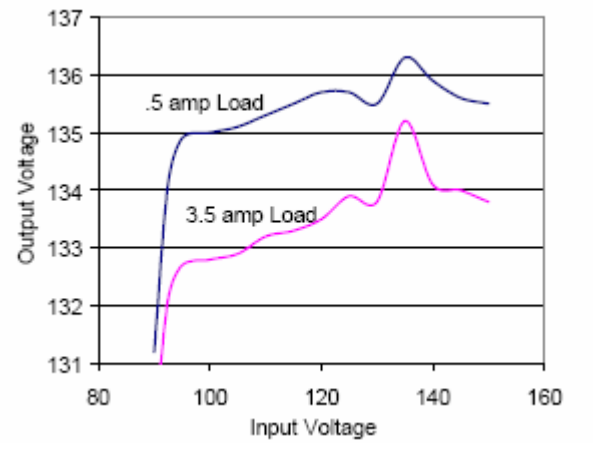


Fig. 9 Open loop Voltage Regulation

As discussed so far, the ESBBR is a stiff voltage source, and to be useful in a real system the ability to limit currents into an overload is required. Two circuits were added to accomplish this. The first an instantaneous over-current sense at twice rated current to shut off all switches and limit the peak current into the output filter.

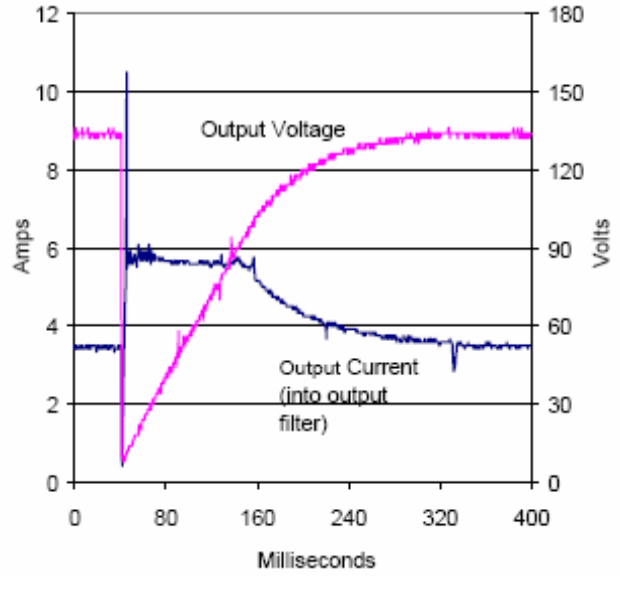


Fig. 10 Transient Overload Response

The second is a linear proportional pus integral current regulation loop at 1.5 times rated current to control the PWM angle until the output voltage recovers and the voltage regulator takes over. Implementation of these loops was successful as illustrated in Fig. 10, which shows the recovery from applying a large capacitor as a transient load while running at rated current. The capacitor is 100 times the output filter capacitance so the output voltage collapses almost completely. The instantaneous over-current sensing limits the initial spike of current (into the filter), and then the ESBBR runs under current limit until the capacitor is charged. The current sensing and the data plotted are the current into the output filter. This is the same as the currents that is in the output switches, and it is the critical current to be controlled for protection of the ESBBR

V. Summary

The ESBBR concept has been developed as a breadboard to demonstrate its capabilities as a combined bus regulator and remote power controller incorporating switching and current limiting functions. The simulated circuit operates at over 98% efficiency over a broad input-output voltage and load range.

VI. References

[1] Bryant, Brad, Kazimierczuk, Marian K., "Voltage Loop of Boost PWM DC–DC Converters With Peak Current-Mode Control ", IEEE Transactions On Circuits And Systems—I: Regular Papers, vol. 53, no. 1, January 2006 pp. 99-105

[2] Redl, R. and Sokal, N. O., "Current-Mode Control, Five Different Types, Used With the Three Basic Classes of Power Converters: Small-Signal AC and Large-Signal DC Characterization, Stability Requirements, and Implementation of Practical Circuits," IEEE Power Electronics Specialists Conference - 1985 Record, pp. 771-785.

[3] Erickson, R.W. and Maksimovic, D., *"Fundamentals of Power Electronics"*, 2nd Ed., Kluwer Academic Publishers, 2001.

[4] Bryant, Brad, Kazimierczuk, Marian K., "Open-Loop Power-Stage Transfer Functions Relevant to Current-Mode Control of Boost PWM Converter Operating in CCM", 2158 IEEE Trans. On Circuits And Systems—I: Regular Papers, vol. 52, no. 10, October 2005 pp. 2158-2164

[5] Brayant, B. and Kazimierczuk, M.K., "Small-Signal Duty Cycle to Inductor Current Transfer Function for Boost PWM DC-DC Converter in Continuous Conduction Mode," Proc. IEEE ISCAS, Vancouver, BC, Canada, 2004,pp.856-859.

[6] Bryant, Brad, Kazimierczuk, Marian K.," Modeling the Closed-Current Loop of PWM Boost DC–DC Converters Operating in CCM With Peak Current-Mode Control", IEEE Trans. On Circuits And Systems—I: Regular Papers, vol. 52, no. 11, November 2005 pp. 2404-2412

[7] Rodrigues, O. and Ghosh, P., "Series Interconnection of DC-DC Converters for Output Control", IEEE Trans. On Power Electronics, vol. 21, no. 22, pp 133-137, October, 2004

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Advances on IPM Technology for Hybrid Cars and Impact in Developing Countries

Dr. M. A. Rahman, *Life Fellow, IEB and IEEE*
Memorial University of Newfoundland
St. John's, Newfoundland, A1B 3X5 Canada

Abstract—The past thirty years have been an exciting period for tremendous advances in the development of interior permanent magnet (IPM) electrical machines. Over the course of this time, interior permanent magnet synchronous machines (IPMSM) have expanded their presence in the commercial marketplace from few specialized niche markets such as machine tool servo drives to mass-produced applications including high-efficiency electric traction drives for the latest generation of hybrid-electric vehicles (HEV). Power ratings of available IPM motor drives have dramatically expanded by approximately three orders of magnitude during this period, now reaching power levels up to 1 MW ratings. What are the factors that made such impressive progress possible? Closer examination reveals that several different knowledge-based technological advancements and market forces have combined, sometimes in fortuitous ways, to accelerate the development of the impressive IPMSM drives technology that we find available today. The purpose of this paper is to provide a broad explanation of the various factors that lead to our current state-of-the-art IPM technology. This highly efficient energy conversion technology has enormous impacts on the world electrical energy supply and demand utilizing conventional fossil fuel sources like oil, coal and gas. Examples will illustrate commercial successes of Toyota's hybrid electric vehicles like PRIUS, utilizing the latest developments in knowledge based highly efficient and smart automobiles now and in the very future.

I. INTRODUCTION

Electric power system forms the backbone of modern society. Electricity and its accessibility are the greatest engineering achievements in the past century. In the 21st century, global warming has become an important issue. Carbon dioxide (Co_2) gas emissions should be reduced to preserve the correct air quality as per Kyoto protocol, implemented on February 16, 2005 by most of the countries. Modern human beings, who need electric energy technologies for climate controlled home and work place environments via air conditioners and mass transportation using cars as necessities, cannot put up with the inconveniences of the past. In order to maintain and develop this energy consuming technologies, availability of sustainable energy sources and their effective uses through efficiency improvements are of paramount importance. Power electronics and electric motor drives are the enabling technologies crucial for industrial competitiveness in the world market place. One of the most valuable achievements in power electronics is to introduce degree of freedom to variable frequency from the fixed value of the generated ac power supplies. Over 60 % of the generated energy is consumed by electric motors. Variable ac speed drive, which regulates the speed of the motor by controlling the frequency, can significantly reduce the energy consumption, particularly in heavy-duty cycle fans, pumps, compressors and traction in hybrid electric vehicles. Thus improvements in efficiency of the electric motor drive systems are the most effective measures to reduce primary energy consumption; and thereby reduce Co_2 gas emissions, which cause global warming.

The objective of this invited paper is to provide a brief introduction to the recent emergence of high efficiency and high performance interior permanent magnet (IPM) synchronous motors. Highlights of IPM motor drives include wide spread application in Japanese hybrid electric vehicles, which are just one of many items of ac motor drive in passenger automobiles to save precious electric energy.

II. ANALYSIS

The principle of operation of any rotating electric motor is derived from Lorenz force. A current carrying conductor placed in a magnetic field is acted upon by a force by way of the *BLI* rule. For a conventional synchronous motor the stator is fed from 3-phase balanced voltage source and rotor field winding is supplied by dc excitation current through slip rings. The machine starts as an induction motor, and when it attains near synchronous speed, the dc excitation current is switched on. The rotor is snapped into synchronism and it runs at synchronous speed. The obvious disadvantage is that the motor needs two sources of power; one ac from the stator and dc through the rotor involving brushes and slip rings. Unlike in induction motor, the speed of a synchronous motor is constant irrespective of loads. But an induction machine is a singly fed motor. The rotor is squirrel cage, simple and robust. The disadvantage is that it cannot operate at synchronous speed N_s, and the rotor speed N_r decreases with load.

Thus an induction motor is an inherently inefficient motion control device, because the ideal efficiency is 1-S, where $S = (N_s\text{-}Nr)/N_s$. These disadvantages of both the conventional doubly fed synchronous motor and the singly fed induction motor can be overcome by means of a permanently excited singly fed IPM motor. An IPM is an induction start but synchronously run high efficiency motor. It is sometimes referred as induction-synchronous motor. It must overcome the magnet brake torque at line starting. However, there are many challenges to overcome. Some are given as follows:

- Create variation of d-q axis inductances without varying air gap.
- Vary and control of excitation of permanently excited rotor of IPM.
- Optimum variation of PM torque and reluctance torque for specific applications.
- Reduction of cost, weight and size of IPM motor.
- Intelligent converter and inverter for IPM drive.

The developed power P_d in a 2-pole 3-phase salient pole synchronous motor can be given as;

978-1-4244-2014-8/08/$25.00 (c)2008 IEEE

$$P_d = \frac{3V_pE_0}{X_d}\sin\delta + \frac{3V_p^2(X_d - X_q)}{2X_dX_q}\sin 2\delta \qquad (1)$$

Where V_p is terminal voltage/per phase, E_o is excitation voltage/per phase; X_d and X_q are d-q axis reactances per phase, respectively and δ is angle between V_p and E_o.

In conventional salient pole synchronous machines, the airgap length at the direct (d) axis is small and the airgap length at the quadrature (q) axis is large. Thus there exists physical variation of the airgap, which in turn causes reluctance changes of the motor as the rotor rotates.

The equation (1) can be rewritten as;

$$P_d = P_e \sin\delta + P_r \sin 2\delta \qquad (2)$$

Where, $P_e = [3V_pE_o]/ X_d$ and $P_r = [3V_p^2(X_d - X_q)]/ 2 X_d X_q$

P_e is the peak power component due to dc field excitation and P_r is the peak power component due to reluctance variation at the airgap. The latter is called the reluctance power. The contribution of each power component to the total power P_d is significant for the optimum design of a salient pole synchronous motor. For fixed parameter values it is obvious that the first term of Eqn (2) is maximum when δ is 90^0, and the second term of Eqn (2) is maximum for $\delta = 45^0$. The salient pole synchronous motor develops more stable power for a given excitation level, because the total developed peak power $P_{d\,peak}$ is greater than each of the P_e and P_r components individually.

The challenge for designers for an IPM motor is to create reluctance variation of the motor by keeping airgap length constant. This has been done by inserting permanent magnets in various arrangements and magnet polarity orientations below the conduction cage of the IPM rotor such that the machine reluctance variations are made possible but keeping the airgap length uniformly constant [1]. For some specific applications the squirrel/conduction cages can be dispensed with for new IPM rotors for air conditioners and hybrid electric vehicles.

The developed torque T_d is obtained by dividing Eqn (2) by angular synchronous speed. An IPM motor develops its driving torque due to both the permanent magnet excitation and reluctance variation.

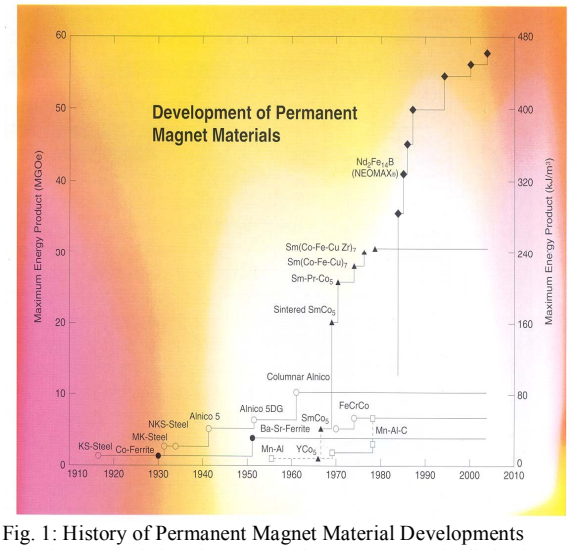


Fig. 1: History of Permanent Magnet Material Developments

The history of development of IPM motors is linked to the advancement of high-energy permanent magnet materials over the past 50 years. Fig.1 illustrates the brief history of the development of permanent magnet (PM) materials. In the 1950s the most promising material was the Alnico (Aluminum Nickel Cobalt) magnet with $(BH)_{max}$ at around 5 MG Oe. Next, Barium Ferrite magnets came by 1960s, and Samarium Cobalt magnets appeared in the 1970s with $(BH)_{max}$ at about 4 and 6 MG Oe., respectively. The latest quantum jump occurred in early 1980s,when Neodymium Boron Iron (NdBFe) magnets with $(BH)_{max.}$ at 14 MG Oe. became commercially available. Now a days NdBFe magnets with $(BH)_{max.}$ at 58 MG Oe. are routinely manufactured and marketed by the Japanese manufactures like Neomax Co, Japanmagnets Inc., Aichi Steels Co, TDK Co., etc .The critical properties of permanent magnets for IPM motors are very high coercive force H_c, high residual magnetic flux density B_r and highest $(BH)_{max.}$ energy product. All PM materials except NdBFe magnets are found not quite suitable for high efficiency IPM motor drives. Merrill introduced an earlier IPM motor using Alnico-5 in 1955[2]. Binn, Barnard, Jabbar presented a series of flux focused IPM motor using ferrite PM materials in 1978[3]. Rahman designed and built the first large 45 kW high efficiency IPM motor utilizing NdBFe magnets in 1982 [4,6]. Rahman, Little and Slemon provided analytical models for IPM in 1985[7-8]. Jahns incorporated the flux-weakening regime in 1987[11,13]. Sebastian and Slemon presented inverter driven IPM drives in 1987[15]. Fratta, Vagati and Villata provided design criteria of IPM for field weakening operation in 1990 [20]. Zhou and Rahman presented the finite element analysis of IPM motor incorporating field and circuit coupling in 1994[28]. Sustained and extensive research, development, analysis, control and application of IPM motors are progressing in leaps and bounds for the past two decades [12-47], perhaps even exceeding Merrill's dream [2] and Alger's expectation.

III. DESIGN REQUIREMENTS

The key requirements of IPM motors and generators for traction applications in hybrid electric vehicles are:

- Large torque and higher power density,
- High torque at low speeds for starting and uphill climb
- High power at high cruising speeds
- Maximum efficiency over wide speed and torque ranges
- Wide speed range with constant power mode, exceeding 2- 4 times the base speed
- Optimum compromise between motor peak torque and inverter volt-ampere ratings
- Short term overload capability, typically twice the rated torque over short duration
- Low cogging torque, low ripple and low acoustic noise
- Optimum stator winding design
- New rotor design with magnets orientation for maximum variation of d-q inductances
- Reduction of magnetic saturation due to cross-coupling Limits to open circuit voltage and total harmonic contents
- Low copper and iron losses at high speeds
- High reliability for all operating conditions
- Minimum weight and smallest size
- Low fuel consumption rate (litre/km),
- Clean and environmentally benign
- Quiet, smooth and comfortable ride
- Better battery power and self-charging
- Smart sensors and interfaces
- Least magnet flux leakage
- Magnet demagnetization withstand with respect to armature reaction
- Temperature and surface corrosion constraints of magnets
- Minimum gear and more direct drive
- Regenerative braking and short charging cycle
- No plug-in and hybrid transmission
- Plug-in in off peak periods
- Solar panel body and hybrid transmission
- Seamless transfer between engine and electric traction
- Minimum maintenance and high efficiency

- Lowest initial and operating cost

IV. MOTOR TORQUE

The developed torque T_d for an IPM motor can also be expressed for synchronously revolving d-q axis reference frame as [6];

$$T_d = \frac{3p}{2}\left[\lambda_m i_q + (L_q - L_d) i_d i_q\right] \qquad (3)$$

Where, λ_m is flux linkage due to permanent magnet excitation, L_d and L_q are d-q axis inductances, respectively; i_d and i_q are d-q axis currents, respectively and p is number of pole pairs. It is also to be noted that the torque equation (3) is quite non-linear, because λ_m , L_d, L_q, i_d and i_q are not usually constants. All these five quantities vary during dynamic operating conditions.

It is to be noted that the first term of equation (3) is identical to the separately excited dc motor. It is important for indirect vector control of an IPM motor. The second term is the reluctance torque. Efficient utilization of this reluctance torque component of equation (3) is most critical for intensive flux weakening operations and efficiency improvements in hybrid electric vehicles (HEV) and electric traction drives [45].

Finite element (FE) analysis is a requirement for fine-tuning the parameters determination of the IPM motors for optimum efficiency in high-speed operation using smart inverter and control systems. Figure 2 shows the finite element based d-q axis magnetic flux distribution due to flux focusing arrangements of rotor permanent magnets [26]. Design optimization of the IPM motor drive system can also be carried out by various methods.

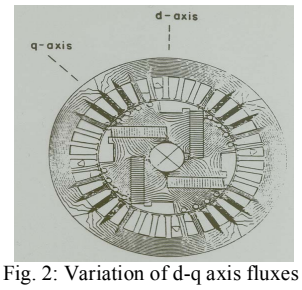


Fig. 2: Variation of d-q axis fluxes

V. STEADY STATE OPERATING MODES

The operation of a synchronous motor is conventionally explained by using the Thevenin's per phase equivalent circuit model. The applied phase voltage V_p and the excitation voltage E_o at the airgap due to dc field current in the rotor of the motor is connected by series reactance X, neglecting stator resistance drop. The phasor voltage triangle is governed by the Kirchoff's voltage law. For the sake of better insight of dc field current supplied in its rotor, the Thevenin's equivalent can be replaced by its dual Norton's equivalent circuit model.

Fig.3 shows the per phase Norton's equivalent circuit of an IPM motor. The phasor current triangle of the Norton's equivalent circuit of a synchronous motor is governed by the Kirchoff's current law of **If + Is = Im**. Note that **Is** is the stator current per phase**, Im** is the magnetising current per phase and $\mathbf{I_f}$ is the phasor current arising out of the rotor permanent magnet excitation. It is quite well known that a conventional synchronous motor can be operated at variable power factor modes by regulating its dc field current I_{fdc}. It is well known that the dc excitation current **If** is varied by controlling the rotor field current I_{fdc} to operate the motor at unity, leading and lagging power factor modes of operation. It is not possible for IPM synchronous motors.

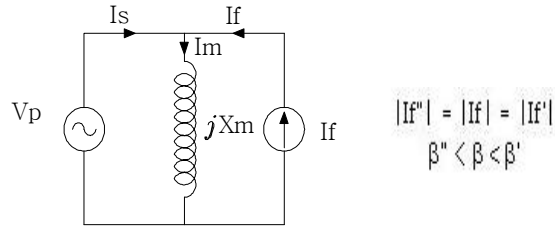


Fig.3: Norton's Equivalent Circuit of IPM Motor

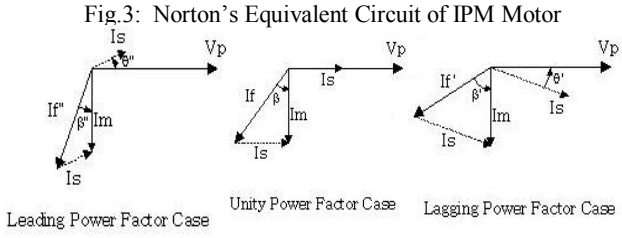


Fig.4: Current Phasor Diagram of IPM Motor

Unlike conventional wire-wound synchronous machines, the rotor of an IPM motor is permanently excited. The rotor permanent magnet can be modeled by equivalent current source, as indicated by **If** in Fig. 3. The excitation current **If** due to rotor permanent magnets for IPM synchronous motor is constant.

However, the IPM motor can be operated by controlling the angle β between the magnetizing current **Im** and constant excitation current **If**. This is explained by means of the current phasor triangle. An IPM motor can be operated in leading, unity and lagging power factor modes of operation by varying the angle β, as shown in Fig.4. This eliminates another constraint for its wide spread applications in industry as the singly fed permanently excited variable power factor IPM synchronous motor. Fig 5 shows an alternate method of varying power factor of an IPM synchronous motor. The direct and quadrature axis (d-q) components of the stator current may be controlled by the

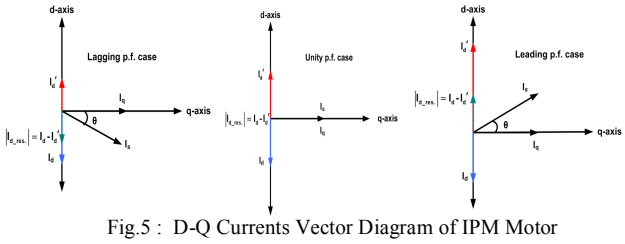


Fig.5 : D-Q Currents Vector Diagram of IPM Motor

vector control (field weakening) technique, in which the d-axis current is varied to operate the IPM motor in leading, lagging and unity power factor modes of operation.

VI. ROTOR DESIGN FOR LINE START IPM

The earlier design of the rotors for IPM motors using Ferrite magnets was geared to increase the air gap flux by arranging the magnets and their orientation. Different old topologies had been tried by Binns [3-4]. Modern NdBFe magnets having high B_r and very large H_c lead the recent trend for new rotor designs.

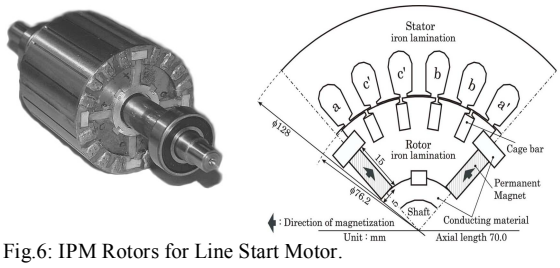


Fig.6: IPM Rotors for Line Start Motor.

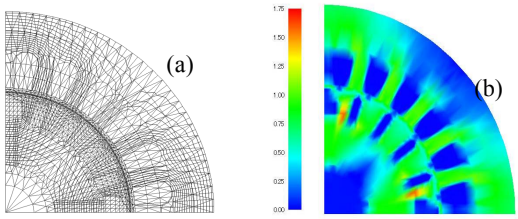


Fig. 7: Magnet flux distribution
(a) FE grids (b) magnetic flux contours

Fig.6 shows the experimental rotor for an IPM motor with line start provision [41]. The left hand picture depicts the IPM rotor, and the right hand figure provides details of dimensions and permanent magnet orientation over one quadrant of 4-pole IPM rotor [41]. Figure 7 shows the finite element grids and magnet flux density contours for one quadrant of the IPM rotor of Fig.6.

The design data for an experimental 3-phase, 4-pole, Y-connected, 200V, 1 hp IPM motor with NdBFe magnets (Neomax –32) are given as; stator : OD= 128mm, ID = 77mm, stack length = 70 mm, number of slots = 24, conductors/slot = 56. rotor: OD = 76.2 mm, core length = 70 mm, inertia = 0.0015 kgm^2 and load inertia = 0.0263 kgm^2. The rotor consists of 2 cages of aluminum bars with lower cage of 7.5 mm depth [41].

Table-1 shows the comparative test performance results of the IPM synchronous motor and induction motor (IM). Both the motors were run at 140V (L-L) and 200V (L-L) voltages. It is quite evident from the test results of Table-1 that the IPM motor outperformed in every category of performances.

Table-1: Performance Results of IPM and Induction Motors

Quantity	IPM Rotor	IM Rotor
Input voltage: V (V)	140	200
Input current: I (A)	2.91	3.43
Input power: W (W)	696	818
Rotor speed: n (rpm)	1500	1434
Torque: T (Nm)	3.82	4.00
Efficiency: η (%)	86.2	73.3
Power factor: pf (%)	98.6	68.8
Output power: P (W)	600	600
Off. X pf product (%)	85.0	50.4
Max output: Pmax. (W)	1115	1240

The significant conclusion is that the efficiency and power factor as well as their product of the IPM motor is over 35% better than that for an identical rated induction motor. The energy efficiency aspect is a key factor for wide spread applications of high performance IPM motor drives.

VIII. APPLICATIONS

IPM motors with intelligent power module (IPM) are now widely used for heavy duty cycle loads, which include ventilation fans, blowers, air conditioner, heat pumps, compressors, cranes, elevators/escalators, blood pumps, ship propellers, locomotive traction drives, electric and hybrid electric vehicles (HEV). The ratings span from few watts to few megawatts range.

Double IPM motors are now increasingly used for energy saving applications in hybrid electric vehicles. The key requirements of IPM propulsion motors for HEV applications include the following [25,30,32,38]: high torque and power density, high torque at low speeds for staring and uphill climb but high power at high cruising speeds, maximum efficiency over wide speed and torque ranges including at low torques, wide speed range with constant power mode, exceeding 2-4 times the base speed, optimum compromise between motor peak torque and inverter volt-ampere ratings, short term overload capability, typically twice the rated torque over short duration, low acoustic noise, low cogging torques, low torque ripples, optimized stator distributed winding with minimum total harmonic distortion factor, innovative rotor design topology with magnets orientation for maximum variation of d-q axis inductances, reduction of cross-coupling magnetic saturation, least magnet flux leakage, magnet demagnetization withstand with respect to armature reaction, temperature and surface corrosion constraints, excessive open circuit back–emf, load and no load stator iron loss at high speeds, high reliability and robustness for various operating conditions, minimum weight and smallest size, low fuel consumption rate (litre/km), clean, quiet, smooth, powerful, efficient and low cost. It is obvious that many of the above mentioned design requirements are complex, some times conflicting and interlinked for specific HEV applications. Furthermore, these design criteria cannot be isolated from their control strategy including power electronic converter and battery. Figure 8 shows the per unit torque/power and efficiency over wide speeds for hybrid electric vehicles [30].

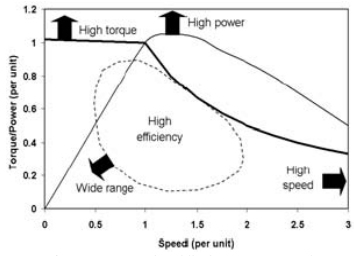


Fig. 8: Torque/power and efficiency requirements for HEV

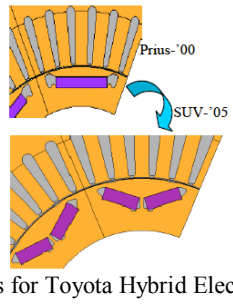


Fig.9: IPM Rotors for Toyota Hybrid Electric Vehicles [45]

Figure 9 shows the IPM rotors for Toyota Prius 2000 model and sports utility vehicles (SUV2005) model of hybrid electric vehicles [45]. Table-2 shows the utilization of IPM motors for Toyota Prius and sports utility vehicles (SUV) to create variation of d-q inductances of the rotor magnets topology for reluctance torque. However, the V type arrangements are preferred for hybrid electric vehicle applications, where the reluctance torque component is critical for high-speed operation in flux weakening regime.

Table-2: d-q axis inductances (mH) [45]

Axis	Straight IPM (Prius)	V- IPM (SUV)
d-axis Ld	1.06	0.86
q-axis Lq	2.26	2.23
Lq –Ld	1.20	1.37
Lq/Ld	2.13	2.59

For SUV 2005 Toyota models the reluctance torque component is about 63% of the total driving torque at a speed of 12,400rpm. The torque to weight ratio significantly improves by operating the IPM motor at 650 Vdc from smart dc-dc converter. The light load stator iron loss also decreased primarily by employing high-grade silicon steel for IPM motors.

Figure 10 provides an illustration of a 123kW IPM motor/generator set for the Toyota hybrid electric car. The back wheel IPM motor is rated at 50 kW for the 4-wheel drive model. The sophisticated and intelligent control in a hybrid electric

vehicle forms the key to successful utilization of IPM motors. Smart power electronic modules as well as new nickel metal hydride battery are the enabling technology for the popular Toyota 'Prius' sedans and 4-wheel SUV models. The design of Toyota hybrid system (THS) includes gasoline engine, new transmission system, IPM traction motor, IPM generator, converter/inverter module, battery and control units. Figure 11 shows the complete transmission layout of the Toyota hybrid system (THS) for its popular hybrid electric vehicle models. This innovative THS transmission is geared to achieve maximum fuel efficiency and a high degree of driving comfort.PM traction drive motors are crucial for fulfilling the power characteristics required for high performance automobiles.

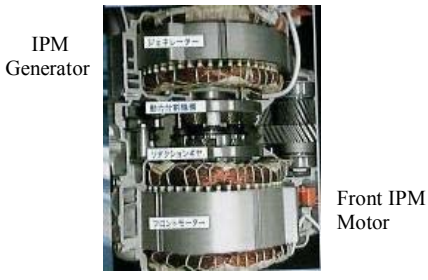


Fig 10: IPM Motor / Generator Set for Toyota Hybrid Car

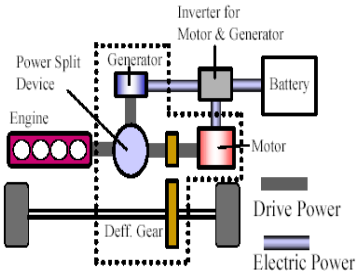


Fig. 11: Toyota Hybrid System (THS) Transmission

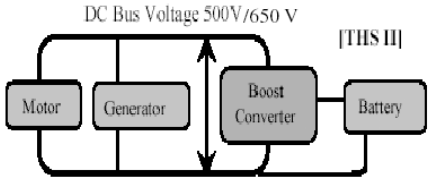


Fig. 12: Layout for Motor/Generator, dc/dc converter

Figure 12 shows the layout of IPM traction motor, IPM generator, dc/dc converter and batter systems. Fig.13 graphically shows the contributions of electric torque due to permanent magnet and the reluctance torque produced within the IPM rotor for the 2000 Prius and 2005 SUV models. This confirms the better choice of IPM motors technology for HEVs. Perhaps it ends the debate of using either the induction motors or reluctance motors for efficient traction drives for mass transportations. Fig. 14 shows the efficiency contours of the Toyota SUV 2005 models at extended speeds of operation, respectively. It is to be noted that Table-3 contains the uses of IPM motor drives in Japanese hybrid electric vehicles.

VIII. CONCLUSIONS

This paper gives a brief introduction to the emergence of high efficiency interior permanent magnet (IPM) synchronous motors. A list of references provides a state of the art survey of significant as well as few incremental contributions in chronological order of appearance over the past 50 years. Simple expressions for developed power and torque are given.

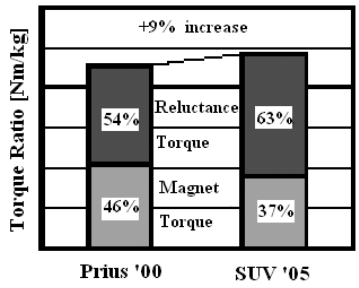


Fig. 13: Ratio of Magnet and Reluctance Torques

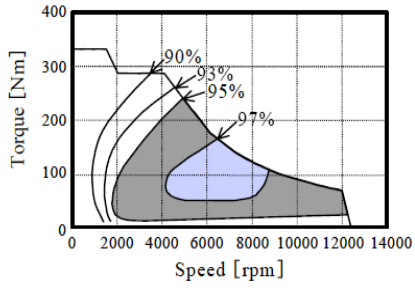


Fig. 14: Efficiency Contour for SUV 2005

Table-3: Application of IPM Motor drives in Japanese Hybrids

Year	Company	Brand	Vehicle type	Power	Voltage	km/liter
1997	Toyota	Prius	Sedan	30kW	274V	22
2000	Toyota	Prius-1	Sedan	33kW	288V	22.5
2004	Toyota	Prius-2	Sedan	50kW	500V	25.0
2005	Toyota	Camry	Sedan	60kW	650V	25.6
2005	Toyota	Kluger*	V6-SUV	123kW	650V	17.8
2005	Toyota	Estima*	V6-Van	123kW	650V	18.6
2005	Toyota	Harrier*	V6-SUV	123kW	650V	17.8
2007	Toyota	Lexus	Sedan	147kW	650V	20.0

* Japanese 4WD, front motor/generator, 123 kW, 12400 rpm, rear motor, 50 kW, nickel metal hydride battery

Rotor design features for specific applications are briefly covered. On line and soft starting provisions are included. Operation of IPM motors at variable power factor is illustrated with the help of Norton's equivalent circuit and phasor diagrams. Comparative performances of IPM and induction motors are summarized. An example of successful traction application is given. Highlight of IPM motor drives includes its wide spread application in Japanese hybrid electric vehicles, which are just one of many items of ac motor drives in passenger automobiles to save precious energy. The paper opens up the debate on plug-in, solar and auto-charged smart hybrid electric vehicles [48-49]. It concludes by hinting on the economic, environmental and social impacts in poor countries.

REFERENCES

1. M.A. Rahman, "Combination Hysteresis, Reluctance, Permanent Magnet Motor", US.Patent No.5,187,401;Issue date:Feb.16, 1993.
2. F.W. Merrill, " Permanent Magnet Excited Synchronous Motors", AIEE Transactions, vol.74, 1955, pp 1754-1760
3. K.J. Binn, W.R. Barnard and M.A. Jabbar, " Hybrid Permanent Magnet Synchronous Motors", IEE Proc., Pt.B, vol.125, No.3, 1978, pp203-208
4. M.A. Rahman, "Permanent Magnet Synchronous Motor - A Review of the Design Art", Proceedings of International Conference of Electrical Machines-ICEM'80, Athens, Greece, Sept. 15-17, 1980, pp. 312-319.
5. V.B. Honsinger, " The Fields and Parameters of Interior Type AC Permanent Magnet Machines", IEEE Transactions on Power Apparatus and Systems, vol. PAS-101,No.4, 1981, pp.867-876

6. M.A.Rahman, NSERC-PRAI-CGE Project on IPM, MUN, 1982.
7. M.A. Rahman, T.A. Little and G.R Slemon, "Analytical Models for Interior-Type Permanent Magnet Synchronous Motors", IEEE Trans. on Magnetics, vol. MAG-21, No.5, 1985, pp. 1741-1743.
8. T. Sebastian, G.R. Slemon and M.A. Rahman, " Modelling of Permanent Magnet Synchronous Motors", IEEE Transactions on Magnetics, vol. MAG-22, No.5, 1986, pp. 1069-1071
9. T.J.E. Miller, "Single Phase Permanent Magnet Motor Analysis", IEEE Transactions on Industry Applications, vol.21, No.4, 1986, pp. 651-658.
10. M.A.Rahman and G.R.Slemon, "Tutorial on Permanent Magnet Motor Design", IEEE Intermag-86, Phoenix, April 16, 1986.
11. T.M. Jahns, G.B. Kliman and T.W. Neumann, " Interior PM Synchronous Motors for adjustable Speed Drives", IEEE Trans. on Industry Applications, vol.22, No.4, 1986, pp. 738-747.
12. B.J. Chalmers, S.A. Hamed and G.D. Baines "Parameters and Performance of High-field Permanent Magnet Motors", IEE Proc. Pt. B, vol.32, No. 3, 1986.
13. T.M.Jahns, " Flux-weakening Regime Operation of an Permanent Synchronous Motor Drive", IEEE Transactions on Industry Applications, vol.23, No.4, 1987, pp. 681-689.
14. M.A. Rahman and A.M. Osheiba, "Performance Analysis of Single Phase Permanent Magnet Synchronous Motors", proceeding on Electric Energy Conference, Adelaide, Australia, Oct. 6-9, 1987, pp. 514-519.
15. T. Sebastian and G.R. Slemon, "Operating Limits of Inverter Driven Permanent Magnet Synchronous Motor Drives", IEEE Transactions on Industry Applications, vol.23, No.2, 1987, pp. 327-333.
16. M.A. Rahman and A.M. Osheiba, "Performance of Line-start Single Phase Permanent Magnet Synchronous Motors", IEEE IAS Annual Meeting, Atlanta, Oct. 19-23, 1987, pp. 104-108.
17. P. Pillay and R. Krishnan, "Control Characteristics and Speed Controller for a High Performance Permanent Synchronous Motor Design, Proceeding of IEEE PESC, 1987, pp. 598-606
18. G.R. Slemon and T. Li, " Reduction of Cogging Torques in Permanent Magnet Synchronous Motor", IEEE Transactions on Magnetics, vol.24, No.6, 1988, pp. 2901-2903.
19. B.K. Bose and P.M. Szczesny, " A Micro-controller based Control and Simulation of an advanced IPM synchronous Machine Drive System for Electric Vehicle Propulsion", IEEE Transactions on Industrial Electronics, vol. 35, No. 4, 1988, pp. 547-559.
20. A. Fratta, A. Vagati and F. Villata, "Design Criteria of an IPM Machine Suitable for Field-weakening Operation", Proceedings, ICEM, MIT Cambridge, 1990, pp. 1059-1065.
21. S. Morimoto, Y.Tadaka, T.Hirasa and K.Taniguchi, " Expansion of Operating Limits for Permanent magnet Motor by Optimum Flux-Weakening8, IEEE Transactions on Industry Applications, vol.26, no. 5, 1990, pp. 966-871.
22. A. Consoli and C. Antonio, " A DSP based Sliding Mode Field Oriented Control of an Interior Permanent Magnet Synchronous Motor Drive", Proceeding of IPEC, Tokyo, April 3-6, 1990, pp. 296-303.
23. R.F. Schiferl and T.A. Lipo, "Power Capability of Salient Pole Permanent Magnet Synchronous Motors in Variable Speed Drive Applications", IEEE Transactions on Industry Applications, vol.26, No.1, 1991, pp. 115-123.
24. A.B. Kulkarni and M. Ehsani, " A Novel Position Sensor Elimination Technique for Interior Permanent Magnet Synchronous Motor Drive", IEEE Transactions on Industry Applications, vol.28, No.1, 1992, pp. 141-150
25. Z.Q. Zhu and D. Howe, "Influence of design parameters on Cogging Torque in Permanent Magnet Machines", IEEE Transactions on Energy Conversion, vol. 15, No. 5, 1992, pp 407-412
26. M. A. Rahman and Ping Zhou, "Field Based Analysis of Permanent Magnet Motors", IEEE Transactions on Magnetics, vol. 30, No.4, 1994, pp. 3664-3667
27. S. Morimoto, M. Sanada and Y. Takeda, " Effects of Compensation of Magnetic Saturation in Flux-weakening Controlled Permanent Magnet Synchronous Motor Drives", IEEE Transactions on Industry Applications, vol.30, No.6, 1994, pp. 1632-1637.
28. Ping Zhou, M. A. Rahman and M. A. Jabbar, "Field and Circuit Analysis of Permanent Magnet Machines", IEEE Transactions on Magnetics, vol. 30, No. 3, July 1994, pp. 1350-1359.
29. M. Ooshima, A. Chiba, T. Fukao and M. A. Rahman, "Design and Analysis of Radial Force in a Permanent Magnet Type Bearingless Motor", IEEE Transactions on Industrial Electronics, vol. 43, No. 2, 1996, pp. 292-299
30. Y.Honda, T. Nakamura, T. Higaki and Y. Takeda, " Motor Design Consideration and Test results of an Interior Permanent Magnet Motor for Electric Vehicles", IEEE–IAS Annual Meeting, New Orleans, vol. 1, 1997, pp.75-82.
31. M.A. Rahman and M.A. Hoque, "On-line Adaptive Artificial Neural Network based Vector Control of Permanent Magnet Synchronous Motor", IEEE Transactions on Energy Conversion, vol. 13, No. 4, December 1998, pp. 311-318.
32. Y.Honda, T. Higaki, S. Morimoto and Y. Takeda, " Rotor Design Optimization of a multi-layer Interior Permanent Magnet Synchronous Motor ", IEE Proc., Electric Power Applications, vol. 145, No.2, 1998, pp. 119-124
33. L Zhong, M.F. Rahman, W.Y.Hu, K.W. Lim and M.A. Rahman, "A Direct Torque Controller for Permanent Magnet Synchronous Motor Drive", IEEE Transactions on Energy Conversion, vol.14, No.3, December 1999, pp.637-642.
34. S. Vaez, V.I. John and M.A. Rahman, "An on-line Loss Minimization Controller for Interior Permanent Magnet Motor Drives", IEEE Transactions on Energy Conversion, vol.14, No.4, 1999, pp.1435-1440.
35. A.Consoli, G. Scarcella and A. Testa, " Industry Application of Zero –Speed Sensorless Control Techniques for PM Synchronous Motors", IEEE Transactions on Industry Applications, vol.37, No.2, 2001, pp. 513-521.
36. W.L. Soong and E. Ertugrul, " Field Weakening Performance of Interior Permanent Magnet Motors", IEEE Transactions on Industry Applications, vol.38, No.5, 2002, pp. 1251-1258.
37. N. Bianchi and A. Canova, "FEM Analysis and Optimization Design of an IPM Motor", Proc. of PEMD'02, Bath, UK, April 2002, pp.16-18.
38. C.C. Chan, " The State of the Art of Electric and Hybrid Vehicles", Proc. IEEE, vol.90, No. 2, 2002, pp. 247-275.
39. M.A. Rahman, M. Vilathgamuwa, M.N.Uddin and K.J.Tseng, " Non-linear Control of Interior Permanent Magnet Synchronous Motor", IEEE Transactions on Industry Applications, vol.39, No.2, 2003, pp.408-416.
40. J. Ha, K.Ide, T. Sawa and S. K. Sul, " Sensorless rotor position estimation of Interior Permanent Magnet Motor from Initial states", IEEE Transactions on Industry Applications, vol.39, No.3, 2003, pp. 761-767.
41. K. Kurihara and M.A. Rahman, "High Efficiency Line-Start Interior Permanent Magnet Synchronous Motors", IEEE Transactions on Industry Applications, vol. 40, No.3, 2004, pp. 789-796.
42. N. Bianchi and T.M. Jahns, etal, Tutorials on "Design, Analysis and Control of Interior PM Synchronous Machines", IEEE IAS Annual Meeting, Seattle, Oct. 12, 2004
43. Yu-seok Jeong, R.D. Lorenz, T.M. Jahns and S.K. Sul, " Initial Position Estimation of an IPM Synchronous Machine using Carrier-frequency Injection Methods", IEEE Transactions on Industry Applications, vol.41, No.1, 2005, pp. 38-45
44. M.N.Uddin, M.A. Abido and M.A.Rahman' " Real-Time Performance Evaluation of a Genetic Algorithm Based Fuzzy Logic Controller for IPMSM Drives", IEEE Transactions on Industry Applications, vol.41, No.1, 2005, pp.246-252
45. M. Kamiya, "Development of Traction Drive Motors for the Toyota Hybrid System", Proc. IPEC-2005, Niigata, April 4-8,2005, Paper S43-2
46. M. A. Rahman, "Advances in Interior Permanent Magnet (IPM) Motor Drives", Proceedings of Inaugural IEEE PES 2005 Conference and Exposition in Africa, Durban, South Africa, 11-15 July 2005, pp.372-377.
47. W.D. Jones, "News, Take This Car And Plug It", IEEE Spectrum, July 2005, pp. 10-13.
48. M.A. Rahman, "Advances of Modern IPM Motor Drives for High Performance Applications", Proceedings of International Conference on Electric Machine Systems (ICEMS-2006), Nagasaki, Japan, November 22, 2006.
49. J. Voelcker, "IEEE-USA promotes Plug-in Hybrid", The Institute, IEEE, November 2007, www.theistitute.ieee.org.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Analysis of Efficiency Optimization for PFM Mode Switching DC-DC Boost regulator

[1] Khondker Zakir Ahmed, [2] Moakhkhrul Islam, [3] Syed Mustafa Khelat Bari, [4] Didar Islam, [5] Mohiuddin Hafiz and [6] Quazi Deen Mohd Khosru

[1, 2, 3, 4] Power IC Ltd., Dhaka, Bangladesh, [5, 6] Dept of EEE, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
E-mail: zakir.ak@gmail.com

Abstract – A mathematical analysis of efficiency optimization for PFM (pulse frequency modulation) mode boost regulator has been presented in this paper. Based on the load demand, input voltage and output voltage a PFM mode booster can operate in single pulse or multi pulse operation. The presented analysis reveals a relationship between operating efficiency and mode of operation considering the external and internal parameters for which loss occurs. This paper also presents a relationship between maximum load delivering capacity and inductor size where separate modes of operation are distinguished.

I. Introduction

Integrated power supplies are critical building blocks in the state-of-the-art portable applications. Among the various techniques used in power management ICs, PFM (pulse frequency modulation) mode offers a relatively higher quality with much simplicity in design. PFM mode booster requires simple design techniques and delivers very high efficiency for a wide range of load current [1]. Especially for light load condition PFM mode booster offers much higher efficiency than other topologies [2]. Although many works has been done on the optimization of efficiency among different modes of topology, an accurate model for the load capability of a PFM mode booster is yet to explore. This paper presents a mathematical model for efficiency optimization of a PFM mode boost regulator based on the mode of operation at any particular instance. Load capability of PFM booster at given inductor size and input voltage is also presented which distinguish separate operating condition of aforementioned booster.

II. Theory and modelling

PFM mode boost regulators are usually suitable for light load conditions. Due to the application in light load conditions, PFM mode boosters usually operate in discontinuous conduction mode (DCM), where the inductor current ramps down to zero at the end of each cycle. When the load demand increases the idle time between two DCM pulse decreases – thus increasing the average load current at the output. When the load is increased enough so that the average load current demand is more than the current provided by the contiguous DCM pulses with no idle time, the booster automatically switches to multi-pulse operation. In multi-pulse operation inductor current doesn't ramp down to zero after each switching cycle but ramps higher at each successive pulse until ramps down to zero after the second or the third pulse. This analysis focuses on multi-pulse operation where after two contiguous pulses the current ramps down to zero before starting another group of two-pulses. Between the two groups of such contiguous pulses there exist an idle time. Fig. 1 shows the inductor current wave shapes in both the single pulse and multi-pulse operation as described above.

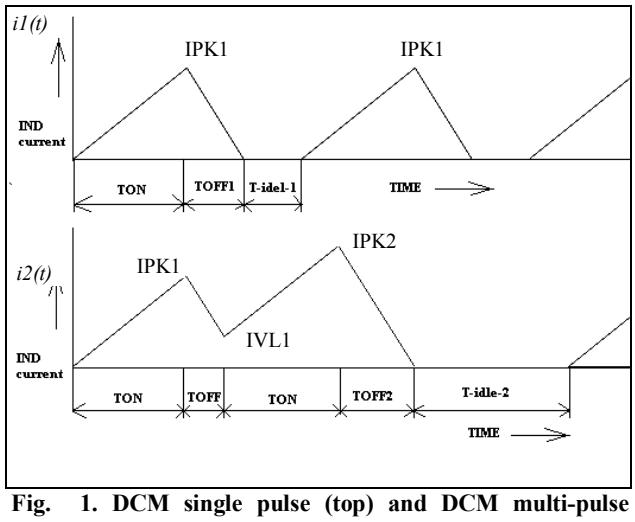


Fig. 1. DCM single pulse (top) and DCM multi-pulse (bottom) operation for a PFM booster.

A basic booster configuration is shown in Fig. 2. The main function of the booster is to control the ON time and OFF-time of the NMOS. During the ON-time, the gate of the NMOS remains high and current builds up through the inductor and NMOS. The Schottky diode is reverse biased and no current flows from SW-node to VOUT-node. During OFF-time, as the NMOS transistor is switched OFF, the inductor current force the SW voltage to jump high and the Schottky diode turns on. During OFF time the output capacitor is charged up by the inductor current.

To analyze efficiency, major loss producing elements are studied. Four major loss producing elements are identified as the Equivalent Series Resistance (ESR) of the inductor (R_{LESR} = 0.3Ohm), the channel resistance of the NMOS in triode region (R_{DSON} = 0.5Ohm), the switching loss due to gate capacitance of the NMOS (C_G = 230pF) and finally the forward diode conduction loss of the Schottky (V_F = 0.3V). The value of ESR of the inductor and the forward Schottky drop are for standard elements used in board and the channel resistance and gate capacitance of the NMOS transistor are measured value for a fabricated PFM mode booster.

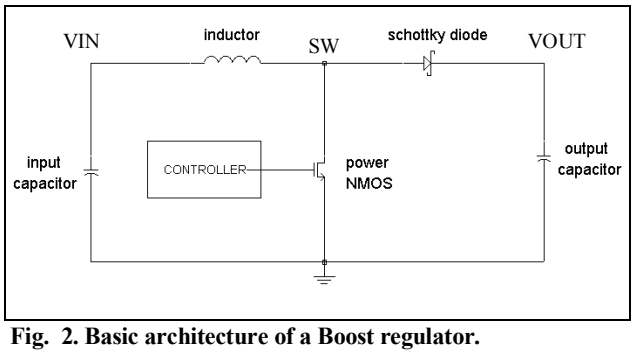


Fig. 2. Basic architecture of a Boost regulator.

As the inductor carries current for both the ON time and OFF time during the operation, the calculation for total energy loss due to ESR takes a form as shown in Eqn.(1).

$$E1 := \int_0^{T1} R_{LESR} \cdot (i1(t))^2 \, dt \quad \text{---------- (1)}$$

Where $T_1 = T_{ON} + T_{OFF} + T_{idle1}$ referring to Fig. 1. For the loss due to the channel resistance of the NMOS, similar equation is used except the value of the resistance which is 0.5Ohm and since the transistor conducts current during ON-time, the integration is done for ON-time only. This is expressed in Eqn. (2).

$$E2 := \int_0^{TON} R_{DSON} \cdot (i1(t))^2 \, dt \quad \text{---------- (2)}$$

The Schottky conducts current during the OFF-period and so the energy lost across Schottky diode can be calculated integrating the current wave during OFF-time. This is expressed in Eqn. (3)

$$E3 := \int_{TON}^{TON+TOFF} V_F \cdot i1(t) \, dt \quad \text{---------- (3)}$$

Finally, the loss due to the gate capacitance is calculated using Eqn. (4).

$$E4 := \frac{1}{2} C_G \cdot VDD^2 \quad \text{---------- (4)}$$

Where VDD is the input supply voltage.

Similarly for the DCM multi pulse operation, the total energy associated with the loss producing elements can be equated as the following equations.

Eqn. (5) expreses the loss due to the inductor ESR.

$$E5 := \int_0^{T2} R_{LESR} \cdot (i2(t))^2 \, dt \quad \text{---------- (5)}$$

Where, $T2 = 2*T_{ON}+T_{OFF}+T_{OFF2} + T_{idle2}$ referring to Fig. 1. Eqn. (6) describes the loss due to the NMOS ON resistance during Two ON periods for the multi pulse situation.

$$E6 := \int_0^{TON} R_{DSON} (i2(t))^2 \, dt + \int_{TON+TOFF}^{2TON+TOFF+TOFF2} R_{DSON} (i2(t))^2 \, dt$$

---------- (6)

Eqn. (7) states the loss due to Schottky diode during two OFF phases.

$$E7 := \int_{TON}^{TON+TOFF} V_F \cdot i2(t) \, dt + \int_{2TON+TOFF}^{2TON+TOFF+TOFF2} V_F \cdot i2(t) \, dt$$

---------- (7)

And Eqn. (8) describes switching loss component due to the gate capacitance. Since for the described multi pulse case, the there are two switching pulse phases the capacitance loss is doubled compared to single pulse case.

$$E8 := 2 \cdot \left(\frac{1}{2} \cdot C_G \cdot VDD^2 \right) \quad \text{---------- (8)}$$

III. Results and Discussion

The analysis has been performed considering boost operation with input supply voltage (V_{DD}) 1.5 volts, output voltage (V_{OUT}) 3.3 volts and inductor (L_{IN}) 10 micro-Henry. T_{ON} and T_{OFF} are set by the internal circuitry with values 0.750 micro-seconds and 0.250 micro-seconds. External parameters I_{PK1}, I_{PK2}, I_{VL1}, T_{OFF1}, T_{OFF2} , T_{idle1} and T_{idle2} are set by the input voltage, output voltage and the inductor value. We calculated these parameters for both single pulse and multi pulse operation assuming in both cases the booster is providing 20mA load at output. The list of calculated parameters is shown in Table 1. These parameters are used in calculating energy consumption by each element using Eqn. (1) through Eqn. (8). After the total energy consumption at each mode of operation is calculated, we finally calculated average Power loss for all the elements. Table 2 shows the calculated power loss by each element separately. For single pulse operation the average power calculated is 4.012mW and for multi pulse operation the average power calculated is 4.260mW. For multi pulse operation, the average power is 6.2% higher than the single pulse operation.

Table 1 List of calculated parameters.

Parameter	Sinlge Pulse	Multi Pulse
TON	0.750uS	0.750uS
TOFF	0.250uS	0.250uS
TOFF1	0.625uS	-
TOFF2	-	1.000uS
Tidle1	2.225uS	-
Tidle2	-	9.250uS
T1	3.600uS	-
T2	-	12.000uS
IPK1	112.5mA	112.5mA
IPK2	-	180.0mA
IVL1	-	67.5mA

Table 2 Element wise Power Loss

Elements	Sinlge Pulse	Multi Pulse
	Energy (nano Joules)	Energy (nano Joules)
Rlser	1.74	8.49
Rdson	1.582	7.718
VF Shottky	10.54	33.75
CG switching loss	0.58	1.16
Total Energy (nJ)	14.442	51.118
Time (uS)	3.6	12
Pavg (mW)	4.012	4.260

To analyze we need to look at the averge contribution of the loss producing elements. Table 3 shows average contributions of each element separately. As we see, for multi pulse operation, gate capacitance switching loss and loss due to Schottky diode drop is lower than single pulse operation. Although for multi pulse operation the booster generates two switching pulses compared to one pulse in

Table 3 Average contributon to total Power Loss

Elements	Sinlge Pulse	Multi Pulse	
	Average Contribution	Average Contribution	Differe-nce
	(mW)	(mW)	(mW)
Rlser	0.483	0.708	+0.224
Rdson	0.439	0.643	+0.204
VF shottky	2.928	2.813	-0.115
CG switching loss	0.161	0.097	-0.064
Pavg (mW)	4.012	4.260	+0.248

single pulse operation, total pulse repititaion time for multi pulse is almost three times than single pulse operation (ratio of T2 and T1). So the average switching loss decreases for multi pulse operation. For Schottky diode loss, since the average ON time for which the diode conducts current is smaller in multi pulse operation this one is also expected to decrease. Rest of elements produce resistive losses. Since resistive loss increases as the square of the current, multi pulse operation registers a higher loss than single pulse operation due to the higher peak current.

For practical test data we present the following figures. Here the booster is seen to operate in two modes of operation at different load condition. This is different than the condition we used in calculation where the load was assumed fixed for both mode of operations. But for practical purpose, once a chip is build the topology is fixed and it is expected to switch between different modes of operation autonomously depending on the external parameters. Here we will compare the difference in efficiency due to the change of mode of operation assuming nothing else changes significantly with load for which the IC changes modes of operation. Fig. 3 shows the booster operating at single pulse operation while supplying an external load of 10mA.

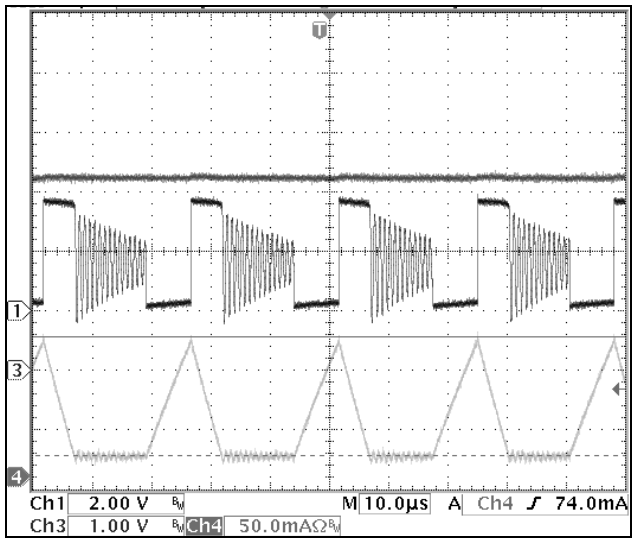


Fig. 3. With 10mA load, the booster is operating at DCM single pulse operation. Channel-1: SW (2V/div) Channel-3: VOUT (1V/div), Channel-4: Inductor Current (50mA/div)

Fig. 4 shows the booster operating at DCM multi pulse operation while supplying 20mA load current. At this case, inductor current doesn't go to zero value at each pulse but builds up to a higher level than the single pulse operation. The pattern of current repeat after two successive pulses.

We have found in our analysis that for multi pulse operation average power loss is higher than single pulse operation. Now as the booster is supplying 10mA load it is operating at single pulse operation (Fig. 3) and when it is supplying 20mA load it operates at multi pulse operation (Fig. 4). Based on the analysis it is expected that for multi pulse operation the efficiency will be lower

since the loss is higher. And that is exactly what we have found in actual test data. . Fig. 5 shows the efficiency data for the booster. The efficiency decrement above 10mA load can be fully explained by the result of the analysis along with Fig. 3 and Fig. 4. As the load is increased, the booster changes it mode of operation that results in higher loss which in turn shows up as lower efficiency.

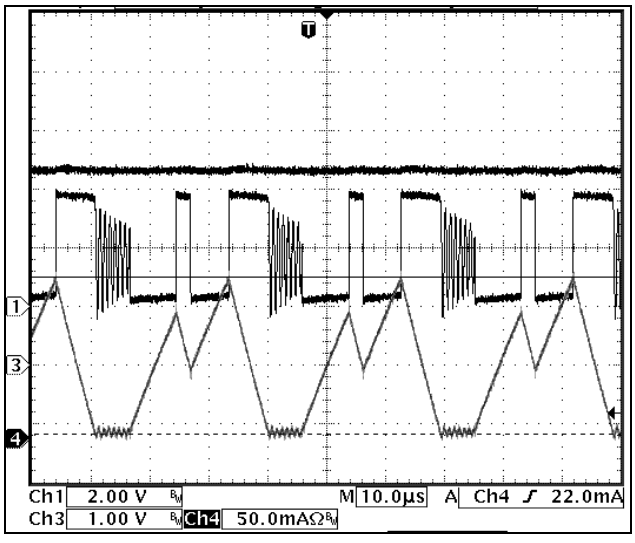


Fig. 4. With 20mA load, the booster is operating at DCM multi pulse operation. Channel-1: SW (2V/div) Channel-3: VOUT (1V/div), Channel-4: Inductor Current (50mA/div).

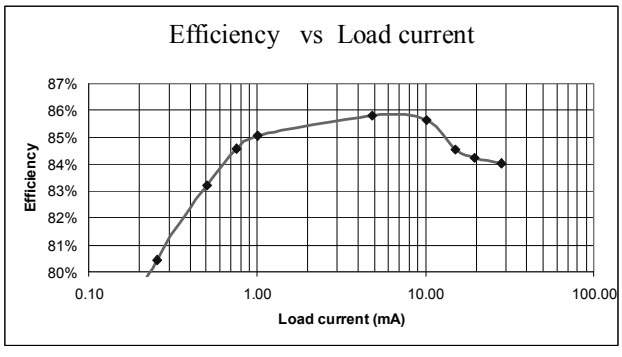


Fig. 5. Efficiency decreases after 10mA where the booster switches from single pulse operation to multi pulse operation.

Fig. 6 shows a graphical analysis of load capability of a booster as a function of inductor size at different input supply. The dashed line at middle of the curve is the boundary of modes of operation. For a given inductor size the booster might operate in single pulse operation or in multi pulse operation depending on the value of input voltage. This curve along with the result of the analysis is very much useful in selecting the operating region of a booster for efficiency optimization. Since single pulse operation offers higher efficiency than multi pulse operation, it is always desirable to operate in single pulse mode which will ensure maximum efficiency. This can be done by analyzing Fig. 6 and then selecting right input set of parameter which will ensure single pulse operation.

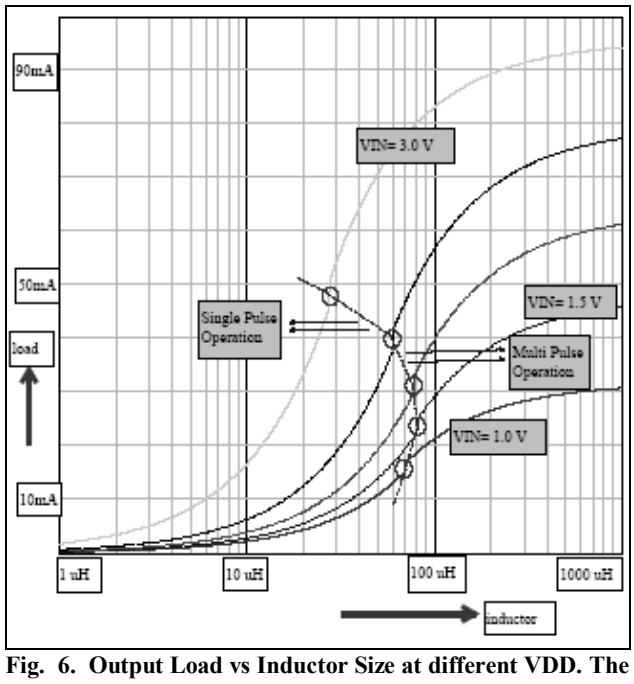


Fig. 6. Output Load vs Inductor Size at different VDD. The dashed line at the middle corresponds to the edge between two modes of operation. Left to the line is single pulse operation region and right to that is multi pulse operation region.

IV. Conclusion

In this paper we have presented an analysis for efficiency optimization for PFM mode boost regulators. By calculating the contribution of the individual loss producing elements the analysis reveals a relationship between modes of operation and efficiency of the booster. The analysis shows that for PFM mode boosters, single pulse operation is preferable than multi pulse operation as multi pulse operation produces higher loss than its' counterpart. Test data has been presented which verifies the result of the analysis. In addition to the loss calculation, load capacity of PFM mode booster is also presented which distinguish the modes of operation for a given set of input voltage and inductor size.

References

[1] B. Sahu, G. A. Rincon-Mora, "A High-Efficiency, Dual-Mode, Dynamic, Buck-Boost Power supply IC for portable applications", *vlsid*, pp. 858-861, 18th international conference on VLSI Design held jointly with 4th International Conference on Embedded Systems Design (vlsid'05), 2005.

[2] H. Deng, X. Duan, N. Sun, Y. Ma, A. Q. Huang. D. Chen "Monolithically Integrated Boost Converter based on 0.5-/spl mu/m CMOS process", IEEE transactions on power electronics, vol. 20, no. 3, pp. 628-638, September 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Efficient Design of Power Transistor in Switching Regulator

Mohiuddin Hafiz[1], Syed Al-Kadry[2], Tania Ansari[3], and Khondker Zakir Ahmed[4]

1. Research Institute for Nano Device & Bio Systems, Hiroshima University, Japan; 2. Dept. of EEE, University of California, Riverside; 3.Dept of EEE, Bangladesh University of Engineering & Technology; 4. Power IC Limited.
E-mail: hafiz2431@hotmail.com

Abstract - The details of design methodology of integrated switches for switching converters of a widely-used topology, along with various aspects, have been presented in this paper. Mathematical model has been established to predict the contributions of various parasitic resistances of the topology to its performance. The switch sample has been fabricated with a standard 0.5-µm dual metal process. The developed model shows a good match with the simulated and test data.

I. Introduction

In power management systems, the DC-DC switching regulator, providing some advantages compared to linear regulators like higher efficiency, requirements of smaller components and less thermal management, possibility of transforming of input voltage to higher output voltage (as in the case of boost regulator) [1], has become an indispensable part and as the trend of voltage scaling is not only limited to digital circuits, but also spreading to other applications like display panels of future cellular phones and portable devices, it's getting more and more importance[2],[3]. In the switching regulators, one of the largest power-loss contributing factors for switchers is the rectifying diode. The power dissipated is equal to the forward voltage drop multiplied by the current going through it. The reverse recovery for silicon diodes can also create loss. These power losses reduce overall efficiency and require thermal management in the form of a heat sink or fan [1]. To minimize the losses and maximize the efficiency, integrated MOSFET switches are used instead of diodes. However, the MOS transistors used for the purpose of regulating large amount of power, need to be designed specifically and they are called power transistors [4]. MOS transistors can conduct large currents at very low drain-to-source voltages, while operating in linear region. This fact reveals the resistive behavior of the MOS in that region, having a resistance of $R_{DS(on)}$, which is related to its geometry as,

$$R_{DS(on)} \cong \frac{1}{\beta(V_{GS} - V_t)} \qquad (1)$$

Here, $\beta \alpha g_m (W/L)$.

Hence, the on-resistance $R_{DS(on)}$ varies inversely with the device's W/L ratio, trans-conductance and the effective gate to source voltage. In order to increase the efficiency of the converter, one of the requirements is to reduce the I^2R loss of the switches and for that purpose the total resistance of the switch needs to be reduced. Moreover, for the efficient transformation of magnetic energy in the switching regulators, we want to store all the volt-second across the inductor in charge phase [4]. Practically, it's not possible to have such condition, because of the voltage drop across the switch. Hence, the resistance of the switch should be minimized as much as it's possible. Though theoretically the on-resistance can be reduced arbitrarily to any small amount by increasing W/L ratio, in reality considerations such as die size and cost, parasitic metallization resistance, orientation of the device, bondwire resistance place limitations upon $R_{DS(on)}$. Discrete MOS transistors can be available with on-resistances of only few milliohms, but when the integrated power devices need to fit in small areas, satisfying some conditions assigned by foundries, their resistances vary from 25mΩ to 1Ω [5]. Moreover the metallization resistance becomes significant for on resistances of less than an Ohm and equation (1) for $R_{DS(on)}$ then becomes,

$$R_{DS(on)} \cong \frac{1}{\beta(V_{GS} - V_t)} + R_M \qquad (2)$$

Where, R_M is the sum of the resistance of the source and drain metallization. This metallization resistance depends on the geometry and orientation of the transistor and is difficult to calculate. Moreover, for the better predictability of the behavior of the power MOS switches, it's expected to make the contribution of the metal resistance less to the overall resistance. In section 2, a mathematical model for power switches, addressing several design issues, has been developed. Section 3 deals with the simulated and measured data. Finally, concluding remarks of the topology have been included in section 4.

II. Topological Analyses

The total resistance includes channel resistance, parasitic resistances of metal layers, metal to metal contact, metal to diffusion contact and diffusion resistance. However, this reduction, for a given silicon area, depends on some parameters like the aspect ratio of the switches,

orientation of the switches, the pitch of the segments, the number of metal layers, the number of contacts among different metal layers or between metal layer and diffusion etc. We want to have the total resistance dominated by the channel resistance, as it is the more

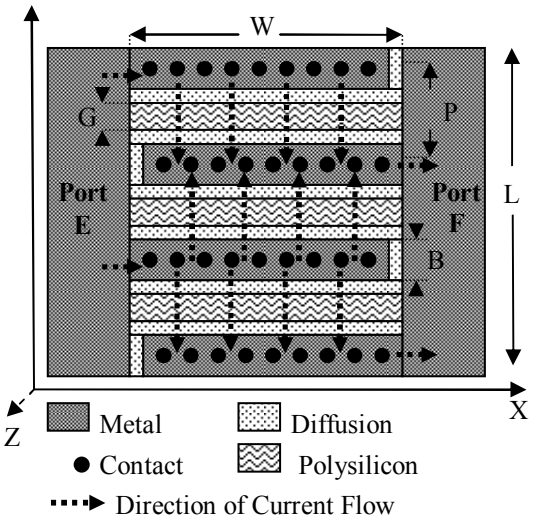


Fig. 1 A sample segment of the switch. The direction of the current flow has been illustrated, as well.

predictable part than any other sorts of resistance of the switch. Referring to Fig 1, a segment of switch with only 4 layers has been illustrated. Here metal layer (metal 1) is the top most layers, and then contact, then polysilicon and diffusion lies in the bottom most position. The segment consists of four fingers of breadth B, two of which are connected to Port E and the remaining two connected to Port F. At first, the current enters the two fingers connected to port E, through the metal layer in the X direction. It then flows downward in the Z direction through the contacts to the diffusion and reaching the diffusion area it then flows in the Y direction through the channels in that plane, underneath the poysilicon gates, comes upward through the contacts to the metal layer of the fingers connected to Port F. Hence, for the channel resistance we've,

$$R_{ch} \propto \frac{G}{W}$$

As W of each finger is increased or gate length L is decreased or both, the channel resistance is decreased and vice-versa. Again, for the parasitic resistance of metal and diffusion, we've,

$$R_{ch} \propto \frac{W}{B}$$

Hence, with W the parasitic resistance of the switch increases. Hence, to have total resistance minimized and the contribution of channel resistance to the total resistance, maximized, an optimization is needed. As shown in the Fig. 1, this is a fingered design where current is fed from one side and collected from the other side. Hence, with W the parasitic resistance of the switch increases. Hence, to have total resistance minimized and the contribution of channel resistance to the total resistance, maximized, an optimization is needed. As

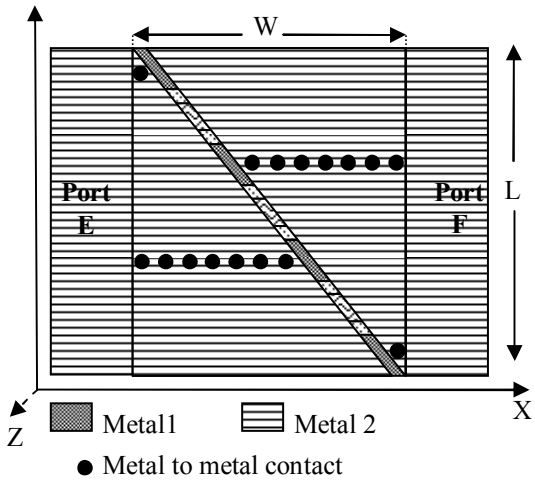


Fig. 2 The segment of the switch, illustrated with the top most metal.

shown in the Fig. 1, this is a fingered design where current is fed from one side and collected from the other side. Every alternate finger is a drain and there is a source finger in the middle of each drain finger. Between each consecutive drain and source the active area is defined by a gate strip under which the channel forms.So each source is fed by two consecutive drains. In the resistance calculation this comes into effect. The alternate drains must be fed with current and when the currents from each segment reach the source side, they need to be collected. In Fig. 2, the placement of metal 2 has been shown. It's laid in diagonal form; the vias are placed in one diagonal part in such a way that all the drains are fed. The other diagonal part is connected to the source fingers underneath. Another way to stack several metal layers in parallel is to place similar metal strips of different layers of metal and connect those by vias [5], but in this method top metal layer has larger conductive area and thus provide effectively lower resistance. The calculation for such metal piece is based on sheet resistance basis which has many approximations. But owing to the fact that its contribution is low as compared to the total resistance, such approximations are still good for practical purpose and as will be showed later that they are actually in good match with measured data. The diagonal metal 2 placement has another advantage that the pad may be placed very close to the pickup at the wider edge making it easy for ESD placement and high current routing at both drain and source side. The objectives are to design a switch with die area so that optimum range of ON resistance can be attained with a reasonable percentage of channel resistance or vice-versa. For this first we need to set a target range of ON resistance value. Then we start with an initial value of area for iteration. The optimum aspect ratio will also be determined so that total resistance is optimized for that area. Before proceeding further, let's introduce a useful parameter called pitch. The pitch of a fingered MOS is defined by the distance of midpoint of a drain finger to the midpoint of the source finger, as illustrated by P in Fig. 1. This pitch is not only limited by the gate length but also by the metal finger width, metal to metal minimum distance and required

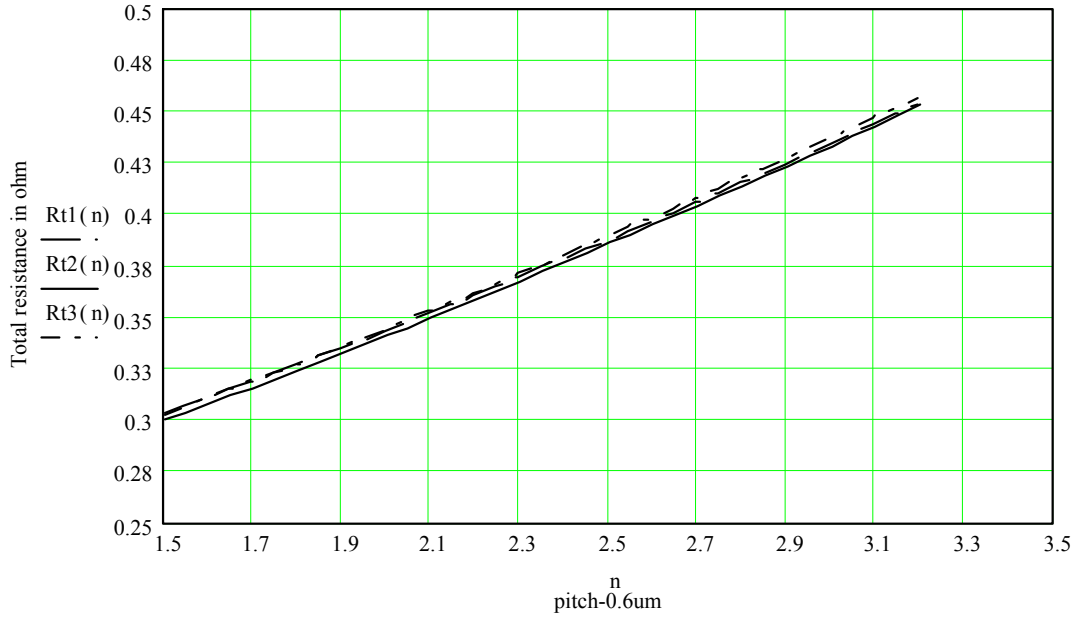


Fig. 3 Variation of total resistance (in Ω) with the pitch (in µm) for different aspect ratios.

DRC spacing of diffusion to metal contact from the polysilicon gate. Metal width is kept such that it meets not only the requirements of the foundry regarding its minimum width and clearance from contact at each edges but also the safe current carrying limit of the metal itself. Hence, pitch= metal (here, metal 1) width + greater length of the required metal to metal spacing or metal to diffusion contact spacing from polysilicon gate set by DRC. The polysilicon gate lies within the pitch. Because of this pitch, another parameter called the diffusion resistance from the channel to contact comes into consideration. For any other MOS it is insignificant. But for power switch the contribution of this resistance is taken into account for better accuracy in calculation. Aspect Ratio AR is the ratio of length to width, and area A is the multiplication of length with width, hence from Fig. 1

$$W = \sqrt{\frac{A}{AR}} \tag{3}$$

$$L = \frac{A}{W} \tag{4}$$

The number of the fingers N, can be calculated as,

$$N = \frac{L}{P} \tag{5}$$

Once the number of the fingers is determined, we've,

$$\beta = \mu_n C_{ox} \frac{N \times W}{G} \tag{6}$$

it's to be noted that N×W stands for the effective gate width. The channel resistance can be calculated from equations (1) and (6), which becomes;

$$R_{ch} = \frac{G}{(\mu_n C_{ox}) \times N \times W \times (V_{GS} - V_t)} \tag{7}$$

In Fig. 1, metal 1 strips are placed along the fingers and if ρ_1 is the resistivity of metal 1 in Ω/□, following the direction of current flow, we've

$$R_{M1} = \frac{W \times \rho_1}{B \times N} \tag{8}$$

The diffusion region lies in the pitch except the polysilicon gate region. If ρ_{diff} is the resistivity of diffusion in Ω/□, the total diffusion resistance R_{diff} is defined as,

$$R_{diff} = \frac{\rho_{diff} \times (P - G)}{N \times W} \tag{9}$$

The contact total contact resistance depends on the number of the contacts put in parallel in each finger and the number of fingers. If R_c is the resistance of a contact and N_c is the number of contacts per finger, we've the contact resistance R_{cont} as,

$$R_{cont} = \frac{R_c}{N_c \times N} \tag{10}$$

Metal 2, as shown in Fig. 2, is placed in the diagonal manner such that the two corners have the lowest number of contacts and this number increases gradually as we traverse along L. Moreover, if we consider the direction of current in metal 2, we'll find that, current goes down to lower layers via contacts and passes beneath each of the gates and ultimately gets collected on the other diagonal portion. However, in case of diagonal placement, the number of contacts is larger in the wider part than that in the narrower part. In Fig. 2, we've more contacts at the source side to pick up current than those in the drain side (at the top end). This phenomenon is the opposite at the bottom side, where there is less contact for pickup in the narrower part of source side compared to those in the drain side, from where current is injected. At the middle point of the power switch we have almost equal number of contacts in the drain and in the source side. Current density will therefore be higher in the middle part than the top and bottom edge. This inequality in current distribution effectively increases the metal 2 resistance than that calculated by multiplying the number of squares

with the sheet resistance it. With a good degree of accuracy, this can be taken as an increase of factor 2, from the original calculated resistance. Now it's to be noted that current flows along the entire L and thus metal 2 will have a effective number of square of (L/W). If ρ_2 is the resistivity of metal 2 in Ω/□, we can write the total metal 2 resistance R_{M2} as,

$$R_{M2} = \frac{\rho_2 \times L \times 2}{W} \qquad (11)$$

Another resistance is the bond wire resistance R_B and it comes from the bond wires which connect the pad of the chip to the post of the pin. When we consider the flow of current from one side to another (drain to source), all the resistance comes in series. Hence, the total resistance R_t comes to be as,

$$R_t = R_{ch} + R_{M1} + R_{diff} + R_{cont} + R_{M2} + R_B \qquad (12)$$

Now if the equation (12) is plotted as a function of pitch for different aspect ratios, we'll have the curve as shown in Fig.3. It's evident from Fig. 3 that for aspect ratio of 2:1 we've the optimum result, i.e. minimum total resistance. It's to be mentioned that, for lower pitch we'll have lower resistance, but it's not always possible to select the lower one, as the metal width of each segment needs to have the required current carrying capability. Thus for a given die area, the total resistance can be optimized. The reverse is also true, i.e. if we want to have a particular range of total resistance, the chip area and aspect ratio required for attaining that resistance can be calculated.

III. Modeling of the Switch

At first, calculations were performed for a specific die area of 58900 μm^2 and aspect ratio of 2:1, using MathCAD. Following the procedures, mentioned in the previous section, the design parameters are found to be as W (width of a finger) =172 µm, l (length of a gate) =0.7µm and N (number of fingers/MOS) = 113. The optimum pitch is found to be 2.9 µm. The resistance of each bond wire is almost 50mΩ and two bond wires are connected in series with the switch, one from the pin SW to the drain side of the switch and the other from the pin GND to the source side of it. Hence almost 100mΩ more will be added to the total resistance. For the purpose of simulation, the power switch is modeled as shown in Fig. 4. Here, the power switch has been modeled using an NMOS named NSW. The parasitic resistances found for different components are 53 mΩ from metal1, 60 mΩ from metal2, 15mΩ from diffusion, 18mΩ from contact and 100 mΩ from bond wires. The total parasitic resistance is limped and divided equally between the drain and source sides of the power switch. The other MOS NCS is kept for current sensing purpose.

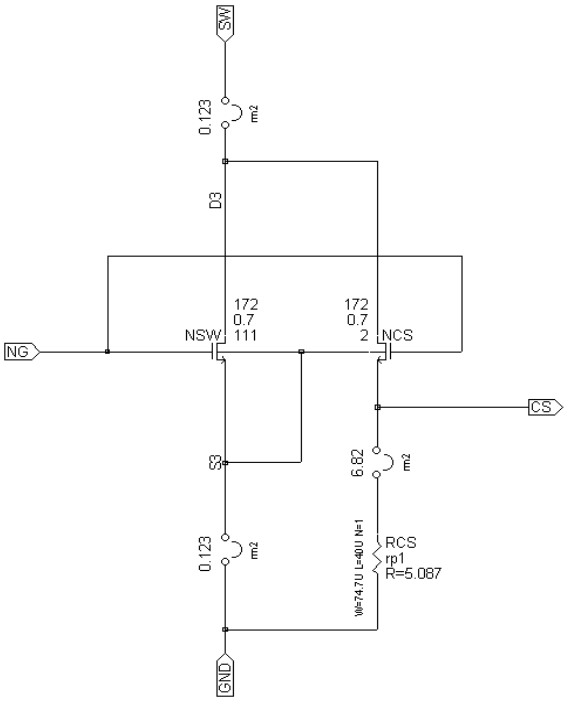


Fig. 4 Model of the NMOS power switch for simulation, with parasitic resistance added as lumped parameters on both the sides.

IV. Simulation and Test Results

The total resistance has been plotted as a function of effective voltage i.e. $(V_{GS} - V_t)$ as shown in Fig. 5. It's to be mentioned here that in the calculation of the channel resistance the value of $\mu_n C_{ox}$ has been taken as constant. But the mobility degradation due to higher values of V_{GS} causes the effective $\mu_n C_{ox}$ to decrease at those values, as the effective mobility μ_{eff} is related the applied V_{GS} as [6],

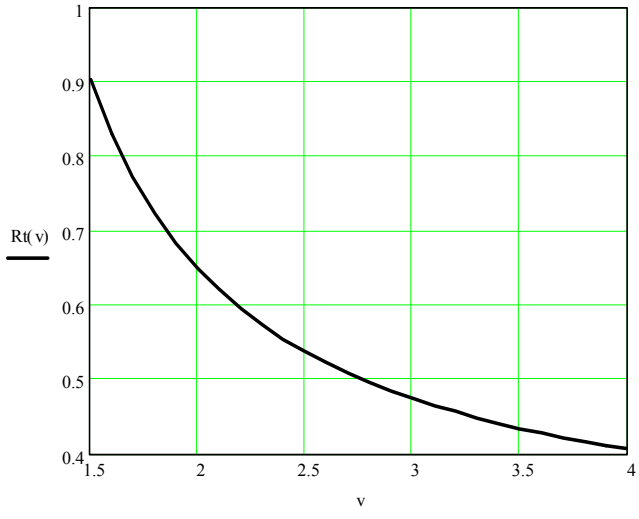


Fig. 5 The variation of total resistance as a function of effective voltage $(V_{GS} - V_t)$ for $\mu_n C_{ox}$ =50µA/V^2

$$\mu_{eff} = \left[\frac{\mu_n}{1 + \theta(V_{GS} - V_t)}\right]$$

Hence, $\mu_n C_{ox}$ is higher in the lower values of V_{GS} and this makes the channel resistance, which is the most significant part of the total resistance, lower than that predicted by equation (7). This fact is further illustrated in Fig. 6, the I_D vs. V_{DS} curves of the power NMOS of W=172µm, l=0.7µm and N=113 (parasitic resistances not

included in this case) plotted for different V_{GS}. For the lower range of V_{GS}, a small change of V_{DS} causes a huge flow of drain current I_D and makes the power MOS to enter the saturation region quickly.

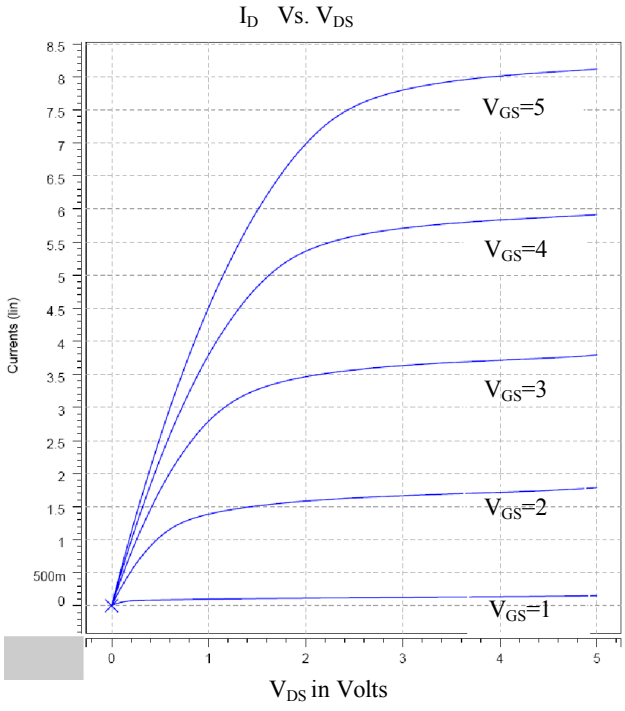


Fig. 6 I_D vs. V_{DS} curves for different V_{GS} of the power MOS of size W=172µm, l=0.7µm, N=113

Thus the total resistance plotted in Fig. 5 will deviate from the actual value in the lower range of V_{GS}. If the total resistances are recalculated using higher value of $\mu_n C_{ox}$ using the same procedure, we've the plot as in Fig 7, showing smaller total resistance than that in Fig. 5, in the lower V_{GS} range.

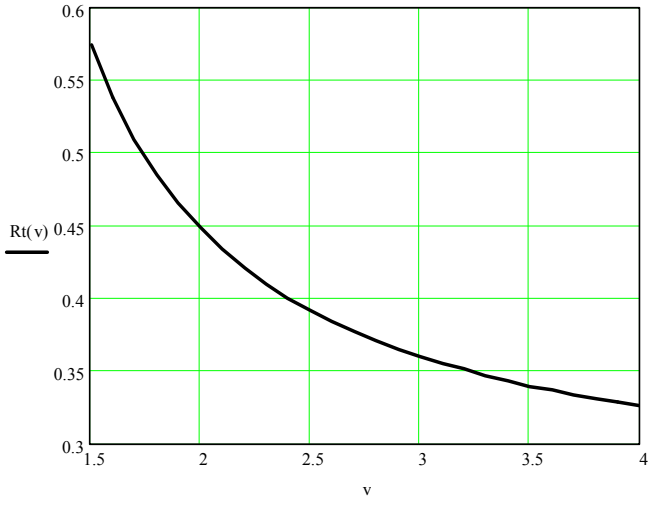


Fig. 7 The variation of total resistance as a function of effective voltage ($V_{GS} - V_t$) for $\mu_n C_{ox}$ =100µA/V^2

The power switch is simulated using the model shown in Fig. 4, by HSPICE. The NMOS switch has been fabricated in 0.5-µm dual metal, dual poly process and the chip micrograph has been shown in Fig. 8.

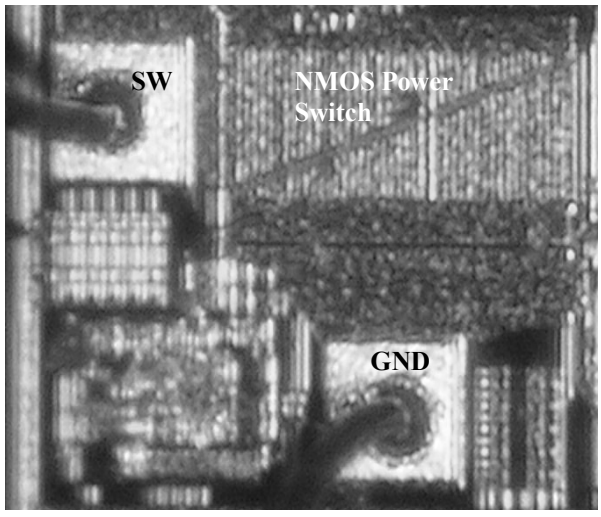


Fig. 8 The chip micrograph of NMOS power switch, placed in between SW and GND pads.

The chip sample is then tested and the simulation and test data have been plotted in Fig. 9.

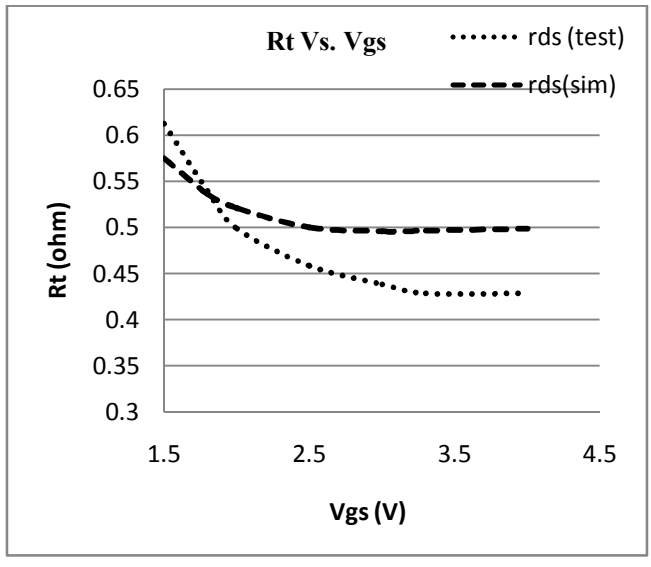


Fig. 9 The simulation and test data of the total resistance as a function of V_{GS}.

Hence it's evident from the simulation and test data that total resistance values resemble those of Fig. 7 for the lower range of V_{GS}. For higher values of V_{GS}, these data correlate to those in Fig. 5.

V. Conclusion

In this paper, a mathematical model of an integrated switch has been presented. While doing so, various issues coming from the geometrical aspects and parasitic

components have been addressed. The simulation and test data show a good correlation with the developed model.

Acknowledgements

The authors are grateful to the top management of POWER IC LTD., the first semiconductor company of Bangladesh, for their support in developing the chip.

References

[1]. *MAXIM Application Note 2031*, Oct19, 2000, http://www.maxim-ic.com.

[2]. J. M. Chang and M. Pedram, ‘‘Energy minimization using multiple supply voltages,’’ *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*,vol. 5, no. 4, pp. 436–443, Dec. 1997.

[3]. H.P Le, C. S. Chae, K.C. Lee, S. W. Wang, G.H. Cho, and G. H. Cho, ‘‘A Single-Inductor Switching DC–DC Converter With Five Outputs and Ordered Power-Distributive Control,’’ *IEEE JSSC,* vol. 42, no. 12, Dec. 2007.

[4]. A. I. Pressman, ‘‘Switching Power Supply Design’’, McGraw-Hill Professional, 2nd Edition.

[5]. A. Hastings, ‘‘The Art of Analog Layout’’, *Pearson Education International*, 2nd Edition, 2006.

[6]. P. R. Gray, P. J. Hurst, S. H. Lewis and R. G. Meyer, ‘‘ Analysis and Design of Analog Integrated Circuits’’, John Wiley & Sons, Inc., 4th Edition, 2001.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Analysis of an OFDM System in the Presence of Carrier Frequency Offset, Phase Noise and Timing Jitter over Rayleigh Fading Channels

Shankhanaad Mallick, *Member IEEE* and Satya Prasad Majumder, *Member IEEE*

Department of Electrical & Electronic Engineering,
Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
Email: shankhanaadmallick@hotmail.com , spmajumder@eee.buet.ac.bd

Abstract – A theoretical analysis for evaluating the performance of an Orthogonal Frequency Division Multiplexing (OFDM) under the combined influence of Carrier Frequency Offset (CFO), phase noise and timing jitter over rayleigh fading channels is presented. An exact closed form expression for the Signal-to-Interference plus Noise Ratio (SINR) is derived and the combined effects of these synchronization impairments are exhibited by the Bit Error Rate (BER) performances of a BPSK-OFDM system over rayleigh fading channels. Results show that OFDM system suffers significant SINR penalty due to CFO and jitter, however, the effect of phase noise is the dominant one.

I. Introduction

Being very efficient in combating multipath fading as well as Inter Symbol Interference (ISI) and in the use of available bandwidth, Orthogonal Frequency Division Multiplexing (OFDM) has been widely adopted and implemented in wire and wireless communications, such as Digital Subscriber Line (DSL), European Digital Audio Broadcasting (DAB), Digital Video Broadcasting-Terrestrial (DVB-T) and its handheld version DVB-H, and IEEE 802.11a/g standards for Wireless Local Area Networks (WLANs) [1]-[2] etc.

Unfortunately OFDM is very much sensitive to the synchronization errors such as Carrier Frequency Offset (CFO), phase noise or timing jitter [3]. The CFO arises mainly due to the Doppler shifts introduced by the channel which causes frequency difference between the transmitter and receiver oscillators. The deleterious effects caused by the CFO are the reduction of the signal amplitude and introduction of Inter-Carrier-Interference (ICI) from the other carriers which are then no longer orthogonal to the filter [4]. Phase noise results from the imperfections of the Local Oscillators (LO) used for the conversion of a baseband signal to a passband (or vise-versa). Phase noise has two effects on an OFDM system: rotation of the symbols over all subcarriers by a Common Phase Error (CPE) and the occurrence of ICI which introduces a blurring of the constellation like thermal noise [5]. Timing errors would occur either when the clock signal is not correctly recovered, or when sampling is not performed at precise sampling instants. Because of the non-ideal nature of the sampling circuit the amplitude of the signal is affected by timing jitter and it introduces additional source of additive noise [6].

The individual effects of CFO and phase noise have been analyzed by several authors and the degradation introduced in the system has been characterized for some particular cases in [3]-[5], [7]-[11]. The effect of timing jitter on the performance of discrete multitone system was also investigated in [12] and in [13]. However, a closed form analytical result that shows the exact quantitative effect of the combination of these three impairments even for Additive White Gaussian Noise (AWGN) channels has not been well addressed.

The purpose of this paper is to analyze, via mainly an analytical approach, the impact of the combined effects of CFO, phase noise and timing jitter to the performance of OFDM systems in rayleigh fading environment. The exact Signal to Interference plus Noise Ratio (SINR) expression in a closed form is derived which provides a quantitative understanding of how system behavior changes with certain parameters. We evaluate the Bit Error Rate (BER) performances of a BPSK-OFDM system over rayleigh fading channels considering the combined influence of these synchronization impairments.

The rest of the paper is organized as follows: In Section II, CFO, along with phase noise and timing jitter process is reviewed and the OFDM system model is given in the presence of CFO, phase noise and timing jitter over rayleigh fading channel. In Section III the exact SINR expression for the combined effects is derived. Section IV gives the results of system performance analysis and finally Section V finishes the paper by giving conclusion.

II. System Model and Description

Consider the m^{th} symbol of an N -subcarrier OFDM system in the presence of normalized CFO, ε , phase noise, $\varphi_m(n)$, and timing jitter, ξ_n as shown in Fig. 1.

A. Carrier Frequency Offset (CFO) Model

The absolute value of the actual CFO f_ε , is either an integer multiple or a fraction of Δf , or the sum of them.

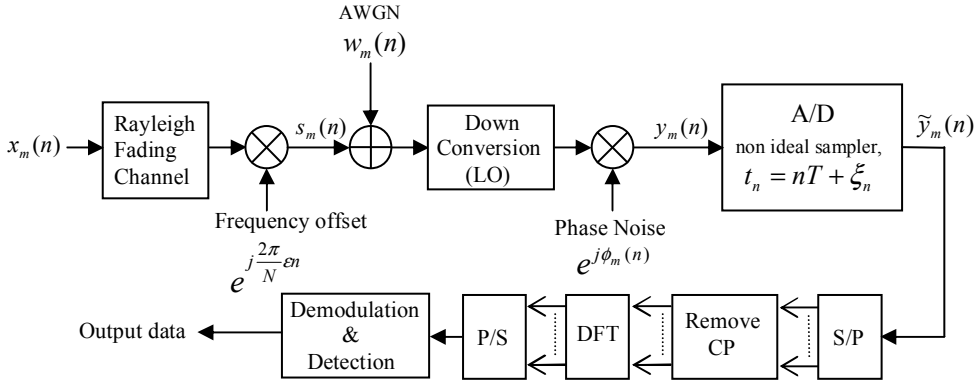


Fig. 1 OFDM system model (receiver) in the presence of CFO, phase noise and timing jitter over rayleigh fading channel

If f_ε is normalized to the subcarrier spacing Δf, then the resulting normalized CFO of the channel can be generally expressed as

$$\varepsilon = \frac{f_\varepsilon}{\Delta f} = \delta + \in \tag{1}$$

where δ is an integer and $|\in| \le 0.5$ .The influence of an integer CFO on OFDM system is different from the influence of a fractional CFO. In the event that $\delta \neq 0$ and $\varepsilon = 0$, symbols transmitted on a certain subcarrier, e.g., subcarrier k , will shift to another subcarrier k_δ , $k_\delta = k + \delta \bmod N - 1$ (2)

As we focus on the ICI effect, we will consider normalized CFO, $\varepsilon = \frac{f_\varepsilon}{\Delta f} = \in$ since no ICI is caused by an integer CFO. We assume relative CFO (ε) to be a Gaussian process, statistically independent of the input signal, with zero mean and variance σ_ε^2 .

B. Phase Noise Model

Phase noise $\varphi_m(n)$, generated at both transmitter and receiver oscillators, can be modeled as [5]

$$\begin{aligned}\varphi_m(n) &= \varphi_{m-1}(N-1) + \sum_{i=-N_g}^{n} u[m(N+N_g)+i] \\ &= \sum_{i=0}^{m(N+N_g)+N_g+n} u(i) = C_m + \sum_{i=0}^{n} u(T_m + i)\end{aligned} \tag{3}$$

where C_m and T_m are defined by $\sum_{i=0}^{m(N+N_g)+N_g-1} u(i)$ and $m(N+N_g)+N_g$ respectively. N_g is the length of cyclic prefix and $u(i)'s$ denote mutually independent Gaussian random variables having zero mean and variance $\sigma_u^2 = 2\pi\beta T / N = 2\pi\beta / R$, where β denotes the two-sided 3-dB linewidth of the Lorentzian power density spectrum of the free running carrier generator [8], T and R denote OFDM symbol period and the transmission data rate, respectively.

C. Timing Jitter Model

In the sampling circuit at the receiver additional error may occur in the determination of the best sampling phase. This means that the sampling instants are non-ideal and is given by [12]

$$t_n = nT + \xi_n \tag{4}$$

where ξ_n is the timing jitter of the n^{th} sampling instant normalized by symbol duration T . Timing jitter can be modeled as a stationary Gaussian random process statistically independent of the input signal with zero mean and variance σ_ξ^2 [6].

D. OFDM System Model

As shown in Fig. 1, the transmitted OFDM signal for the m^{th} symbol is given by the N point complex modulation sequence

$$x_m(n) = \sum_{k=0}^{N-1} X_m(k) e^{j\frac{2\pi}{N}nk} \tag{5}$$

where n ranges from 0 to $N + N_g - 1$.

After passing through a rayleigh fading channel and LO, the received signal impaired by AWGN and phase noise can be modeled as

$$y_m(n) = [\sum_{k=0}^{N-1} X_m(k) H_m(k) e^{j\frac{2\pi}{N}n(k+\varepsilon)}] e^{j\varphi_m(n)} + w_m(n) \tag{6}$$

or, $y_m(n) = s_m(n) e^{j\varphi_m(n)} + w_m(n)$ (7)

where, $s_m(n) = \sum_{k=0}^{N-1} X_m(k) H_m(k) e^{j\frac{2\pi}{N}n(k+\varepsilon)}$

In (6) $H_m(k)$ is the transfer function of the rayleigh fading channel at the frequency of the k^{th} carrier and $w_m(n)$ is the complex envelope of AWGN with zero mean and variance σ^2 .

Assuming $\varphi_m(n)$ is small [9] so that

$$e^{j\varphi_m(n)} \approx 1 + j\varphi_m(n) \tag{8}$$

Substituting (8) into (7) yields

$$y_m(n) = s_m(n) + s_m(n) j\varphi_m(n) + w_m(n) \tag{9}$$

After DFT and by dropping the subscript 'm' (9) yields

$$Y(k) = S(k) + S(k) \otimes j\Theta(k) + W(k) \tag{10}$$

where, $S(k)$, $\Theta(k)$ and $W(k)$ are the DFT responses of $s_m(n)$, $\varphi_m(n)$, and $w_m(n)$ respectively and $\otimes$ denotes the circular convolution operation.

III. Exact SINR Expression for the Combined Effect

Let, $S(k) = S_1(k) + I_1(k)$

where

$$\begin{aligned} S_1(k) &= \frac{1}{N} X(k)H(k) \sum_{n=0}^{N-1} e^{\frac{j2\pi n\varepsilon}{N}} \\ &= X(k)H(k) \frac{\sin(\pi\varepsilon)}{N\sin(\frac{\pi\varepsilon}{N})} e^{j\pi\varepsilon(\frac{N-1}{N})} \end{aligned} \tag{11}$$

The first component, $S_1(k)$, is the modulation value $X(k)$ modified by the channel transfer function. It experiences an amplitude reduction and phase shift due to CFO,(ε). As N is always much greater than ($\pi\varepsilon$), $N\sin(\frac{\pi\varepsilon}{N})$ is replaced by ($\pi\varepsilon$).

The second term, $I_1(k)$, is the ICI caused only by the CFO and is given by

$$\begin{aligned} I_1(k) &= \frac{1}{N} \sum_{\substack{r=0 \\ r\neq k}}^{N-1} X(r)H(r) \sum_{n=0}^{N-1} e^{\frac{j2\pi n\varepsilon}{N}} \\ &= \sum_{\substack{r=0 \\ r\neq k}}^{N-1} X(r)H(r) \frac{\sin(\pi\varepsilon)}{N\sin(\frac{\pi(r-k+\varepsilon)}{N})} e^{j\pi\varepsilon(\frac{N-1}{N})} e^{-j\pi(\frac{r-k}{N})} \end{aligned} \tag{12}$$

Assuming $E[X_k] = 0$ and $E[X_k X_r^*] = |X|^2 \delta_{rk}$ and average channel gain $E\{|H_r|^2\} = |H|^2$ we can obtain from (11)

$$E[|S_1(k)|^2] = |X|^2 |H|^2 sinc^2(\pi\varepsilon) \tag{13}$$

$E[|I_1(k)|^2]$ is evaluated in [4] as

$$E[|I_1(k)|^2] \le 0.5947 |X|^2 |H|^2 (\sin\pi\varepsilon)^2; \quad |\varepsilon| \le 0.5 \tag{14}$$

Let, $I_2(k) = S(k) \otimes j\Theta(k)$ (15)

In $I_2(k)$, the term that causes ICI due to the joint effect of phase noise and CFO is given by

$$\begin{aligned} I_2'(k) &= (\frac{j}{N} \sum_{\substack{r=0 \\ r\neq k}}^{N-1} X(r)H(r) \sum_{n=0}^{N-1} e^{\frac{j2\pi n\varepsilon}{N}}) \Theta(k-r) \\ &= j \sum_{\substack{r=0 \\ r\neq k}}^{N-1} X(r)H(r) \frac{\sin(\pi\varepsilon)}{N\sin(\frac{\pi\varepsilon}{N})} \Theta(k-r) \end{aligned} \tag{16}$$

In (Appendix) the energy of $\Theta(r)$ is given by

$$E[|\Theta(r)|^2] = \frac{\sigma_u^2}{2N\sin^2(\frac{\pi r}{N})} \tag{17}$$

Thus $E[|I_2'(k)|^2]$ can be evaluated as

$$E[|I_2'(k)|^2] = |X|^2 |H|^2 sinc^2(\pi\varepsilon) \frac{\sigma_u^2}{2N\sum_{r=1}^{N-1} \sin^2(\frac{\pi r}{N})} \tag{18}$$

As channel SNR γ_{ch} is defined by, $\gamma_{ch} = \frac{|X|^2|H|^2}{E[|W(k)|^2]} = \frac{E_{in}\alpha^2}{N_0} = \gamma_{in}\alpha^2$, where $\gamma_{in} = \frac{E_{in}}{N_0}$ is the input SNR, E_{in} is the averaged transmitted energy of the individual carriers, $\frac{N_0}{2}$ is the power spectral density of the AWGN in the fading transmission channel and α is the rayleigh fading channel attenuation/gain parameter. Therefore the SINR expression in the presence of CFO and phase noise in rayleigh fading environment may be expressed as

$$SINR(\varepsilon, \sigma_u^2, \alpha) \ge \frac{\gamma_{in}\alpha^2 \{sinc^2(\pi\varepsilon)\}}{1 + \gamma_{in}\alpha^2 [0.5947(\sin\pi\varepsilon)^2 + \{\frac{\sigma_u^2}{2N} sinc^2(\pi\varepsilon) \sum_{r=1}^{N-1} \frac{1}{\sin^2(\frac{\pi r}{N})}\}]} ; |\varepsilon| \le 0.5 \tag{19}$$

In the non-ideal sampling circuit, the amplitude of the samples is affected by a random timing jitter (ξ) which ultimately causes E_{in} to degrade by a factor of $(1-\xi)$ over a time slot and on the other hand it increases additive noise energy by $E_{in}\xi$. Considering the effect of jitter the SINR expression of (19) is modified as follows

$$SINR(\varepsilon, \sigma_u^2, \xi, \alpha) \ge \frac{\gamma_{in}\alpha^2(1-\xi)\{sinc^2(\pi\varepsilon)\}}{1 + \gamma_{in}\alpha^2(1-\xi)[0.5947(\sin\pi\varepsilon)^2 + \{\frac{\sigma_u^2}{2N} sinc^2(\pi\varepsilon) \sum_{r=1}^{N-1} \frac{1}{\sin^2(\frac{\pi r}{N})}\}] + \gamma_{in}\alpha^2\xi} ; \quad |\varepsilon| \le 0.5, \quad |\xi| \le 1 \tag{20}$$

(20) indicates that, in the presence of CFO, phase noise, timing jitter and rayleigh fading, several parameters affect OFDM system performance, resulting in severe performance degradation which is unacceptable in practice. In the absence of CFO ($\varepsilon = 0$) the SINR expression of (20) reduces to

$$SINR(\sigma_u^2,\xi,\alpha) \geq \frac{\gamma_{in}\alpha^2(1-\xi)}{1+\gamma_{in}\alpha^2(1-\xi)[\{\frac{\sigma_u^2}{2N}\sum_{r=1}^{N-1}\frac{1}{\sin^2(\frac{\pi r}{N})}\}]+\gamma_{in}\alpha^2\xi} \quad ;|\xi| \leq 1 \tag{21}$$

which reflects the combined influences of phase noise and timing jitter in rayleigh fading environment. In the case of perfect carrier-phase synchronization, $\Theta(r)$ becomes a Dirac delta function and σ_u^2 approaches zero, while the SINR expression of (21) reduces to

$$SINR(\xi,\alpha) \geq \frac{\gamma_{in}\alpha^2(1-\xi)}{1+\gamma_{in}\alpha^2\xi} \quad ;|\xi| \leq 1 \tag{22}$$

Finally for an ideal sampling circuit ($\xi = 0$) and for a non-fading environment ($\alpha = 1$) SINR expression of (22) becomes input SNR, γ_{in}.

From the SINR expression of (20), the conditional probability of bit error, $P_b(e|\varepsilon,\xi,\alpha)$ conditioned on a given value of ε , ξ and α , for a given input SNR, γ_{in} and phase noise variance, σ_u^2 can be obtained as [14]

$$P_b(e|\varepsilon,\xi,\alpha) = \frac{1}{2}erfc(\sqrt{SINR(\varepsilon,\xi,\alpha)}) \tag{23}$$

Then the average BER for BPSK-OFDM system over rayleigh fading channels can be evaluated as

$$BER = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_b(e|\varepsilon,\xi,\alpha)P_\varepsilon(\varepsilon)P_\xi(\xi)P_\alpha(\alpha)d\varepsilon\, d\xi d\alpha \tag{24}$$

here the probability density functions (PDFs) $P_\varepsilon(\varepsilon)$ and $P_\xi(\xi)$ are assumed Gaussian whereas the PDF of α , $P_\alpha(\alpha)$ is rayleigh.

IV. System Performance Analysis

In the presence of CFO, phase noise and timing jitter over fading channels, (24) and (20) indicate that *BER* and *SINR* are functions of these three impairments as well as some critical system parameters. These relations are depicted respectively in Figs. 2-7. If not mentioned in the figures the following parameters are used for computation in this section:

Table 1: System and Channel Parameters

No. of Subcarriers (N)	2048
Cyclic Prefix Length (N_g)	128
Modulation	BPSK
Data Rate(R)	64/7 MHz
Channel type	Rayleigh fading
Fading variance of channel	0.33
Input SNR	20 dB

Fig. 2 illustrates the catastrophic effect of CFO, phase noise and timing jitter on the BER performance of a BPSK-OFDM system over rayleigh fading channel. As shown, the performance degrades with the increase of the combinational variances of $\sigma_\varepsilon^2, \sigma_u^2$ and σ_ξ^2. At high SNRs there exists BER floors resulting from the combined ICI effects lowering the effective SNRs, and that BER floor runs up when variance level increases. It also implies that OFDM systems with high SNR are more sensitive to ICI, though higher SNR leads to better performance.

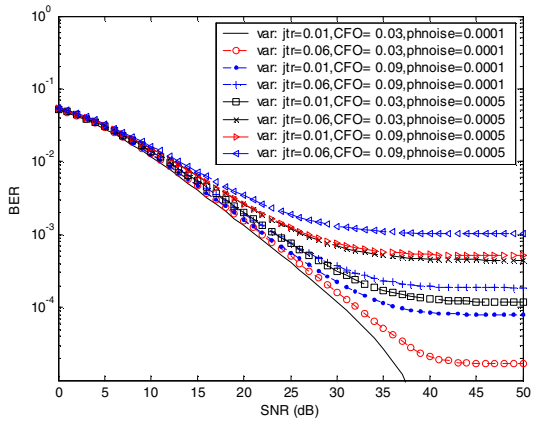


Fig. 2 BER performance of an OFDM system over rayleigh fading channel for different combination of variances (var) of CFO, phase noise (phnoise) and jitter (jtr).

Fig. 3, 4 & 5 demonstrates the SINR penalty suffered by the system as a function of $\sigma_\varepsilon^2, \sigma_\xi^2$ and σ_u^2 respectively for 20dB input SNR.

In Fig. 3, first of all, we consider the variance of CFO (σ_ε^2) as variable and σ_ξ^2 , σ_u^2 as constants. When the variances of jitter and phase noise are zero, ICI occurs only due to CFO. With the increase of σ_ε^2 , SINR penalty

increases and we can see that more than 7dB penalty is suffered by the system for $\sigma_\varepsilon^2 = 0.1$. If either or both jitter and phase noise are present along with fading then the system suffers a penalty even at $\sigma_\varepsilon^2 \approx 0$. The SINR penalty curve shifts upwards with the increase of the constant variances of jitter and phase noise.

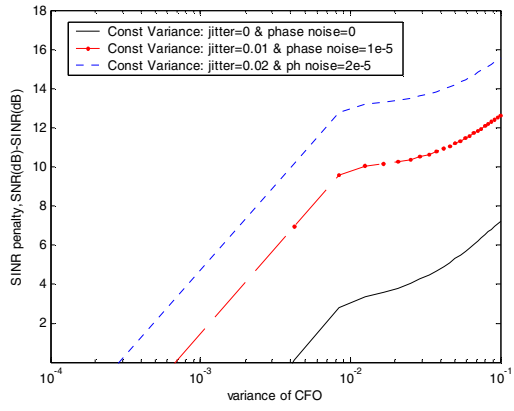


Fig. 3 SINR penalty as a function of the variance of CFO for different Constant Variances of jitter and phase noise.

Similar curves are obtained in Fig. 4, for considering σ_ξ^2 as variable and $\sigma_\varepsilon^2, \sigma_u^2$ as constants. From Fig. 3 & 4, we can see that nearly 1-2dB additional penalty is suffered by the system for jitter than CFO when their respective constant variances have relatively low values.

In Fig. 5 we see that SINR penalty drastically changes for a very small change in the variances of phase noise. It is also noticeable that SINR is strongly dependent on the number of sub-carriers (N) as ICI due to phase noise is a function of N. Larger number of N leads to shorter subcarrier spacing distance, hence more sensitive to phase noise and as a result of that SINR penalty increases. From Fig 3, 4 & 5 we can conclude that OFDM is more sensitive to phase noise than CFO and jitter.

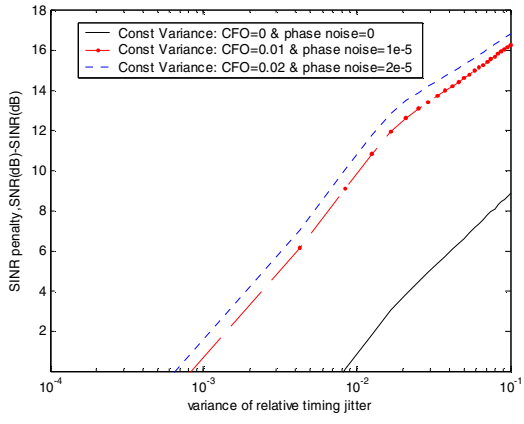


Fig. 4 SINR penalty as a function of the variance of relative timing jitter for different Constant Variances of CFO and phase noise.

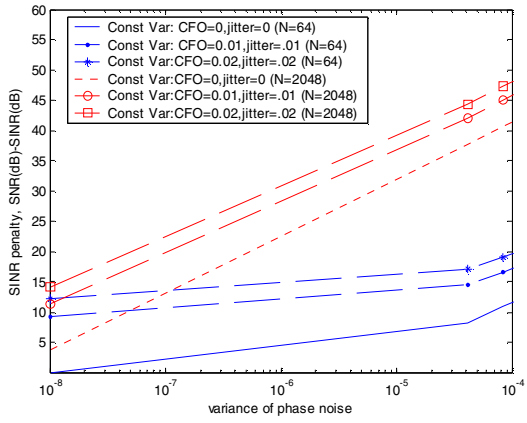


Fig. 5 SINR penalty as a function of the variance of phase noise for different constant variances of CFO and jitter with $N = 64 \& 2048$.

For low variances of CFO and jitter, SINR is strongly dependent on the values of β and N at higher values of SNR. Here, the sample rate, R is kept constant. As shown in Fig. 6, when β is very small compared to the subcarrier spacing, i.e., $\beta\frac{N}{R}$ is of the order of 10^{-4} or less, the ICI due to phase noise is negligible. As a result, there is a constant SINR penalty due to the presence of CFO and timing jitter along with fading. Meanwhile, for high phase noise levels with $\beta\frac{N}{R} \geq 1$, SINR penalty exceeds the value of SNR itself, which implying that the ICI overwhelms the desired signals. Higher transmission data rate, R results in better system performance.

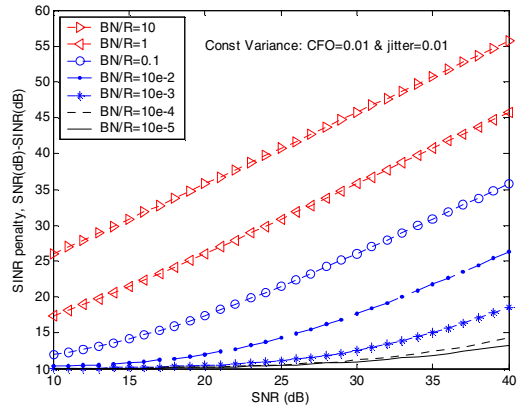


Fig. 6 SINR penalty as a function of SNR for different ($\beta\frac{N}{R}$)settings.

Fig. 7 reflects the BER performances of an OFDM system over rayleigh channels with different fading variances where the variances of all three synchronization impairments are assumed constant. It is noticeable that the BER performance improves with the increase of the fading variance. This is because attenuation of the input SNR decreases at high variances of rayleigh fading channels.

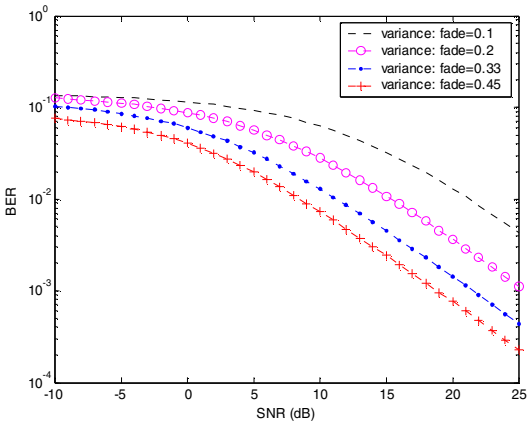


Fig. 7 BER performances of an OFDM system for different variances of rayleigh fading channel with $\sigma_\varepsilon^2 = 0.03, \sigma_\xi^2 = 0.01, \sigma_u^2 = 1\times10^{-5}$

V. Conclusion

In this paper, an analytical technique is provided for evaluating the performance of an OFDM system impaired by CFO, phase noise and timing jitter over rayleigh fading channels. An exact close-form expression for the SINR is derived and the BER performances of a BPSK-OFDM system are evaluated considering these effects. It is noticed that the OFDM system suffers significant penalty due to CFO and jitter, however, the effect of phase noise is the dominant one. It is shown by analysis that the system performance also depends on several critical parameters such as number of subcarriers, phase noise linewidth, transmission data rate, input SNR and the fading characteristics of the channel.

Appendix

Energy of $\Theta(r)$

The DFT response of $\varphi_m(n)$ is given by

$$\Theta(r) = \frac{1}{N}\sum_{n=0}^{N-1}\varphi_m(n)e^{-j\frac{2\pi}{N}nr} = \frac{1}{N}\sum_{n=0}^{N-1}[C_m + \sum_{i=0}^{n}u(T_m+i)]e^{-j\frac{2\pi}{N}nr}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{i=0}^{n}u(T_m+i)e^{-j\frac{2\pi}{N}nr}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}u(T_m+n)\sum_{i=n}^{N-1}e^{-j\frac{2\pi}{N}ir} = \frac{1}{N}\sum_{n=0}^{N-1}u(T_m+n)\cdot\frac{e^{-j\frac{2\pi}{N}nr}-1}{1-e^{-j\frac{2\pi}{N}r}}$$

$$= \frac{-1}{N\sin(\frac{\pi r}{N})\cdot e^{-j\frac{\pi}{N}r}}\sum_{n=0}^{N-1}u(T_m+n)\cdot\sin(\frac{\pi rn}{N})e^{-j\frac{\pi}{N}nr}$$

Note that $\frac{1}{N}\sum_{n=0}^{N-1}e^{-j\frac{2\pi}{N}nr} = 0$ for $r = 1,2,\ldots,N-1$. Hence according to the mutual independence of Gaussian random variables $u(i)$'s, we can calculate the energy of $\Theta(r)$ as

$$E[|\Theta(r)|^2] = \frac{\sigma_u^2}{N^2\sin^2(\frac{\pi r}{N})}\sum_{n=0}^{N-1}\sin^2(\frac{\pi rn}{N})$$

$$= \frac{\sigma_u^2}{2N\sin^2(\frac{\pi r}{N})}$$

Since for $r \neq 0$ and for even N, $\sum_{n=0}^{N-1}\sin^2(\frac{\pi rn}{N}) = \frac{N}{2}$

References

[1] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz Band," *IEEE Std.* 802.11a, 1999.

[2] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band," *IEEE Std.* 802.11g, 2003.

[3] H. Steendam and M. Moeneclaey, "Synchronization Sensitivity of Multicarrier Systems," *Euro. Trans. Telecomms.*, vol. 15, pp. 223-234, 2004.

[4] P.H.Moose, "A Technique for Orthogonal Frequency-Division Multiplexing Frequency Offset Correction," *IEEE Trans. Commun.*, vol. 42, pp. 2908-2914, Oct. 1994.

[5] S.Wu and Y. Bar-Ness, "OFDM Systems in the Presence of Phase Noise: Consequences and Solutions," *IEEE Trans. Commun.*, vol. 52, No. 11, pp. 1988-1996, Nov. 2004.

[6] L.Tomba, W.A. Krzymien, "A Model for the Analysis of Timing Jitter in OFDM Systems," in *Proc. IEEE Int. Conf. Commun. (ICC 98),* vol. 3, pp. 1227-1231, June 1998

[7] J. Armstrong, "Analysis of New and Existing Methods of Reducing Intercarrier Interference Due to Carrier Frequency Offset in OFDM," *IEEE Trans. Commun.,* vol. 47, pp. 365-369, March 1999.

[8] T. Pollet, M. van Bladel and M. Moeneclaey, " BER Sensitivity of OFDM Systems to Carrier Frequency Offset and Wiener Phase Noise," *IEEE Trans. Commun.,* vol. 43, pp. 191-193, Feb. 1995.

[9] A.G. Armada, "Understanding the Effects of Phase Noise in Orthogonal Frequency-Division Multiplexing (OFDM)," *IEEE Trans. Broadcast.,* vol. 47, pp. 153-159, June 2001.

[10] L. Tomba, "On the Effect of Wiener Phase Noise in OFDM System," *IEEE Trans. Commun.,* vol. 46, pp. 580-583, May 1998.

[11] Y.Zhang and H.Liu, "MIMO-OFDM Systems in the Presence of Phase Noise and Doubly Selective Fading," *IEEE Trans. Veh. Technol.,* vol. 56, no. 4, pp. 2277-2285, July 2007.

[12] T.N.Zogakis and J.M.Cioffi, "The effect of timing jitter on the performance of a discrete multitone system," *IEEE Trans. Commun.*, vol. 44, no. 7, pp. 799-808, July 1996.

[13] T.Pollet and M.Moeneclaey, "Synchronizability of OFDM signals," in *Proc. Globecom'95*, pp. 2054-2058, Singapore, Nov. 1995.

[14] J.G. Proakis, *Digital Communications,* 4th ed., New York: McGraw-Hill,2001.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Adaptive Resource Allocation Based on Modified Genetic Algorithm and Particle Swarm Optimization for Multiuser OFDM Systems

Imtiaz Ahmed, *Member IEEE* and Satya Prasad Majumder, *Member IEEE*

Department of Electrical & Electronic Engineering,
Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
Email: imtiaz123b@gmail.com , spmajumder@eee.buet.ac.bd

Abstract - Adaptive resource allocation is one of the most challenging tasks for multiuser Orthogonal Frequency Division Multiplexing (OFDM) systems. In this paper, two evolutionary approaches, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) have been applied for adaptive subcarrier and bit allocations to minimize the overall transmit power of a multiuser OFDM system. Each user will be assigned a number of subcarriers with at least one minimum subcarrier even at the worst case. Then the number of bits and the transmit power level for each subcarrier are calculated. Simulation results show that both the evolutionary approaches outperform the conventional static resource allocation schemes considerably in multiuser scenario. The results further reveal that both the algorithms can handle large allocation of subcarriers without significant performance degradation. However the performance of PSO is found to be better than the GA in terms of execution time, simplicity and convergence.

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is considered as one of the most promising transmission techniques in wideband wireless systems because of its bandwidth efficient performance in combating multipath fading as well as Inter Symbol Interference (ISI). The use of adaptive modulation schemes can enhance the system performance by changing its modulation constellation and transmit power according to the instantaneous Channel State Information (CSI). The combination of OFDM and adaptive modulation can utilize the merits of both technologies and is attracting more and more interests.

To meet the increasing user demand for high data rate, allocation of appropriate frequency spectrum as well as efficient use of channel frequencies have become very a vital issue. OFDM, like other wireless systems, requires the proper allocation of the limited resources, e.g. total transmit power and available frequency bandwidth, among the users to meet the users' service requirements. Two classes of resource allocation schemes exist in OFDM systems: fixed resource allocation [1] and dynamic resource allocation [2] [3] [4] [5]. Fixed resource allocation schemes, such as time division multiple access (TDMA) and frequency division multiple access (FDMA), assign an independent dimension, e.g. time slot or subchannel, to each user. A fixed resource allocation scheme is not optimal since the scheme is fixed regardless of the current channel condition.

On the other hand, dynamic resource allocation allocates a dimension adaptively to the users based on their channel gains. Due to the time-varying nature of the wireless channel, dynamic resource allocation makes full use of multiuser diversity to achieve higher performance. In [6], Wong et al proposed iterative searching algorithm that applies Lagrangian relaxation for optimum multiuser subcarrier, bit and power allocation. The algorithm is close to the lower bound with the requirement of high and complex computation. The algorithm proposed in [7], however, over-simplifies the subcarrier allocation but could not fully utilize the multiuser diversity. In [8], an iterated water-filling algorithm is proposed; the algorithm can acquire similar performance as Wong's algorithm and avoids the computational complexity. Y. B. Reddy et al introduced Genetic Algorithm in resource allocation with significant improvements [9].

In this literature, evolutionary approaches have been applied through a proposed algorithm of resource allocation. To our best knowledge, adaptive resource allocation for OFDM systems based on Particle Swarm Optimization (PSO) method is still missing in the literature [10][11]. In this paper, PSO is introduced as a promising technique of adaptive resource allocation for multiuser OFDM systems. The performance of PSO has been compared with a modified Genetic Algorithm (GA). The convergence of the conventional GA [12] has been improved through the method of fractional generation gap. The performance of the modified GA and PSO methods are also compared with some of the existing fixed subcarrier and bit allocation schemes.

The rest of the paper is organized as follows: In Section II, the problem of dynamic resource allocation is formulated in multiuser OFDM system and system model is described. The modified GA and PSO based description for this problem and the parameters for these evolutionary approaches are discussed in Section III. Section IV gives the numerical results and the discussion of the system performance and finally Section V finishes the paper by giving conclusion.

II. System Model

Let we consider a multiuser OFDM system having K (k =

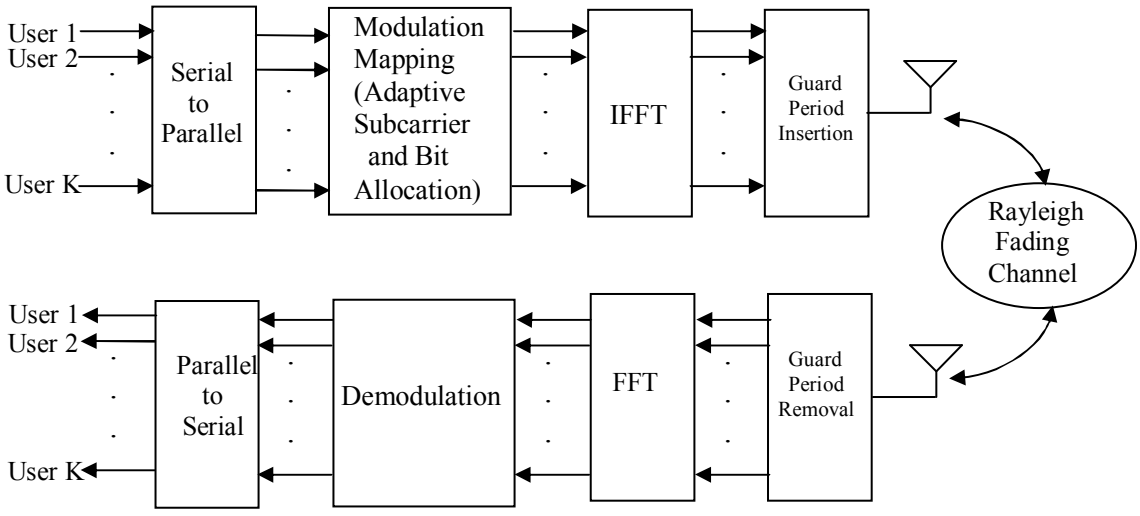


Fig. 1 Basic OFDM system model having the provision of resource allocation section in the transmitter.

1,2, … , K) users and N (n = 1, 2, … , N) subcarriers. The system allots a subset of N subcarriers to a particular user and determines the number of bits per each assigned subcarrier on downlink transmission (Fig. 1).

Let $b_{n,k} \in \{0,1,2,.....,B\}$ signify the number of bits for nth subcarrier and kth user where B denotes the maximum number of information bits that can be transmitted by each subcarrier. Here $b_{n,k}$ determines the mode of adaptive modulation (i.e. BPSK, 16 QAM, 64 QAM or anything else). The system has been assumed to acquire channel state information through its dynamic channel estimation scheme. Let $\alpha_{n,k}$ represents the channel gain for nth subcarrier and kth user. The required transmission power for the specified bit error rate at $b_{n,k}$ bits per symbol is given by [13],

$$p_{n,k} = \frac{f(b_{n,k})}{\alpha_{n,k}^2} \tag{1}$$

In multiuser scenario, not more than one user is considered to share a particular subcarrier. Mathematically it is expressed as

$$\lambda_{n,k} = \begin{cases} 1 \ if \ b_{n,k} \neq 0 \\ 0 \ if \ b_{n,k} = 0 \end{cases} \tag{2}$$

The required total transmission power ($P_{n,k}$) can be written as follows [14]

$$P_{n,k} = \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{f(b_{n,k})}{\alpha_{n,k}^2} \times \lambda_{n,k} \tag{3}$$

where,

$$f(b_{n,k}) = \frac{N_0}{3}[Q^{-1}(\frac{BER_n}{4})]^2(2^{b_{n,k}} - 1) \tag{4}$$

$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt$ and here Q^{-1} denotes the inverse Q function.

III. Application of the evolutionary approaches in OFDM system

The purpose of resource allocation at the base station is to allocate intelligently the limited resources, e.g. total transmit power and available frequency bandwidth, among users to meet users' service requirements. Channel-aware adaptive resource allocation has been shown to achieve higher system performance than static resource allocation, and is becoming more critical in current and future wireless communication systems as the user data rate requirements increase. Furthermore, the subcarrier allocation problem to multiple users has many different permutations, thereby making the solution space very large. Unlike other algorithms, the evolutionary approaches can handle large solution space without any performance degradation. In this paper, the subcarriers and bits are allocated to different users according to the dynamic channel state information. Each user is allocated one or more subcarriers provided that one subcarrier can be used by only one user. The number of bits are then chosen according to the water filling algorithm i.e. the modulation schemes are selected in response of the channel state information of the corresponding user. The optimum arrangement of the users as well as subcarriers can be evaluated by one of the two evolutionary approaches, Genetic Algorithm (GA) or Particle Swarm Optimization (PSO). The flowchart of this problem has been attached to the Appendix (Fig. 6).

A.1 Genetic algorithm based allocation:

Genetic algorithm [9], [12] is inspired by the mechanism of natural selection where stronger individuals are likely to be the winners in a competing environment. The continuing performance improvement of the computational system has made GA attractive for some

types of optimization. As a matter of fact GA is very suitable for the optimization of bit and subcarrier allocation problem in multiuser OFDM system.

A.1.1 Genetic based algorithm

- Generate chromosomes of N elements and population size M for this experiment. Each element in the chromosome is a subcarrier allocated to a user. The population is a 2-D array where the number of rows represents population size and the column of a row represents subcarriers.
- Evaluate the water filling algorithm to allocate each user's bits and subcarriers and calculate the overall transmission power.
- Generate the new population using crossover and mutation using their occurrence probability.
- Execute the program to reach the optimum fitness.

Repeat the last two steps till the system converges. In this paper the GA has been modified slightly over the conventional one by setting a fractional value of the generation gap. The fractional generation gap helps to converge quickly by taking the good genes for the next generation.

A.2 Particle swarm optimization based allocation:

Particle swarm optimization (PSO) is one of the evolutionary computational techniques. Like the other evolutionary computation techniques, PSO is a population-based search algorithm and is initialized with a population of random solutions, called particles. The original PSO algorithm is discovered through simplified social model simulation.

A.2.1 Particle swarm based algorithm

The original PSO algorithm [10] can be described by

$$v_{id} = wv_{id} + c_1 rand()(p_{id} - x_{id}) + c_2 Rand()(p_{gd} - x_{id}) \quad (5)$$

$$x_{id} = x_{id} + v_{id} \quad (6)$$

where c_1 and c_2 are positive constants, w is the inertia weight and rand() and Rand() are two random functions in the range [0,1]; $X_i = (x_{i1}, x_{i2}, ...x_{iD})$ represents the ith particle; $P_i = (p_{i1}, p_{i2}, ...p_{iD})$ represents the best previous position (the position giving the best fitness value) of the i^{th} particle; the symbol g represents the index of the best particle among all the particles in the population; $V_i = (v_{i1}, v_{i2}, ...v_{iD})$ represents the rate of the position change (velocity) for particle i. The processing steps [11] are as follows:

- Initialize a population of N particles which signifies the random positions and velocities on D dimensions of the users' allocations of subcarriers in the problem space.
- Then evaluate the water-filling algorithm to allocate users' bits and subcarriers.
- For each particle, evaluate the optimization fitness function of equation (1) in D variables.
- Compare particle's fitness evaluation with its pbest. If current value is better than pbest, then set pbest equal to the current value, and P_i equals to the current location X_i in D-dimensional space.
- Identify the particle in the neighborhood with the best success so far, and assign its index to the variable g.
- Change the velocity and position of the particle according to equation (5) and (6).
- Loop to step 2 until a criterion is met, usually a sufficiently good fitness or a maximum number of iterations.

IV. Simulation results

In this section, genetic algorithm as well as the particle swarm optimization have been studied extensively along with some relative comparative features. The channel has been assumed here as the quasi static. The channel state information at each subcarrier is generated randomly and subject to 'Rayleigh' distribution. In this simulation, 64 subcarriers have been used for 2, 4, 6 & 8 users. Using water-filling algorithm the subcarriers have been allocated as needed. The simulations have been done with different numbers of initial population and swarms. The bits allocations are allocated according to the state of the channel. The modulation type in this OFDM system has been confined to no modulation (0 bits), QPSK (2 bits), 16 QAM (4 bits) and 64 QAM (6 bits). According to these specifications, the simulations have been performed by modified GA and PSO. The parameters of the used algorithms are shown in Table-1.

Fig. 2 & 3 represent the convergence curves (evaluated by GA & PSO) with different sizes of initial population and swarms. Initial population size greater than 60 gives similar impact whereas its lower value gives few dBm higher than the original result (Fig. 2). But unlike GA, the low value of initial swarm size does not degrade the overall performance too much in PSO. So it is quite sufficient to take the value of 25 for initial size of swarm. This lowers the need of memory size as well as computational complexity for PSO.

Table 1: Parameters of modified GA and PSO

Modified GA		PSO	
Initial Population size	80	Initial Swarm Size	25
Generations	0 to 100	Generations	0 to 100
Crossover	0.6	C_1	1.5
Mutation	0.03	C_2	1.5
Generation Gap	0.8	Initial Inertia weight	0.8

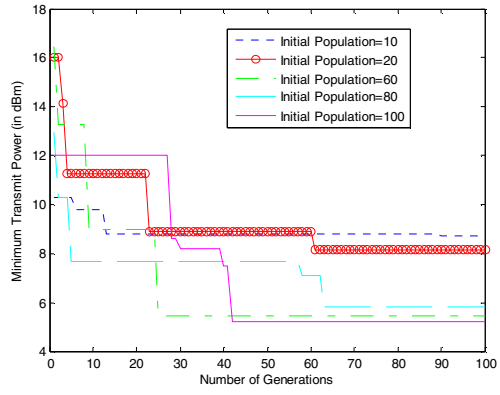


Fig. 2 Convergence curves of modified GA for different sizes of initial population.

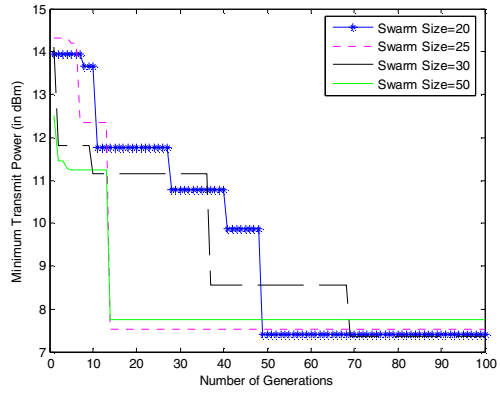


Fig. 3 Convergence curves of PSO for different swarm sizes.

With these values of initial sizes of the algorithms, the modified GA and PSO have been used for simulation to allocate the resources for multiuser OFDM systems. The minimum power has been calculated with these algorithms for 10 times (Table 2). In Fig. 4, both the algorithms have been simulated for 0 to 200 generations. It is clearly evident that PSO converges to lower value compared with GA (also similar in Table 2) although the initial rate of convergence is higher for GA. The relative higher rate in initial convergence may be attributed to higher number of individuals in the population and hence higher number of function evaluated in one generation of GA, which is several times compared with that of PSO.

Table 2: Minimum transmit power evaluation by modified GA and PSO

Run	Minimum Transmit Power (in dBm) (By GA)	Minimum Transmit Power (in dBm) (By PSO)
1	4.462	4.829
2	5.048	4.015
3	4.549	5.767
4	5.9	4.698
5	5.047	4.864
6	4.625	4.654
7	5.738	4.662
8	4.94	5.149
9	6.207	5.757
10	6.628	4.328
Mean	5.3144	4.8722

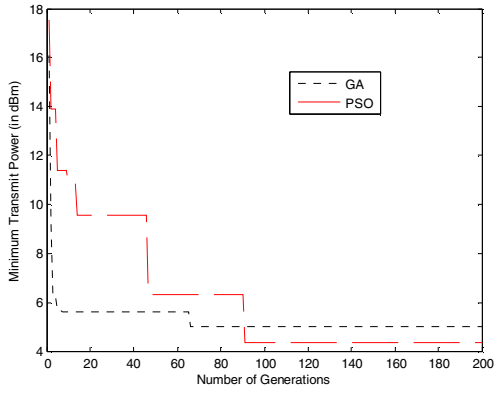


Fig. 4 Comparison of the convergence of GA and PSO

Although there are huge differences between GA and PSO in terms of internal operations to update the solutions, yet both of them are population-based and evaluate objective function and they try to optimize it. As such the number of function evaluated to achieve the target (objective function) value and corresponding central processing unit (CPU) execution time are considered as the performance metrics to carry out a comparative assessment between GA and PSO (in Table-3). Here in all the cases GA needs more time to converge than PSO although the number of generations for GA is sometimes lower than that of PSO. It is mainly due to the fact that GA needs more functions to evaluate to reach an optimum value whereas PSO needs to execute only two simple functions per generation for each swarm.

Table 3: Comparative performance measures of modified GA and PSO for a target value of 3.01 dBm*

	PSO			Modified GA		
Run	**CPU execution time (s)**	**Number of generations**	**Number of functions evaluated**	**CPU execution time (s)**	**Number of generations**	**Number of functions evaluated**
1	2.5428	66	1650	3.978	30	4320
2	2.2932	60	1500	12.8701	100	14400
3	1.2012	30	750	4.1184	32	4608
4	3.666	100	2500	8.2681	64	9216
5	2.7144	70	1750	4.0872	31	4464
6	3.6192	100	2500	12.9481	100	14400
7	1.1388	24	600	4.2744	32	4608
8	3.6972	100	2500	12.4021	95	13680
9	0.3276	6	150	1.1232	9	1296
10	1.6536	43	1075	4.9296	31	4464

***All the simulations have been carried out on a PC (Processor: Intel(R), Core(TM) 2 CPU, 1.73 GHz, RAM: 1022 MB).**

The total transmit power, calculated by modified GA and PSO, has been compared to the initial static algorithms (in Fig. 5). The dynamic allocation algorithms outperform the static ones whereas PSO gives better result for higher numbers of users than the modified GA.

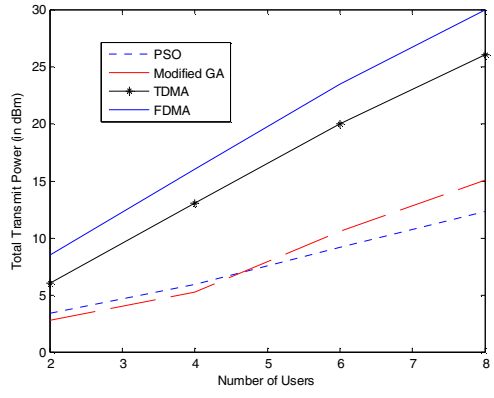


Fig. 5 Total transmit power with different algorithms for bit error rate = 10^{-3}

V. Conclusion

In this paper, the adaptive bit and subcarrier allocation has been developed with modified genetic algorithm and particle swarm optimization. The genetic algorithm has been modified by using a fractional generation gap with the fact in mind that it converges faster than the original one. The performance obtained by particle swarm optimization shows relatively better result than the modified genetic modified in terms of simplicity, coding capability, computational resources, execution time and convergence. In multiuser OFDM systems the proposed algorithm using the evolutionary approaches outperforms the static allocation schemes and the particle swarm approach performs even better than the modified GA for higher number of users.

Appendix

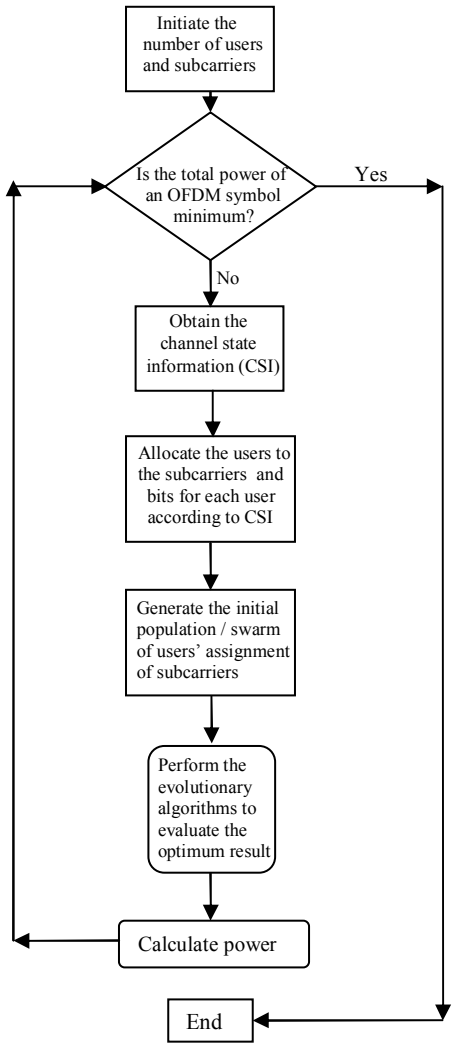


Fig. 6 Flowchart of the proposed algorithm

References

[1] E. Lawrey, “Multiuser OFDM," in Proc. International Symposium on Signal Processing and its Applications, Brisbane, Australia, Aug. 1999.

[2] J. Jang and K. B. Lee, “Transmit power adaptation for multiuser OFDM systems," IEEE Journal on Selected Areas in Communications, vol. 21, no. 2, pp. 171-178, Feb. 2003.

[3] I. Kim, H. L. Lee, B. Kim, and Y. H. Lee, “On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM," in Proc. IEEE Global Communications Conf., vol. 6, pp. 3648-3652, Nov. 2001.

[4] W. Rhee and J. M. Cioffi, “Increasing in capacity of multiuser OFDM system using dynamic subchannel allocation," in Proc. IEEE Int. Vehicular Tech. Conf., Tokyo, Japan, May 2000.

[5] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, “Multicarrier OFDM with adaptive subcarrier, bit, and power allocation," IEEE Journal on Selected Areas in Communications, vol. 17, no. 10, pp. 1747{1758, Oct. 1999.

[6] C. Wong and et al, “Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation,” IEEE JJAC, Vol. 17, No.10, Oct, pp. 1747-1758, 1999.

[7] E. Bakhtiari and B.H. Khalaj , “A new joint power and subcarrier allocation scheme for multiuser OFDM systems” in 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, Beijing, China, Vol. 2, pp.1959 - 1963, Sept.7-10, 2003.

[8] G. Zhang, “Subcarrier and bit allocation for real-time services in multiuser OFDM systems” in 2004 IEEE International Conference on Communications, Paris, France, Vol. 5, pp.2985-2989, June 20-24, 2004.

[9] Y. B. Reddy and N. Gajendar, “Evolutionary approach for efficient resource allocation in multi-user OFDM systems”, Journal of Communications, Vol. 2, No. 5, pp. 42-48, August, 2007.

[10] R. C. Eberhart, and J. Kennedy, (1995). A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, pp. 39-43. , 1995 , Piscataway, NJ: IEEE Service Center.

[11] Y. Shi and R. C. Eberhart, (1998a), “Parameter selection in particle swarm optimization”, Proceedings of the 1998 Annual Conference on Evolutionary Computation, March 1998.

[12] K. F. Man K. S. Tang and S. Kwong, “Genetic Algorithms: concepts and applications”, IEEE transactions on Industrial Electronics, Vol. 43, Issu: 5, Oct. pp 519-534, 1996.

[13] M. Wahlqvist et al. , “Capacity comparison of an OFDM based multiple access system using different dynamic resource allocation” Proc. Vehicular Technology Conf. , vol. 3, pp 1664-1668, 1997.

[14] J. G. Proakis, “Digital Communication” , 4th ed. New York, McGraw-Hill, 2000

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Comparison Study of various Call Admission Control Scheme in WCDMA Network

Syed Foysol Islam, Md. Firoz Hossain,
M.Sc. in Electrical Engineering (BTH, Sweden)
Lecturer, Department of Electronics and Telecommunication Engineering,
University of Development Alternative (UODA), Bangladesh.
E-mail: engg_rumi@yahoo.com, firoz_ce@yahoo.com

Abstract:

The main objective of this research is to derive a mathematical model of call admission control in WCDMA network. Three main call admission algorithm wideband power based (WPB), throughput based (TB) and adaptive call admission control (ACAC) algorithm are investigated throughout this paper and a little comparison between them is presented.

Key Words: Wide Band Code Division Multiple Access (WCDMA), Wideband power based (WPB), Throughput based (TB) and Adaptive call admission control (ACAC)

I. Introduction

When a new call arrives in the system, it needs to check whether to accept the call or not. At first the system has to examine whether the new call is going to degrade the quality of the ongoing calls or the planned coverage area. If it attempts to make degradation in the system, then the system should block the call. In order to maintain the required quality of service of the new incoming call, there are three parameters that have to be checked: required SIR, inter cellular interference, intracellular interference. Based on these parameters the system admits the call in a selective way that does not affect the ongoing calls. This decision making part of the UMTS network is called the call admission control (CAC). In this research we will deeply study three call admission schemes and their performance.

Calculation of SIR:

$$SIR = \frac{Signal\ Power}{Total\ Interface\ Power} \qquad (1)$$

Equatiion (I) can be simplified as

$$SIR = SF.\frac{P_j}{I_{total}} = SF.\frac{P_j}{I_{inter} + I_{intra} + P_n} \qquad (2)$$

Where,

P_j = Received signal power of the user at Node B,

$$I_{total} = I_{inter} + I_{intra} + P_n \qquad (3)$$

I_{inter} = Interference caused by the Intercellular communications, I_{intra}= Interference caused by the Intra cellular communications, P_n= Thermal Noise which is assumed to be -99dBm in the downlink and -103 dBm in the uplink

SF= Spreading Factor

$$\text{Spreading Factor} = \frac{Carrier\ Bandwidth}{Information\ Rate} = \frac{Chip\ Rate}{Data\ Rate} = \frac{W}{R} \qquad (4)$$

II. Call Admission Control Schemes

We have reviewed a lot of papers on this issue. Each method takes different parameter to make the decision criteria. Intercell interference and intracell interference are taken into account to measure the wideband received power based (WPB) admission control and the system throughput based (TB) admission control, service specific admission control, an heuristic method for making the decision of admission control, call admission control depends on the available bandwidth and capacity of the system presented in [1] [7] [5] [6] respectly. An adaptive method for call admission control (ACAC) focused in [4]. In this paper we have investigated on two main call admission control algorithm WPB and TB. A brief discussion on these methods is presented in this paper. A new promising method adaptive call admission control (ACAC) also compared with the previous two methods.

a) WPB Admission Control:

Interference caused by the mobile stations within the own cell and also by the neighboring cells taken into account in this method The system maintains a threshold value both for uplink and downlink for accepting a new call.

UP Link: A new call is accepted only when the new total interference ($I_{total}+\Delta I$) caused by the new call is less than the threshold value (I_{th}) set by radio network planning. If the new resulting total interference that caused by the new call exceeds the

threshold value it should be blocked. The mathematical representation of this formula is given by the equation (5)

$$\underbrace{I_{total\_old} + \Delta I}_{Total\ Interference} \langle I_{th} \qquad (5)$$

Where,

I_{total}= The interference before admitting the new call

ΔI = The estimated interference caused by the new call,

Figure 1 shows the explanation of this method. Let us assume that in a power controlled system the load of the system at any instant is L_{old} and that creates the interference I_{old}. Now consider a new call coming to the Node B for getting admission then the RNC estimates the interference it would create as ΔI which is marked as I_{new}. The admission control algorithm checks whether this total interference ($I_{old}+\Delta I$) would exceed the predefined threshold value I_{th}. If the total interference exceeds the threshold value I_{th} then that call must be blocked.

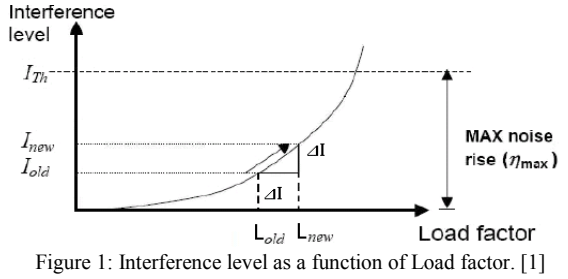


Figure 1: Interference level as a function of Load factor. [1]

As we have seen from the equation *(4)* that the estimated value of interference need to calculated. There are two methods for the calculation of increase interference or power, the derivative method and the integration method. Both take into account the load curve and are based on the derivative of uplink interference with respect to the uplink load factor i.e.

$$\frac{dI_{total}}{d\eta} \qquad (6)$$

We Know Noise rise is given by [1],

$$Noise\ rise = \frac{The\ interference\ before\ admitting\ new\ call}{Thermal\ Noise}$$

$$\approx \frac{I_{total}}{P_n} \approx \frac{1}{1-\eta}$$

$$\therefore I_{total} \approx \frac{P_n}{1-\eta}$$

$$So,\ \frac{dI_{total}}{d\eta} \approx \frac{P_n}{(1-\eta)^2} \qquad (7)$$

The change in the uplink interference can be obtained by the following equations

$$\frac{\Delta I}{\Delta L} \approx \frac{dI_{total}}{d\eta}$$

$$\therefore \Delta I \approx \frac{dI_{total}}{d\eta} \Delta L \qquad (8)$$

Now using equation (7),

$$\therefore \Delta I \approx \frac{P_n}{(1-\eta)^2} \Delta L \qquad (9)$$

Substituting by the value of P_n, equation (8) can be simplified as

$$\Delta I \approx \frac{I_{total}}{1-\eta} \Delta L \qquad (10)$$

The second uplink interference increase estimation based on the integration method in which the differentiation of uplink interference with respect to the load factor is integrated from the old value of load factor ($L_{old} \approx \eta$) to the new value ($L_{new} \approx \eta + \Delta L$) i.e.

$$\Delta I \approx \int_{\eta}^{\eta+\Delta L} dI_{total} \qquad (11)$$

$$\approx \int_{\eta}^{\eta+\Delta L} \frac{P_n}{(1-\eta)^2} \Delta L$$

$$\approx \frac{P_n}{1-\eta-\Delta L} - \frac{P_n}{1-\eta}$$

$$\approx \frac{P_n(1-\eta-1+\eta+\Delta L)}{(1-\eta-\Delta L)(1-\eta)}$$

$$\approx \frac{\Delta L}{(1-\eta-\Delta L)} \cdot \frac{P_n}{(1-\eta)} \qquad (12)$$

Simplified by equation (6)

$$\Delta I \approx \frac{I_{total}}{1-\eta-\Delta L} \Delta L \qquad (13)$$

The value of load ΔL is given by

$$\Delta L \approx \frac{1}{1+\frac{W}{(E_b/N_0)vR}} \qquad (14)$$

Where, E_b/N_0 denotes signal to noise ratio, W is the chip rate, v is the activity factor and R data rate of traffic.

Downlink: In the downlink the same strategies is used but in this case the considering parameter is transmission power. If the new total downlink transmission power does not exceed the threshold power value, then the call is admitted.

$$\underbrace{P_{total\_old}}_{Total\ Power} + \underbrace{\Delta P}_{r} \langle P_{th} \qquad (15)$$

P_{total_old}: The transmission power before admitting the new call, ΔP: Estimated transmission power required for the new call, P_{th}: Threshold value set by radio network planning, *Total Power*: Total

estimated transmission power,The power increase ΔP_{total} is estimated by the initial power.

b) Throughput Based Admission Control

Unlike wide band power based admission control, throughput based admission control takes into account the load. Two different threshold values one for uplink threshold and downlink threshold are used for taking decision.

Uplink:

The new user is not admitted in the system if the new total load exceeds the predefined uplink threshold set by the radio network planning.

$$\underbrace{\eta_{ul} + \Delta L}_{Total\ Load} \rangle\ \eta_{ul\_Th} \qquad (16)$$

Where, η_{ul}: The load before admitting new user L_{old}, ΔL : Estimated load for the new user or call, η_{ul_Th}: Threshold value for the uplink load factor, *Total Power:* Total estimated load for the new user

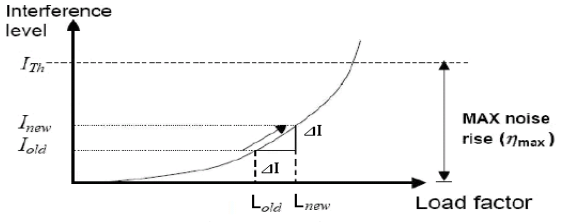


Figure 2: Load Curve

Down Link: The new call is not admitted in the system if the total resulting load exceeds the downlink threshold value.

$$\eta_{DL} + \Delta L \rangle\ \eta_{DL\_Th} \qquad (17)$$

Where η_{DL} can be calculated as

$$\eta_{DL} \approx \frac{\sum_{j=1}^{N} R_j}{R_{\max}} \qquad (18)$$

N is the total no of connections in the system, R_j is the bit rate of user j and R_{max} is the maximum allowed throughput of the cell [1].

c) Adaptive Call Admission Control:

ACAC scheme, the base station updates the total no of users to the RNC in regular intervals (τ). This small interval may call an epoch. With this information the RNC should decide which scheme (WPB or TB) it needs to switch to, by calculating the number of each type of user presented in the system at the end of a previous epoch. If there are more voice users, the ACAC switches to WPB and if there are more data users, it switches to the TB scheme. This prediction depends on α, which is the parameter used to predict the number of calls in the coming epoch and β, keeps the information of total number of calls that have originated in the system since start-up. The values of α and β varies between 0 and 1 and are calculated adaptively through simulations [4], [8]. The predicted no of calls that arrive in the system determined by the following equations

$$\hat{V}_{n+1} = \alpha V_n + (1-\alpha)\hat{V} + \beta V_{total} \qquad (19)$$

$$\hat{D}_{n+1} = \alpha D_n + (1-\alpha)\hat{D} + \beta D_{total} \qquad (20)$$

Where, $\hat{V}_{n+1}$ *:* voice calls arrival in the coming epoch, $\hat{D}_{n+1}$ *:* data calls arrival in the coming epoch, $\hat{V}_n$ *:* voice calls in the previous epoch, $\hat{D}_n$ *:*data calls in the previous epoch, V_n *:* Originated number of voice calls in the previous epoch, D_n *:* Originated number of data calls in the previous epoch.

In a system where (m-k) channels are busy is defined by the following equation

$$\beta(m,k) = \frac{\beta(m-1,k-1)}{1+\frac{1}{m}\sum_{r=0}^{R-1} A_r b_r \beta(m-1,b_r-1)} \qquad (21)$$

Here, *R:* The number of traffic classes (*0 – R –1*), b_r*:* Required data rate, *m:* No of servers in the system and *k >0*

$$A \approx \frac{\lambda_r}{\mu_r}$$

$$\approx \frac{Poisson\ \ distributed\ \ call\ \ arrival\ \ rate\ \ of\ \ class\ \ r}{Exponential\ \ distributed\ \ call\ \ arrival\ \ rate\ \ of\ \ class\ \ r}$$

The initial values of β measured by the following equations

$$\beta(m,0) = \frac{\frac{1}{m}\sum_{r=0}^{R-1} A_r b_r \beta(m-1,b_r-1)}{1+\frac{1}{m}\sum_{r=0}^{R-1} A_r b_r \beta(m-1,b_r-1)} \qquad (22)$$

III. Comparative Result

Contrast between WPB and TB schemes is shown by the figure 3. It has been observed from the graph that more interference will add from the neighboring cells with the increasing value of *i*. The other cell to own cell interference ratio *i* with value 0 means no interference from the neighbor.

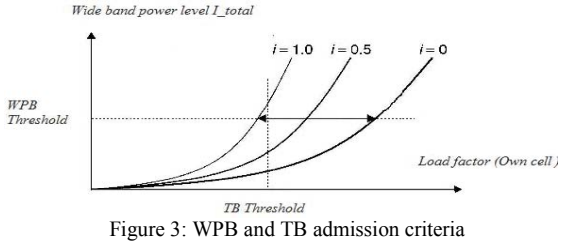


Figure 3: WPB and TB admission criteria

WPB takes the interference from adjacent frequency bands. This could be originated from the other operator's mobile station, which is closer to a base station. So that it could perform an overestimate of the wide band received power. TB does not take inference from the neighboring cells. Rather it concern about the loading of the neighboring cells through the RNC.

Adaptive call admission control (ACAC) combines the WPB and TB schemes. Depending on the total no of voice *(19)* and data users *(20)* it switches between WPB and TB scheme. If there is more voice user in the system ACAC switches to WPB mode and if there is more data users than the voice users the ACAC follow the TB mode. The limitations of WPB and TB overcome by the ACAC scheme. The call blocking probability in ACAC is tends to be zero comparing other two methods. Figure 4 and 5 compares the performance of these three methods by call blocking probability call dropping probability.

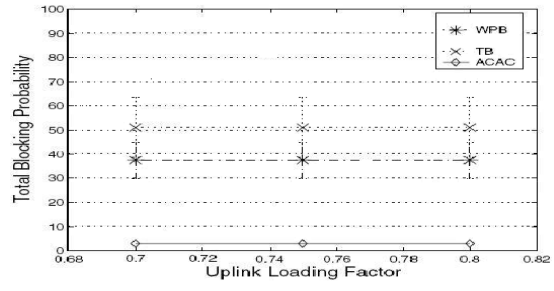


Figure 4: Call blocking probability of WPB, TB and ACAC scheme

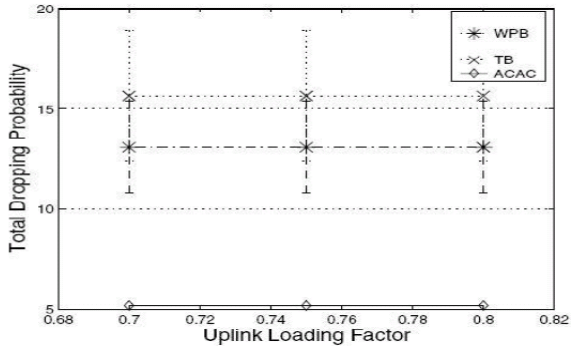


Figure 5: Call dropping probability of WPB, TB and ACAC Scheme

Figure 4 and figure 5 help us to observe that the call blocking probability in ACAC is less than the WPB and TB. The call dropping probability in ACAC is less than the WPB and TB schemes. So we can say that the ACAC is best algorithm.

VI. Conclusion

Call admission control plays the primary role in radio resource management. As it is used in wireless networks to optimize the system performance and guarantee the QoS. By using a perfect admission control algorithm congestion and over load of the network can be eliminated. Two major admission control algorithms WPB and TB are studied in this paper. One of the latest algorithms ACAC is also studied in this paper. We have observed that Adaptive CAC's which is the combination of the above two methods could be a better option for a system design. We have limited our work only within the WCDMA FDD mode.

References:

[1] Harri Holma and Anti Toskala ,"WCDMA for UMTS radio access for third generation mobile communications", *John wiley and sons ltd.*

[2] Il-Min Kim, Byung-Cheol Shin, and Dong-Jun Lee," SIR-based call admission control by intercell interference prediction for DS-CDMA systems", *IEEE COMMUNICATIONS LETTERS, VOL. 4, NO. 1, JANUARY 2000.*

[3] Technical Specification of 3rd Generation Partnership Project on Radio Resource Management "3G TR 25.922 V3.0.0".

[4] Kamala Subramaniam, Arne A. Nilsson,"An analytical model for adaptive call admission control scheme in a heterogeneous UMTS-WCDMA system", *2005 IEEE.*

[5] Jun Ye, Xuemin (Sherman) Shen and Jon W. Mark ," Call admission control in Wideband CDMA cellular Networks by using fuzzy logic", *IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 4, NO. 2, MARCH/APRIL 2005.*

[6] Songsong Sun and Witold **A.** Krzymien, "Call admission policies and capacity analysis of a multi-service CDMA personal communication system with continuous and discontinuous transmission", *1998 IEEE.*

[7] Chae Y. Lee and Jun Jo, "Service specific call admission control in WCDMA system".

[8] Kamala Subramaniam, Arne A. Nilsson, "Tier-based analytical model for adaptive call admission control scheme in a UMTS-WCDMA system".

[9] William Stallings, "Wireless communication and Networks", Pearson Education.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Decision Combining in Relay Networks

Sharmin R Ara and R. Viswanathan *

Department of Electrical and Electronic Engineering, East West University, Dhaka, Bangladesh,*Department of Electrical & Computer Engineering , Southern Illinois University at Carbondale, USA
E-mail: shra@ewu.edu.bd,viswa@engr.siu.edu

Abstract - We consider non-coherent detection of *M-ary* FSK modulated signals transmitted over a slow, Rayleigh fading channel in a wireless relay network. The network consists of a single source-destination pair and a number of relays (*L*), which employ cooperative diversity. Performances of a counting rule and square law combiner are studied. We derive closed form expressions for probabilities of error for equal relay channel average SNR. For unequal relay channel SNRs, we resort to Monte Carlo simulations to estimate the error probabilities. We examine different combinations of *M* and *L* for a range of average SNR values. Although the square law combiner outperforms the counting rule for equal and small average SNRs, the loss in performance is not high. Simplicity of counting rule may be advantageous in these cases.

I. Introduction

Cooperative diversity is a popular and effective technique to mitigate fading in wireless network. This scheme could be effectively used for cellular, satellite and certain wireless local area network (LAN). Cooperative diversity exploits the broadcast nature and inherent spatial diversity of the channel [1]-[5], i.e., transmitted signal can be received and processed by any number of terminals. Based on relaying procedure, cooperative diversity can be broadly categorized as fixed relaying and adaptive relaying. Amplify and forward (AF) and decode and forward (DF) belong to the first category. In a relaying protocol, communication takes place in two phases. In first phase, the signals are transmitted by the source to the destination in a broadcast manner. All relays and destination receive faded noisy versions of these signals. In second phase, the relays retransmit a processed version of the received signals to the destination, and finally the destination combines the signals received in two phases. A comprehensive analysis of statistical properties of AF relay channels such as auto-correlation, level crossing rate (LCR), and average outage duration (AOD) were studied in [6]. Combining procedures for *M-ary* hypothesis testing with diversity link appeared in [7]. In an attempt to find channel aware processing that minimizes the error probability at destination node, an iterative algorithm was presented to find relay schemes that are at least locally optimum [4].

In this paper we are considering amplify and forward relays for a single source-to-destination (S-D) pair of network. Assuming non-coherent detection [8], our aim is to find a simple procedure to calculate probability of error at destination. Since detection is non-coherent, arriving signal phase information is not needed. We will restrict our work to *M-ary* orthogonal FSK, which is appropriate for both slow and fast fading, though we will assume a sufficiently slow, Rayleigh fading channel. Also, fading processes on the *L*-channels are assumed to be mutually statistically independent.

II. Counting Rule and Multinomial Distribution

In this section a simple non parametric procedure such as decision combining, as a method of aggregating the information arriving through different channels, is proposed. A single S-D pair with a large number of relays in a wireless network is considered. The relays are of identical capacity and performance. It is assumed that the symbols are sent over a Rayleigh fading channel using non coherent modulation such as M *-ary* FSK; Each channel is assumed to be frequency- nonselective and mutually statistically independent. The signal received is corrupted by additive white Gaussian noise. Hence, for each identical relay channel, the probability of correct symbol reception at D equals $P_c = (1 - P_e)$, where P_e is the probability of symbol error for a Relay link and probability that any one of the symbols, other than the transmitted one, is selected is $P_e/(M-1)$. Given that there are L relay links and given decisions are arrived at by processing the received signals individually, counting the numbers of links that have decided on 1 through M, is a sufficient statistic. Calculating the probability of final correct decision (PCS) can be conveniently arrived at by using results from ranking and selection problems in statistics [9]. Considering a multinomial distribution with M cells, where the cell π_i has probability p_i, for selecting the cell with largest probability $p_{[M]}$ Bechhofer, Elmaghraby and Morse proposed a fixed sample procedure [9].

If $Y_{i,L}$ denotes the number of observations that arise in the cell π_i, the procedure is given by: Select cell π_i for which

$$Y_{i,L} = \max_{1 \le r \le M} Y_{r,L} \quad (1)$$

With the provision that a tie will be broken by randomization, the probability of correct selection (PCS) is given by

$$P_{cs} = \sum_{y_1 \cdots y_M} \frac{1}{s} \frac{N!}{y_1! \ldots y_M!} p^{y_1}[1] \ldots . p_{[M]}{}^{y_M} , \quad (2)$$

$$[\sum_{i=1}^{M} y_i = L,\ y_M \ge y_i, i = 1,\ldots . M-1\ ;],$$

where s denotes the number of ties ,

$$p_{[M]} = P_c , \quad (3)$$

$$p_{[i]} = \frac{(1-P_c)}{M-1}, \ i = 1,2,..M\text{-}1. \quad (4)$$

Now, if we define a parameter to measure the quality of the link such that,

$$\theta = \frac{p_{[M]}}{p_{[i]}}, i = 1,..... M-1 \quad (5)$$

$$\theta = (M-1)Pc/(1-Pc) \quad (6)$$

For a perfect link, θ tends to infinity, and in the worst case, it approaches 1. For *M-ary* in slow Rayleigh fading channel with non coherent detection, the probability of correct symbol decision in a relay is (pp. 834, [8])

$$P_c = 1 - \sum_{m=1}^{M-1} \frac{(-1)^{m+1} \binom{M-1}{m}}{1+m+m\overline{\gamma_c}} \quad (7)$$

where $\overline{\gamma_c}$ is the average SNR of a relay link. Using (2)-(7), we get the probability of correct selection for different sets of hypothesis (M) and relays (L) as :

For $M = 2$

$Pcs = (\theta^2/(\theta+1)^3)(\theta+3)$, L=3 (8.a)

$Pcs = (\theta^2/(\theta+1)^4)(\theta^2+4\theta+3)$, L=4 (8.b)

$Pcs = (\theta^3/(\theta+1)^5)(\theta^2+5\theta+10)$, L=5 (8.c)

For $M = 4$

$Pcs = (\theta/(\theta+3)^3)(\theta^2+9\theta+6)$, L=3 (9.a)

$Pcs = (\theta/(\theta+3)^4)(\theta^3+12\theta^2+45\theta+6)$, L=4 (9.b)

$Pcs = (\theta^2/(\theta+3)^5)(\theta^3+15\theta^2+90\theta+150)$, L=5 (9.c)

For $M = 8$

$Pcs = (\theta/(\theta+7)^3)(\theta^2+21\theta+42)$, L=3 (10.a)

$Pcs = (\theta/(\theta+7)^4)(\theta^3+28\theta^2+273\theta+210)$, L=4 (10.b)

$Pcs = (\theta/(\theta+7)^5)(\theta^4+35\theta^3+490\theta^2+2145\theta+840)$, L=5 (10.c)

Pcs is evaluated for each combination of M and L in order to find the probability of error, which equals $(1-Pcs)$. This method is applicable only when the average SNR of the diversity channels are equal. When average SNR of each channel is different, a simulation is done to find the probability of error. The simulation will count the number of votes in different cells, for a particular hypothesis, and increment error count after each iteration, if the correct hypothesis doesn't get the maximum number of counts. The ties are broken by randomization, i.e., if number of ties equal to three, including the correct one, then probability of picking up the correct one is 1/3 and so on.

III. Square Law Combiner

Though the complex MRC (maximum ratio combiner) is the optimal combiner when channel phase is known, for non coherent detection, square law combiner is optimal when all the relay links are independent and identically distributed as Rayleigh (see Appendix for a proof). The output of the combiner containing the signal (assumed as U_1 without any loss of generality) is [8],

$$U_1 = \sum_{k=1}^{L} | 2\zeta\alpha_k e^{-j\phi_k} + N_{k1} |^2 \quad (11)$$

where $\{\alpha_k e^{-j\varphi_k}\}, \{N_{k1}\}$, are complex valued zero mean Gaussian random variable.While the output of the remaining *M*-1 combiners are :

$$U_m = \sum_{k=1}^{L} | N_{km} |^2, m = 2,3,4....M \quad (12)$$

Proakis [8] did a detailed analysis of the output of the combiner showing that U_1 will have a Chi-square probability density function with $2L$ degrees of freedom, when all diversity channels have equal SNR

$$p(U_1) = \frac{1}{(2\sigma_1^2)^L (L-1)!} U_1{}^{L-1} \exp\left(-\frac{U_1}{2\sigma_1^2}\right) \quad (13)$$

where,

$$\sigma_1^2 = \frac{1}{2} E\left(| 2\zeta\alpha_k e^{-j\varphi_k} + N_{k1} |^2 \right) = 2\zeta N_0 \left(1+\overline{\gamma_c}\right)$$

and $\overline{\gamma_c}$ is the average SNR per diversity channel.

The output of the other combiners, $U_2, \ldots\ldots, U_M$ are identically distributed with the probability density function given by

$$p(U_2) = \frac{1}{(2\sigma_2^2)(L-1)!} U_2{}^{L-1} \exp\left(-\frac{U_2}{2\sigma_2^2}\right) \quad (14)$$

where,

$\sigma_2^2 = 2\zeta N_0$.

The probability of error is simply 1 minus the probability that $\bigcap_{m=2}^{M} U_1 > U_m$.

Now,

$$P(U_2 < U_1) = \int_0^{U_1} p(U_2)\,dU_2$$

$$= 1 - \exp\left(\frac{-U_1}{2\sigma_2^2}\right)\sum_{k=0}^{L-1}\frac{1}{k!}\left(\frac{U_1}{2\sigma_2^2}\right)^k \quad (15)$$

With U_1 fixed, the joint probability $P(U_2 < U_1, U_3 < U_1, \ldots\ldots U_m < U_1)$ is equal to $P(U_2 < U_1)$ raised to the $(M-1)$th power. The $(M-1)$th power of this probability is then averaged over the probability density function of U_1 to yield the probability of correct decision. If this result is subtracted from unity then the probability of symbol error could be written in the following form

$$P_M = 1 - \int_0^{\infty}\left[\left\{\frac{1}{(2\sigma_1^2)^L (L-1)!}U_1^{L-1}\exp\left(\frac{-U_1}{2\sigma}\right)\right\}\times \left\{1 - \exp\left(\frac{U_1}{2\sigma_2^2}\right)\sum_{k=0}^{L-1}\frac{1}{k!}\left(\frac{U_1}{2\sigma_2^2}\right)^k\right\}^{M-1}\right] dU_1 \quad (16)$$

Using Binomial expansion for the (M-1)th power transform, P_M could be expressed as

$$P_M = \left[\frac{1}{(L-1)!}\sum_{m=1}^{M-1}\frac{(-1)^{m+1}\binom{M-1}{m}}{(1+m+m\overline{\gamma_c})^L}\times \sum_{k=0}^{m(L-1)}\beta_{km}(L-1+k)!\left(\frac{1+\overline{\gamma_c}}{1+m+m\overline{\gamma_c}}\right)^k\right] \quad (17)$$

With no diversity (L=1), the error probability reduces to the simple form

$$P_M = \sum_{m=1}^{M-1}\frac{(-1)^{m+1}\binom{M-1}{m}}{1+m+m\overline{\gamma_c}} \quad (18)$$

To apply the result stated in (17) we need to calculate the coefficient β_{km} for different sets of M and L, using the following equation

$$\left(\sum_{k=0}^{L-1}\frac{U_1^k}{k!}\right)^m = \sum_{k=0}^{m(L-1)}\beta_{km}U_1^k \quad (19)$$

Now we will concentrate for set of diversity channels that have different SNR. If γ_b is the sum of L statistically independent components γ_k, which is the instantaneous SNR of k^{th} channel, the probability density function of γ_b can be written as,

$$p(\gamma_b) = \sum_{k=1}^{L}\frac{\pi_k}{\overline{\gamma_k}}e^{\frac{-\gamma_b}{\overline{\gamma_k}}} \quad (20)$$

where π_k is defined as,

$$\pi_k = \prod_{\substack{i=1\\ i\neq k}}^{L}\frac{\overline{\gamma_k}}{\overline{\gamma_k} - \overline{\gamma_i}}$$

and $\overline{\gamma_k}$ is the average SNR of the k^{th} channel.

The probability of symbol error, conditioned on a specific γ_b, could be expressed as

$$P_M = \sum_{n=1}^{M-1}(-1)^{n+1}\binom{M-1}{n}\frac{1}{n+1}\exp\left[\frac{-nc\gamma_b}{(n+1)}\right] \quad (21)$$

where $c = \log_2 M$. By unconditioning P_M with respect to probability density function of γ_b, we get the final probability of error

$$P_e = \int_0^{\infty}\left(\sum_{n=1}^{M-1}(-1)^{n+1}\binom{M-1}{n}\frac{1}{n+1}\exp\left[\frac{-nc\gamma_b}{(n+1)}\right]\right)p(\gamma_b)d\gamma_b \quad (22)$$

Therefore,

$$P_e = \sum_{n=1}^{M-1}(-1)^{n+1}\binom{M-1}{n}\sum_{k=1}^{L}\frac{1}{1+n+nc\overline{\gamma_k}}\prod_{\substack{i=1\\ i\neq k}}^{L}\frac{\overline{\gamma_k}}{\overline{\gamma_k} - \overline{\gamma_i}} \quad (23)$$

IV. Numerical Results

Two cases of channels having equal SNR and channels having unequal SNR are considered. The counting rule method is compared with the results obtained from square law combiner. Number of hypothesis (M) for which this numerical evaluations are done are two, four or eight. Number of diversity channels (L) are three, four or five.

A. Channels with equal SNR

Graphs are plotted for same M and different L in the same figure, so they can be compared with respect to diversity. We have applied theoretical results obtained from equations (8), (9), (10) for counting rule and (17) for square law combiner. For square law combiner we needed to calculate the coefficients β_{km}, for each set of M and L using equation (19), which becomes tedious for M = 8, and hence M = 8 is omitted from the calculations. Average channel SNR values are assumed to range from 6 dB to 16 dB. Fig. 1 and Fig. 2 show the probability of error comparison of the two methods. For all cases, the

probabilities of error using counting rule are higher than those achieved using square law combiner, although not significantly large when average channel SNR is low. Also, fewer number of hypothesis gives lower error rate, which is true both for counting rule and square law. As the number of relays increases, the probability of error decreases irrespective of the number of hypotheses , both in counting rule and square law.

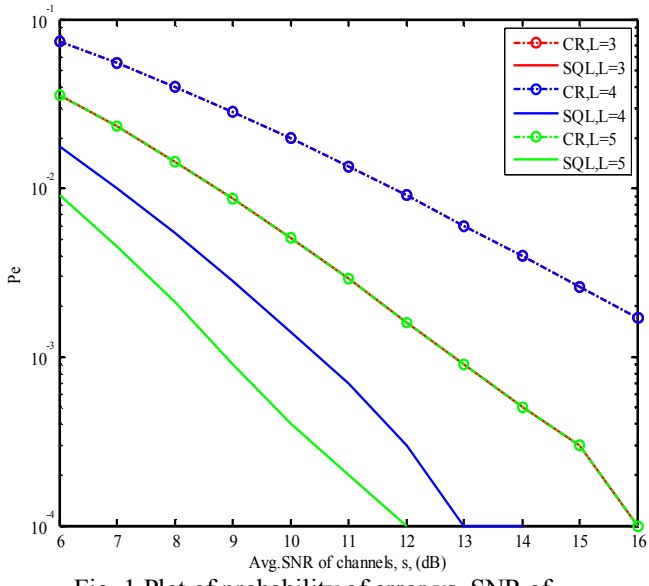


Fig. 1 Plot of probability of error vs. SNR of channels for M=2

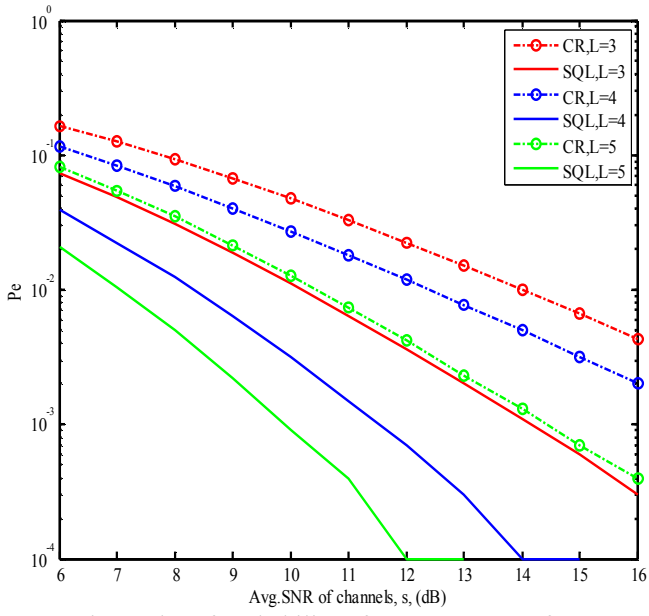


Fig. 2 Plot of probability of error vs. SNR of channels for M=4

B. Channels with unequal SNR

To calculate probability of error for channels having unequal SNR, instead of theoretical approach simulation is done in counting rule method. The channel SNRs are assumed to be,

$$[s-2 \quad s \quad s+2], \qquad L=3$$

$$[s-3 \quad s-1 \quad s+1 \quad s+3], \qquad L=4$$

$$[s-4 \quad s-2 \quad s \quad s+2 \quad s+4], \qquad L=5$$

where s is the average of the SNRs of all channels, which ranges from 6 dB to 16 dB. The spacing between SNR of adjacent channels is 2 dB. Another set of calculations are done for channels spaced 4 dB apart.

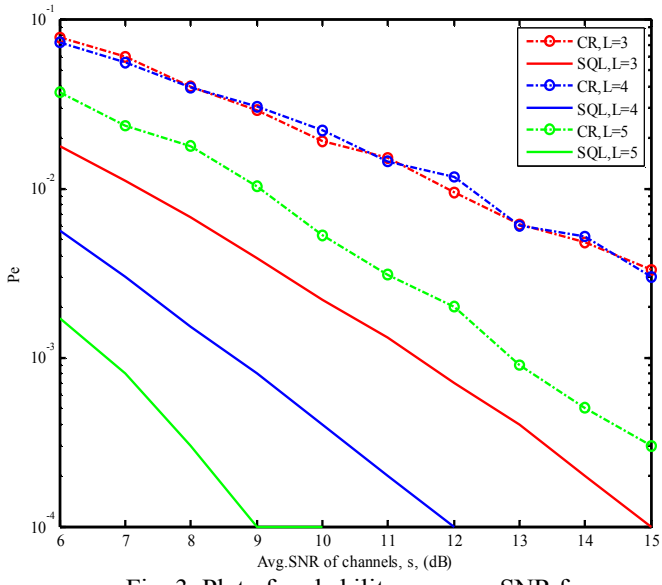


Fig. 3 Plot of probability error vs. SNR for $M=2$, and unequal relay SNRs are 2dB apart

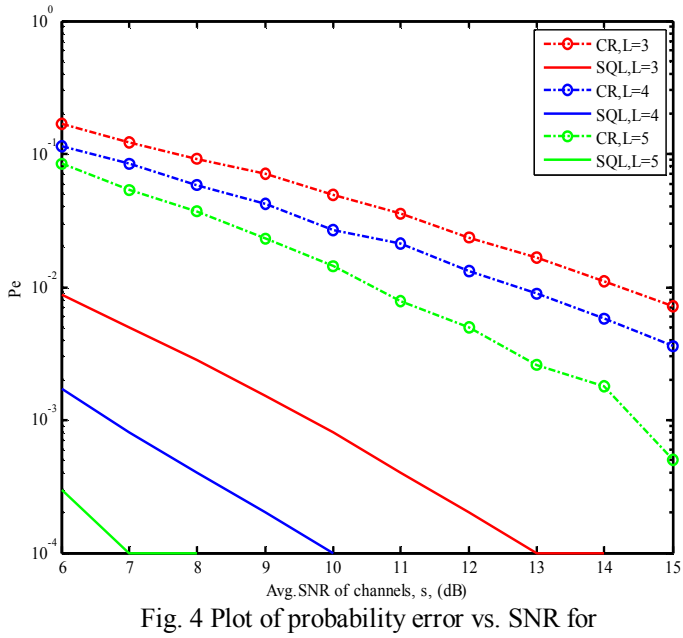


Fig. 4 Plot of probability error vs. SNR for $M=4$, and unequal relay SNRs are 2dB apart

Fig. 3 through Fig. 8 show comparison of error performances of counting rule and square law.
Clearly, for the unequal average SNR case, the square law provides significantly smaller probability of error than the counting rule, for all SNRs, including the small values.
Fig. 3 and Fig. 6 show that for M=2, counting rule gives almost identical probabilities of errors for both L=3 and L= 4, as in the case of equal SNR channel. However, the

results obtained for L=3 and L=4 using square law are quite different. Square law performs much better than counting rule for L=4 and L=5.

V. Conclusion

In this paper we compared two combining procedures, counting rule and square law, for M-ary FSK detection in relay networks. The network consists of a single source - destination pair and L number of relays. The transmission channel is assumed to be slow Rayleigh fading channel and signals are corrupted by AWGN. As expected, the square law performs better for all cases. For equal average channel SNRs, error rates are not too far apart when the diversity channel has small average SNR. Simplicity of counting rule may be advantageous in such cases. Counting rule performs much inferior to square law combiner for unequal relay channel average SNRs as compared to the equal SNR channel.

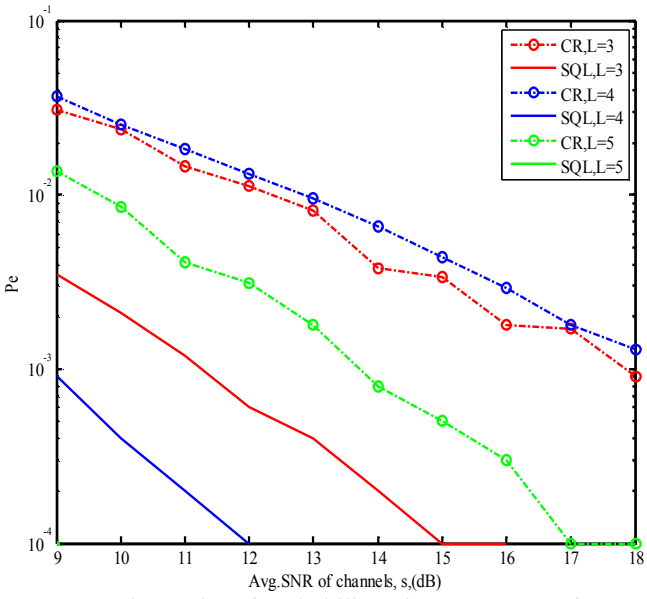


Fig. 6 Plot of probability of error vs. SNR for M = 2, unequal relay SNRs are 4dB apart

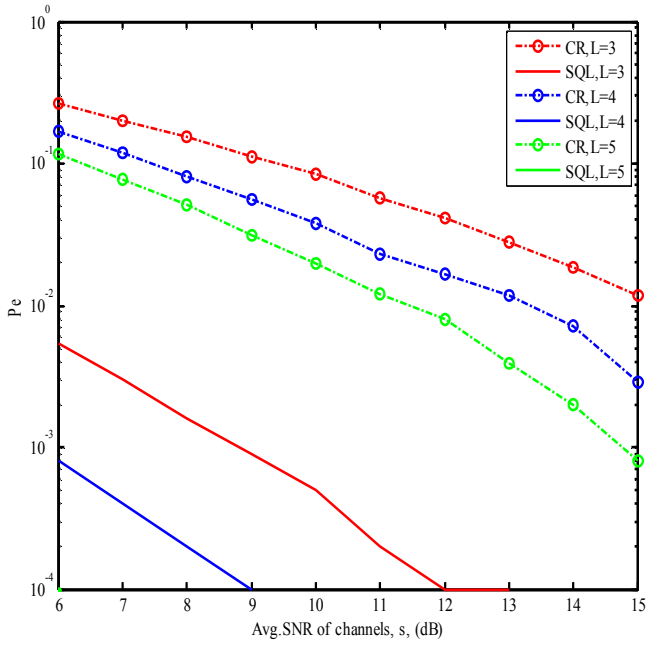


Fig. 5 Plot of probability of error vs. SNR for M = 8, unequal relay SNRs are 2dB apart

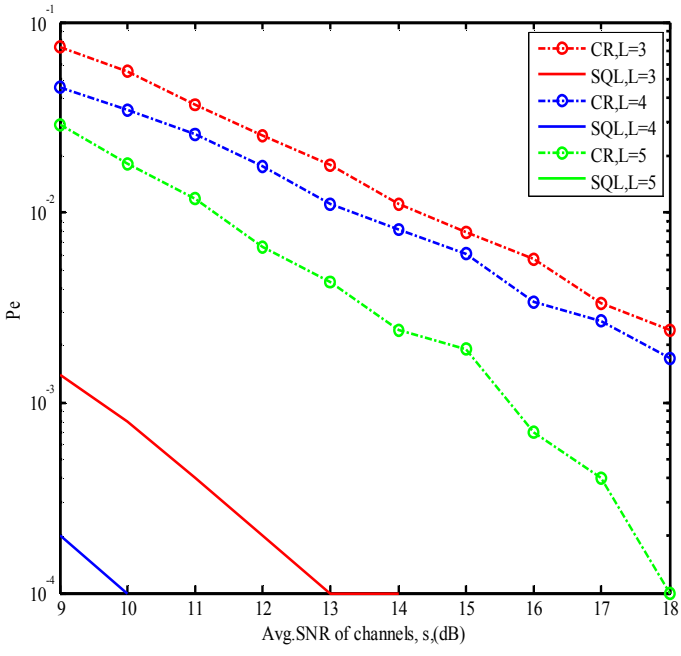


Fig. 7 Plot of probability of error vs. SNR for M = 4, unequal relay SNRs are 4dB apart

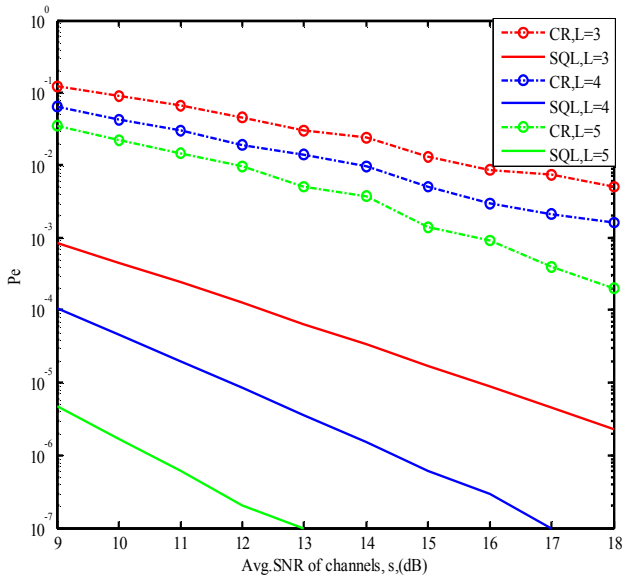


Fig. 8 Plot of probability error vs. SNR for $M = 8$, unequal relay SNRs are 4dB apart

VI. References

[1] Theodore S. Rappaport, *Wireless Communications : Principles and Practice*, Prentice Hall, 2nd Edition, 2002.

[2] J. Nicholas Laneman, *Ph.D. Dissertation*, MIT, Department of Electrical Engineering and Computer Science, 2002.

[3] Andrew Sendonaris, Elza Erkip and Behnaam Aazhang, "User cooperation diversity-part I : System description ", pp. 1927-1938, "User cooperation diversity – Part II : Implementation aspects and performance analysis" *IEEE Transactions on Communication*, Vol. 51, No. 11, pp. 1939-1948, November 2003.

[4] Bin Liu, Biao Chen, Rick S. Blum, "Minimum error probability cooperaitve relay design", *IEEE Transaction on Signal Processing*,Vol. 55, No. 2, pp. 656-664, February 2007.

[5] Naveen Shastry, Raviraj S. Adve, " A theoretical analysis of cooperative diversity in wireless sensor networks", *IEEE GLOBECOM*, pp. 3269-3273, 2005.

[6] Chirag S. Patel, Gordon L. Stuber and Thomas G. Pratt, "Statistical properties of amplify and forward relay fading channels", *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 1, pp. 1-9, January 2006

[7] R. Niu, B. Chen, and P. K. Varshney, "Decision fusion rules in wireless sensor networks using fading channel statistics", in *IEEE Transaction on Signal Processing*, Vol. 54, No. 3, pp. 1018-1027, March, 2006.

[8] John G. Proakis, *Digital Communication*, Mc Graw Hill, 4th edition, 2001.

[9] Shanti S. Gupta, S. Panchapakesan, *Multiple Decision Procedure*, SIAM, Philadelphia, 2002

VII. APPENDIX

Let U_m denote the square law output from the mth frequency filter in the lth relay.

$$U_m = \sum_{l=1}^{L} U_{ml}, m = 1,2,..,M \qquad \text{(i)}$$

Assuming U_k contains the signal, the maximum likelihood test picks

$$\max_k \; p(U_{11},..,U_{1L},U_{21},..,U_{2L},...,U_{M1},...,U_{ML} \mid H_k)$$

$$= \max_k \left((\frac{1}{2\sigma_1^2}\exp(-\sum_{l=1}^{L}\frac{U_{kl}}{2\sigma_1^2})\prod_{\substack{j=1\\ j\neq k}}^{M}\frac{1}{2\sigma_2^2}\exp(-\frac{1}{2\sigma_2^2}\sum_{l=1}^{L}U_{jL}) \right)$$

$$= \min_k \left(\sum_{l=1}^{L}\frac{U_{kl}}{2\sigma_1^{\,2}} + \sum_{\substack{j=1\\ j\neq k}}^{M}\frac{\sum_{l=1}^{L}U_{jl}}{2\sigma_2^{\,2}} \right) \qquad \text{(ii)}$$

If we define θ as $\theta = SNR + 1$, and $U_j = \sum_{l=1}^{L} U_{jl}$, then equation (ii) becomes

$$= \min\left(\frac{1}{\theta}U_1 + \sum_{\substack{j=1\\ j\neq 1}}^{M}U_j, \frac{1}{\theta}U_2 + \sum_{\substack{j=1\\ j\neq 2}}^{M}U_j, .., \frac{1}{\theta}U_M + \sum_{\substack{j=1\\ j\neq M}}^{M}U_j \right) \qquad \text{(iii)}$$

Let us suppose that kth term is the minimum. In order that the receiver picks correctly the kth hypothesis

$$\frac{1}{\theta}U_k + \sum_{\substack{j=1\\ j\neq k}}^{M}U_j < \frac{1}{\theta} + \sum_{\substack{j=1\\ j\neq 1}}^{M}U_j$$

$$\frac{1}{\theta}U_k + \sum_{\substack{j=1\\ j\neq k}}^{M}U_j < \frac{1}{\theta} + \sum_{\substack{j=1\\ j\neq 2}}^{M}U_j$$

$$\vdots$$

$$\frac{1}{\theta}U_k + \sum_{\substack{j=1\\ j\neq k}}^{M}U_j < \frac{1}{\theta} + \sum_{j=1}^{M-1}U_j$$

Or

$$U_1(1-\frac{1}{\theta}) < U_k(1-\frac{1}{\theta})$$

$$U_2(1-\frac{1}{\theta}) < U_k(1-\frac{1}{\theta})$$

$$\vdots$$

$$U_M(1-\frac{1}{\theta}) < U_k(1-\frac{1}{\theta})$$

Since $\theta > 1$, i.e., $\frac{1}{\theta} < 1$

$U_k > U_1$, $U_k > U_2$,..,$U_k > U_M$.

i.e., U_k is the maximum among $(U_1, U_2 .., \ldots U_M)$.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Investigation on Stochastic Tap Delay line Model of UWB Indoor Channel

Jyoteesh Malhotra[1], Ajay K. Sharma[2], R.S Kaler[3]

[1]Department of Electronics & Communication Engineering, G..N.D.U. Regional Campus, Jalandhar
[2]Department of Computer Science & Engineering, National Institute of Technology, Jalandhar
[3]Department of Electronics & Communication Engineering, Thapar University, Patiala
E-mail: jyoteesh@rediffmail.com[1], sharmaajayk@rediffmail.com[2], rskaler@yahoo.com[3]

Abstract - The Stochastic Tap Delay line model of the UWB indoor channel have been used for generating Power Delay Profile (PDP). The local Power Delay profiles for different categories of channel types with range of T-R distance separation as per IEEE 802.15.4a standard have been generated. The important attributes of PDP have been selected for investigating the relative Channel behavior. The Complementary CDFs of mean excess delay and root mean square (RMS) delay spread have been computed for the four channel categories which include CM1 (0-4m LOS), CM2 (0-4m NLOS), CM3 (4-10m NLOS) and extreme NLOS (6-17m). The mean and standard deviation of the ensemble of PDPs at various locations have also been evaluated through simulations.

I. Introduction

In recent years, Ultra Wide-Band (UWB) communications has received great interest from both the research community and industry. The potential strength of the UWB radio technique lies in its use of extremely wide transmission bandwidths, which results in desirable capabilities including accurate position location and ranging, lack of significant fading, high multiple access capability, covert communications, and possible easier material penetration. UWB communications are best suited for short-range communications: Sensor Networks and Personal Area Networks (PANs). In order to build wireless systems that utilize the UWB potential, it is first required to understand Multipath delay profiles for UWB channel characterizations. There has been a enormous activity for measuring delay profiles, undertaken in the last couple of years [1-7]. The three delay profile models widely referenced in the literature are Stochastic tapped delay line model (STDL) [2, 3], multi-cluster model [7, 8] and exponential-lognormal model [1, 4]. All the delay profile models are based on certain parameters characterizing the channel category. The STDL model is least complicated as we need only two parameters to predict the channel behaviour. Moreover, it captures the occurrence of LOS components in many profiles. The cluster and exponential-lognormal models need 5 and 13 parameters to specify the channel category, respectively. Also, the IEEE 802.154a standard for low data rate applications with high precision ranging capability has adopted the STDL due to its high merits. Thus STDL model has been selected for investigation. This paper is organized as follows. In the next section, we will give a description of various Multipath UWB models based on channel sounding measurements along with related parameters. Following this we will describe the Stochastic Tap delay line model to be used in the performance analyses of PAN physical layer proposals. The simulation methodology to generate the Multipath intensity profile of different residential environments using their measured numeric parameters [5, 6] is described in section IV. A set of statistical characteristics that capture the variability of the delay profile over many TR paths and buildings, e.g., the distribution of the mean excess delay, RMS delay spread, ensemble mean PDP and standard deviation of PDP are identified in Section V. Simulation results for the attributes computed of the model are also discussed therein, followed by Section VI, wherein we conclude the paper.

II. UWB delay profile Models for PAN

We regard a delay profile model as consisting of three parts:

- A functional description of power vs. delay.
- A set of probability distributions for those function parameters that vary across T-R paths and/or buildings.
- A set of numerical values for both the function parameters that are fixed and the parameters of the probability distributions.
- The first two parts comprise the structure of the model, and the third comprises the numeric(s).

There has been a great deal of activity on the characterization of multipath delay profiles for UWB channels [4-7]. The widely referenced model postulates a single cluster with an exponential profile [2, 3]. A variation on exponential model based approach and based on an extensive data collection, an exponential profile multiplied by a noise-like variation with log-normal statistics has been proposed [4-6]. The model proposed by Intel for high rate wireless personal area networks is based on the existence of ray clusters, each cluster having an exponential profile of ray power vs. delay [7]. Although there are both frequency domain and time domain models that may be appropriate for UWB systems, we chose to focus our work on evaluating discrete time models. These

models are based upon the following channel impulse response model:

$$h(t) = \sum_{l=0}^{L-1} \alpha_l \delta(t - \tau_l) \quad (1.1)$$

where is the amplitude fading factor on path l (could be complex), is the random delay of path l, L is the number of Multipath components, and is the Dirac delta function. In a given bandwidth, W, sampling theory tells us that the impulse response $h(t)$ (and, by extension, the delay profile) is completely determined by a set of samples spaced by 1/W or less. Therefore, one way to characterize a PDP is via a set of samples spaced by $1/W$ ($\tau_i = i/W$, $i = 0, 1, 2, ..$.), and the result is suitable for any bandwidth of W or smaller. Alternatively, one can model the physical Multipath echoes, which may arrive at arbitrary delays bearing no relationship to integer multiples of $1/W$. The cluster model takes the latter approach, while the exponential and exponential-lognormal models take the former approach. The particular approach used may have implications for simulation, design, generality, and so on, but both are valid. There are several parameters that need to be defined to complete this particular model, and each will be addressed in the following subsections.

A. Delay spread

Typical values for the Multipath delay spread of indoor channels have been reported to be between 15 nsec in a residence to over 100 nsec in an office to a 150 nsec in a commercial building. Other measurements at 10-meter distances suggest RMS delay spreads of 19-47 nsec [9]. Trying to stick with those published in [10], suggest that a fairly conservative RMS delay spread of 25 nsec would be a good initial starting point for PAN type applications with antenna separations of about 10 meters or less. Shorter RMS delay spreads could be considered for shorter ranges (5 meters or less). Results in [10] (few measurements have been done with such short range), found average RMS delay spreads of around 17 nsec at these short ranges. As a result of this wide variation, the final Multipath model should consider a range of RMS channel delay spreads.

B. Large scale Amplitude Statistics

The large-scale fading characterizes the changes in the received signal when the receiver position varies over a significant fraction of the transmitter–receiver (T–R) distance and/or the environment around the receiver changes. This situation typically occurs when the receiver is moved from one room to another room in a building.

C. Small scale Amplitude Statistics

The small-scale effects, on the other hand, are manifested in the changes of the PDP (Power Delay Profile) caused by small changes of the receiver position, while the environment around the receiver does not change significantly. This occurs, for instance, when the receiver is moved over the measurement grid within a room in a building. The small-scale fading causes the differences between the PDPs at the different points of the measurement grid. In "narrowband" models, it is usually assumed that the magnitude of the first (quasi-LOS) Multipath component follows Rician or Nakagami statistics and the later components are assumed to have Rayleigh statistics [11]. However, in UWB propagation each resolved MPC is due to a small number of scatterers, and the amplitude distribution in each delay bin differs markedly from the Rayleigh distribution. The number of Multi path components falling into each resolvable bin is much smaller, and it has been empirically determined that in many environments, alternative amplitude distributions must be used such as Nakagami distribution. This distribution has been used to model the magnitude statistics in mobile radio when the conditions of the central limit theorem are not fulfilled [13]. In fact, the presented analysis showed that the best-fit distribution of the small-scale magnitude statistics is the Nakagami distribution [12], corresponding to a Gamma distribution of the energy gains. Lognormal distribution has also been suggested for use in UWB by [7].

III. Stochastic Tapped Delay line Model

The database for this model [3] is a set of measurements made in a 500-MHz bandwidth using baseband pulses. The measurements were conducted at 14 locations within one office building, with transmit-receive distances ranging from ~6 m to ~17 m. Of the 14 paths measured, two were LOS and 12 were NLOS, and all were statistically modelled as one population. All Small Scale Averaged-PDPs (SSA-PDP) exhibit an exponential decay as a function of the excess delay. As a delay axis translation is used, the direct path always falls in the first bin in all the PDPs. The energy of the subsequent MPCs decay exponentially with delay starting from the second bin. The average energy of the second MPC may be expressed as 'r' a fraction of the average energy of the direct path. The PDP as a function of delay (τ) is represented by discrete samples spaced by $1/W$ i.e.

$$r = e^{\left(-\frac{\tau}{\varepsilon}\right)} \quad (1.2)$$

where τ and ε are in ns and r is the ratio of energy gains relative to first bin .

$$\overline{G_K} = \begin{cases} \dfrac{G_{tot}}{1+rF(\varepsilon)} & for...k=1 \\ \dfrac{G_{tot}}{1+rF(\varepsilon)} r.e^{(-(\tau_k-\tau_2)/\varepsilon)} & for...k=2,...N_{bins} \end{cases} \quad (1.3)$$

Thus, SSA-PDP is completely characterized by averaged energy gain (G_{tot}), the power ratio r, and the decay constant ε. The structure of this PDP model includes specifying how r and ε are distributed over the T-R paths. The assumption made is that both are lognormal, i.e., 10 log r and 10 log ε are both Gaussian random variates over the population of all possible TR paths. The numerics of the model then consist of specifying the mean and standard deviation of these decibel quantities. For the original building measurements [3], the means were -4 and -16.1, respectively, while the standard deviations were 3 and 1.27, respectively. Another set have computed

moments for the measurements carried out in [6] for three channel category and computed parameters are tabulated in table 1.

Table 1 Parameters of the STDL model [6].

Parameter	*CM1*	*CM2*	*CM3*
Mean (power ratio)	-7.04	-2.15	-2.56
Std.deviation (power ratio)	2.65	1.44	2.07
Mean (Decay constant)	7.31	8.07	8.74
Std.dev. (Decay constant)	1.63	0.99	1.81

The small-scale fading causes the differences between the PDPs at the different points of the measurement grid. The variations over the normalized energies in each bin are treated as stochastic Gamma distributed and the statistics (mean, variance) of the energy gain vary with delays.

$$G_K \sim \Gamma\left(\overline{G}_k; m_k\right) \tag{1.4}$$

The parameters m_k, $\overline{G}_k$ in each bin of the Gamma distributions themselves are random variables distributed according to a truncated Gaussian distribution i.e. their distribution looks like a Gaussian for $m > 0.5$ and zero elsewhere.

$$m_k = \Gamma_N\left(\mu_m, \sigma_m^2\right) \tag{1.5}$$

We summarize the exponential model as follows. The average energy gains $\overline{G}_k$ vary both with the excess delay and the large-scale conditions, while the m_k parameters depend only on the excess delay. Given the $\overline{G}_k$ and the m_k, the normalized energy gains of the MPCs arriving in the same room at different excess delays are independent realizations of the gamma distribution.

IV. Simulation Methodology

The simulation method implementation [3] has been depicted in Figure 1. We start out by generating large-scale statistics. First the total mean energy at a certain distance'*d*' is assumed for simplicity as 1. Next, we generate the decay constant and the power ratio as lognormal distributed random numbers. Thus, the SSA-PDP (the average energy gains in a specified delay bin) is completely specified and generated using (1.3). The small scale PDPs also called, as small-scale variations about the SSA-PDP will be generated. We generate the local PDPs by computing the normalized energy gains of every bin and every location as Gamma distributed independent variables. For generating the normalized energy gains required parameter 'm_k' is generated for every delay bin that is found to be another random variable distributed as truncated Gaussian. Thus, normalized gains generated as gamma distribution function of average energy gains and severity parameter m_k.

V. Simulation Results & Discussion

The ultimate test of a PDP model is whether an ensemble of profiles generated from that model yields the same link performance results as the ensemble of profiles generated

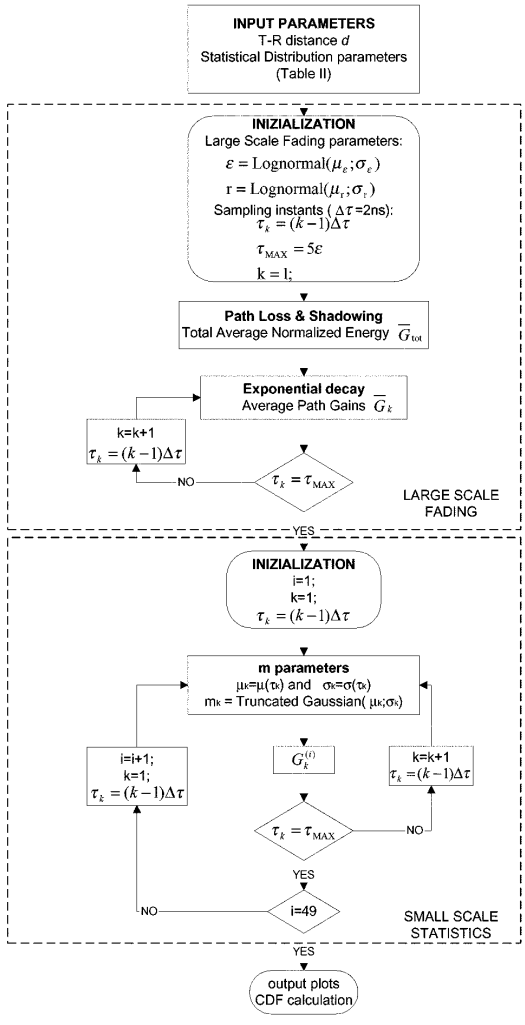


Fig. 1 Simulation Procedure for generating Local PDP

directly from the data. We will compute certain statistical characteristics of the ensemble of PDPs. The starting point in defining characteristics is the PDP at a given location, *p(τ)*. Assuming a statistically generated ensemble of such PDPs, {*p (τ)*}, we compute the following :

- The cumulative distribution function (CDF) of the mean delay, τ_m, across the ensemble.
- The CDF of the rms delay spread, τ_{rms}, across the ensemble.
- The mean across the ensemble of *p(τ),* denoted by $\mu_{p(\tau)}$.
- The standard deviation across the ensemble of *p(τ)*, denoted by $\sigma_{p(\tau)}$.

We computed each of the above four characteristics for each of the four environment [4] categories CM1 (0-4m LOS), CM2 (0-4m NLOS), CM3 (4-10m NLOS) and extreme NLOS environment. In this section, we will first present the results obtained from 40 simulated multipath profiles, generated based on the statistical model described above, in which the local variables are changed according to the small-scale statistics, and the "global" variables such as the total averaged received energy, the decay constant, and the power ratio are fixed. Simulated Power Delay profile gains were obtained and plotted in figure 2.

Figure 6 shows CDFs of mean delay, τ_m, for categories CM1, CM2, CM3, and Extreme NLOS environments and Figure 7 do the same for the CDFs of RMS delay spread τ_{rms}. Figure 8 shows the ensemble mean of the PDP, μ_p, as a function of τ; for all the four categories of channels and Figure 9 do the same for the ensemble standard deviation of the PDP, σ_p. The 40 simulated local PDPs are generated assuming the same distance from the transmitter. The parameters related to the large-scale statistics (i.e., the total averaged received energy, the decay constant, and the power ratio) were set equal to the corresponding parameters extracted from the measurement data [2] described in Table 1. The first order and second order parameters as extracted from statistical analysis of UWB channel model are tabulated in Table 2.

Table 2: Extracted values of Parameters using STDL model in different environments.

Channel	*Mean Delay*	*RMS spread*
CM1	4.3706 ns	4.3602 ns
CM2	7.0681 ns	6.0937 ns
CM3	8.2432 ns	7.1518 ns
Extreme NLOS	33.5233 ns	32.5009 ns

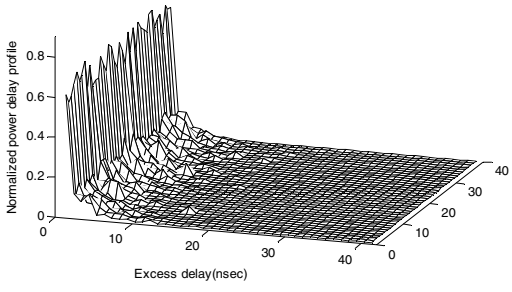


Fig. 2 Simulated Local PDP in CM1 environment using Stochastic Tap Delay line model

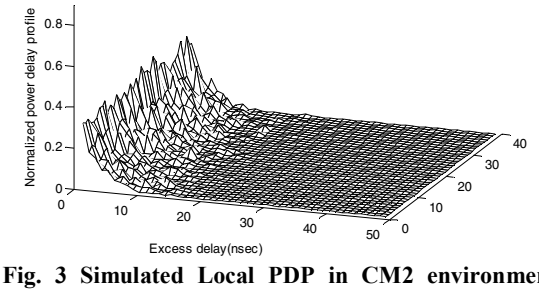


Fig. 3 Simulated Local PDP in CM2 environment using Stochastic Tap Delay line model

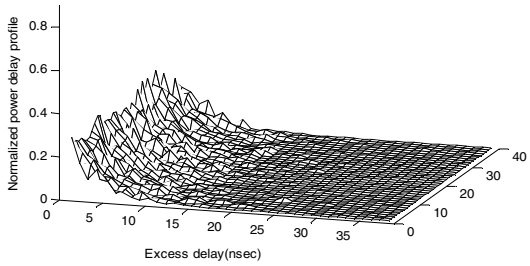


Fig. 4 Simulated Local PDP in CM3 environment using Stochastic Tap Delay line model.

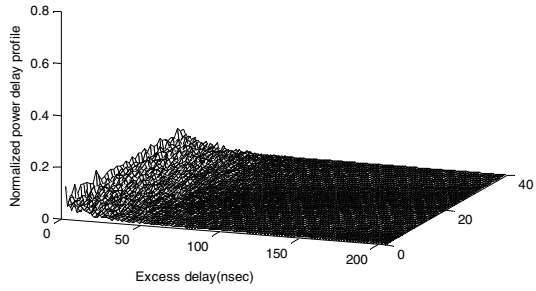


Fig. 5 Simulated Local PDP in extreme NLOS environment

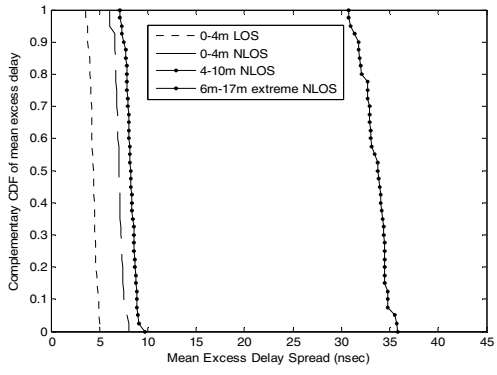


Fig. 6 Complementary CDF of Mean Excess delay in diversified indoor environment categories

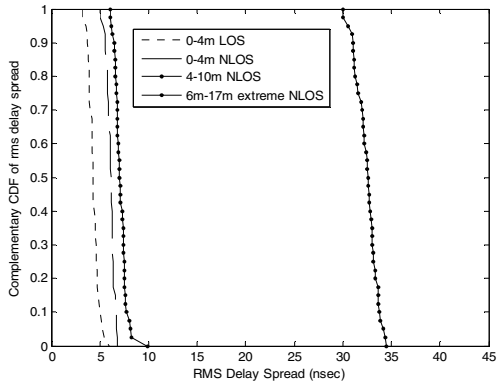


Fig. 7 Complementary CDF of RMS delay spread in diversified indoor environment categories

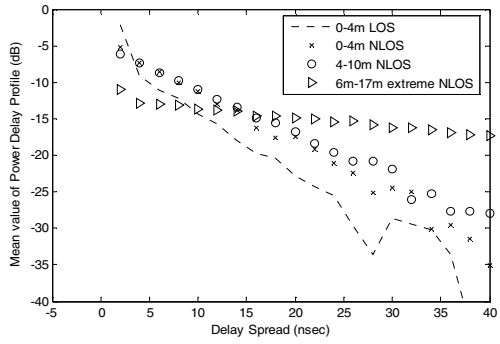


Fig. 8 Ensemble mean of PDP in diversified indoor environment categories using Stochastic Tap Delay line model.

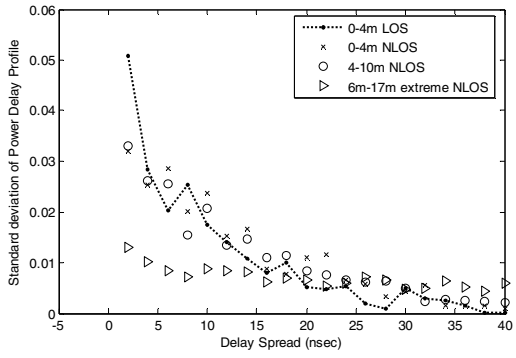


Fig. 9 Ensemble Standard deviation of PDP in diversified indoor environment categories using Stochastic Tap Delay line model

It has been observed that a specular component resulted in relatively more power in the initial delay bins in CM1 (0-4m) environment. CM2, CM3 and extreme NLOS environments have increasing trend of delay spread and absence of specular component in their PDP. The variation of the first and second order statistics of PDP i.e. ensemble mean and standard deviation has been found to be reduced with channel environment changes from CM1 to extreme NLOS. The PDP is exponentially decaying in the beginning in all channel scenarios. This is due to the fading severity parameter m_k of gamma distribution is greater than 1 for the amplitude of the first several arrival delay bins.

VI. Conclusions

The analysis of UWB channel model shows that the well-established tapped-delay-line model, with independent fading of the taps (bins), accurately reproduces the behaviour of the measured channel. In contrast to narrowband models, the energy statistics due to small-scale effects follow a Gamma distribution for all bins, with the '*m*' factor decreasing with increasing excess delay. The variations of the large-scale parameters, such as the total averaged energy, decay constant, and ratio of the energies in the first and second bin, can be modelled as stochastic parameters that change, e.g., from room to room. Thus, the STDL model has the virtues that it is simple, requires relatively few parameters, and captures the occurrence of LOS components in many generated profiles.

References

1. Larry J. Greenstein, et al., "Comparison Study of UWB Indoor Channel Models," IEEE transactions on wireless communications, vol. 6, no. 1, pp. 128-135 January 2007.
2. D. Cassioli, M. Z. Win, and A. F. Molisch, "A statistical model for the UWB indoor channel," in Proc. 53rd IEEE Vehicular Technology Conference, vol. 2, May 2001, pp. 1159–1163.
3. D Cassioli,., M. Z Win, F Vatalaro, A Molisch, "Low Complexity Rake receivers in Ultra Wideband Channels", IEEE Transactions on Wireless Communications, Vol. 6, pp.1265-1275, April 2007.
4. S S Ghassemzadeh, L J Greenstein, T Sveinsson, A Kavcic, and V Tarokh, "UWB indoor delay profile model for residential and commercial environments," IEEE Veh. Technol. Conf.-Fall, Oct. 2003, vol. 5, pp. 3120-3125.
5. S S Ghassemzadeh, R. Jana, C. W. Rice W. Turin, and V. Tarokh, "Measurement and modeling of an indoor UWB channel," IEEE Trans.Commun., vol. 52, no. 10, pp. 1786-1796, Oct. 2004.
6. S S Ghassemzadeh, L. J. Greenstein, T. Sveinsson, and V. Tarokh, "UWB delay profile models for residential and commercial indoor environments," IEEE Trans. Veh. Technol., vol. 54, no. 4, pp. 1235-1244, July 2005.
7. J. Foerster, "Channel modeling sub-committee report final," IEEE P802.15-02/490r1, Feb. 2003.
8. A. F. Molisch, J. R. Foerster, and M. Pendergrass, "Channel models for ultra-wideband personal area networks," IEEE Wireless Commun. Mag., vol. 10, no. 6, pp. 14-21, Dec. 2003.
9. K Pahlavan and A Levesque, Wireless Information Networks, John Wiley and Sons, 2006.
10. H Hashemi, "Impulse Response Modeling of Indoor Radio Propagation Channels," IEEE JSAC, Vol. 11, No. 7, Sept. 1993, pp. 967-978.
11. E Failli, Ed., Digital Land Mobile Radio. Final Report of COST 207, Luxemburg: Commission of the European Union, 1989.
12. M Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in Statistical Method in Radio Wave Propagation, W. C. Hoffman, Ed. Oxford, U.K.: Pergamon, 1960, pp. 3–36.
13. W R Braun and U. Dersch, "A physical mobile radio channel model," IEEE Trans. Veh. Technol., vol. 40, pp. 472–482, May 1991.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Secrecy capacity of MIMO channels

Mohammad Rakibul Islam [1], Jinsang Kim [1], Md. Shamsul Arefin [2]
[1]*Dept. of Electronics and Radio Engineering, Kyung Hee University*
1 Seocheon, Kihung, Yongin, Gyeonggi, 449-701, Korea
[2]*Dept. of Electrical and Electronic Engineering*
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
Email: rakibultowhid@yahoo.com

***Abstract*—A Gaussian multiple input multiple output (MIMO) channel is considered where a transmitter is communicating to a receiver in the presence of an eavesdropper. The transmitter is equipped with multiple antennas, while the receiver and the eavesdropper also contains multiple antennas. We present a technique for determining the secrecy capacity of the MIMO channel under Gaussian noise. To do so, we transform the channel into multiple single input multiple output (SIMO) Gaussian wiretap channel and then use scalar approach using standard techniques of communications theory.**

***Index Terms*—Secret communication, MIMO, wiretap channel, capacity, entropy.**

I. Introduction

Wireless communications is open from its inherent property and is vulnerable to eavesdropping and jamming attacks. This vulnerability leads us to conserve secure communications. The eavesdropping attack was first studied by Wyner in [1], where he considers a single-user wire-tap channel. The measure of secrecy is the message equivocation rate at the wire-tapper, which is defined as the entropy of the message at the wire-tapper, given the wire-tappers observation. Wyner models the wire-tappers channel as a degraded version of the channel from the transmitter to the legitimate receiver, which is a reasonable assumption in a wired channel. For this channel, Wyner identifies the rate-equivocation region and therefore, the secrecy capacity. Wyners result was extended to the Gaussian wire-tap channel in [2], and it was shown that Gaussian signalling is optimal. The secrecy capacity was found to be the difference between the capacities of the main and the eavesdropping channels. Csiszar and Korner [3] studied the general, i.e., not necessarily degraded, single-transmitter, single-receiver, single-eavesdropper, discrete memoryless channel with secrecy constraints, and found an expression for the secrecy capacity, in the form of the maximization of the difference between two mutual informations involving an auxiliary random variable. The auxiliary random variable is interpreted as performing pre-processing on the information. The explicit calculation of the secrecy capacity for a given channel requires the solution of this maximization problem in terms of the joint distribution of the auxiliary random variable and the channel input. The use of multiple transmit and receive antennas has been shown to increase the achievable rates when there are no secrecy constraints [4]. The Gaussian multiple-input multipleoutput (MIMO) wire-tap channel is a special case of the single-transmitter, single-receiver, single-eavesdropper wire-tap channel. Since the Gaussian MIMO channel is not degraded in general, finding its secrecy capacity involves identifying the optimum joint distribution of the auxiliary random variable representing pre-processing and the channel input in the Csiszar-Korner formula. However, solving this optimization problem directly for non-degraded channels is difficult, forcing researchers typically to follow a two-step solution, where in the first step a feasible solution is identified (an achievable scheme), and in the second step a tight upper bound that meets this feasible solution is developed (tight converse). The first paper studying secrecy in MIMO communications is [5], which proposes an achievable scheme, where the transmitter uses its multiple transmit antennas to transmit only in the null space of the eavesdroppers channel, thereby preventing any eavesdropping. Reference [6] studies the Gaussian single-input multiple-output (SIMO) wire-tap channel, and shows that it is equivalent to a scalar Gaussian channel, and gives the secrecy capacity using the results of [2]. An achievable scheme has been proposed for the Gaussian multipleinput single-output (MISO) wire-tap channel in [7], and independently and concurrently in [8]. In both of these papers, the achievable secrecy rate is obtained by restricting the 2 channel input to be Gaussian, with no pre-processing of information. The secrecy rate found in [7,8] is shown to be the secrecy capacity of the Gaussian MISO wire-tap channel in [9,10]. Further, [9, 10] allow the eavesdropper to have multiple antennas (MISOME). In all these papers, no one has tried to find the secrecy capacity of MIMO wiretap channel using the scalar gaussian channel.

In this paper we are trying to find out the secrecy capacity of a MIMO channel using a scalar gaussian approach. We consider the MIMO channel a combination of several SIMO channel and find out the secrecy capacity. We use the following notations throughout this paper: Bold face lower and upper case letters are used to represent vectors and matrices, respectively. $\mathbf{x}^\dagger$ denotes the conjugate transpose of the complex vector $\mathbf{x}$. Whether a variable is deterministic or random will be clear from the context.

The remainder of this paper is organized as follows: In section 2, wiretap channel is analyzed where gaussian wiretap channel is also discussed. In section 3, System model for the MIMO channel is introduced where the mathematical model is also developed. Then section 4 concludes this paper.

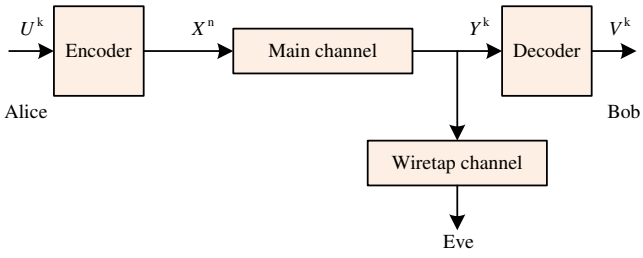


Fig. 1. Wiretap channel

II. WIRETAP CHANNEL

The wiretap channel is a degraded form of a broadcast channel, where the goal is to maximize the transmission rate in the main channel while making negligible the amount of information leaked to the cascade (wiretapper) channel. The secrecy capacity is the maximal rate at which this goal is achieved. We are considering two discrete memoryless channels (DMC) as the ones depicted in Figure 1. The encoder takes the input sequence $U^k = (U_1, ..., U_k)$ and transforms it into a n symbols sequence $X^n = (X_1, ..., X_n)$. The rate of the code is $R = k/n$. The security engineer must design an encoder/decoder pair that ideally maximizes the transmission rate of the legitimate user, subject to the constraint that the rate at which the wiretapper learns the sequence is as small as possible. The wiretapper knows the encoding used, and its ignorance about the source depends only on the noise realization present in the channels. The source is assumed to be stationary and ergodic, and takes its values over a finite alphabet. The probability of block decoding error is denoted by

$$P_e = Pr\{U^k \neq V^k\} \tag{1}$$

The wiretapper uncertainty about the source is measured by the equivocation

$$H(U^k|Z^n), \tag{2}$$

after observing the output of the channel.

Definition 2.1: The fractional equivocation

$$\delta = \frac{H(U^k|Z^n)}{H(U^k)} \tag{3}$$

The rate of the code

$$R = \frac{H(U^k)}{n} \tag{4}$$

Definition 2.2: (Achievability) The pair (R^*, d^*) is said to be achievable if for all $\epsilon > 0$ there exists an enconder/decoder pair such that

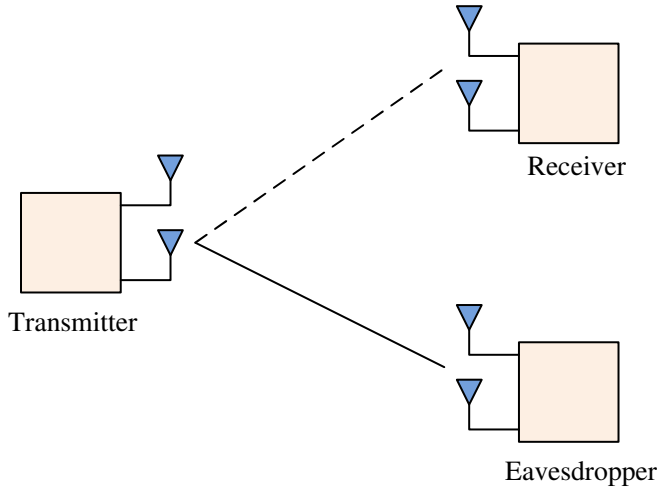


Fig. 2. A communication system with multi-antenna transmitter, receiver and eavesdropper

$$\begin{aligned} R &\geq R^* - \epsilon \\ \delta &\geq d^* - \epsilon \\ P_e &\leq \epsilon \end{aligned} \tag{5}$$

Theorem 2.1: [1]. The set of achievable pairs (R, d) can be characterized as follows

$$\mathcal{R} = \{(R, d) : 0 \leq R \leq C_M, 0 \leq d \leq 1, R_d \leq C_S\} \tag{6}$$

where C_S is the secrecy capacity and its value is

$$C_S = C_M - C_W \tag{7}$$

where C_M is the main channel capacity and C_W is the eavesdropper channel capacity.

Although Wyner's work only considered discrete-time channels, Leung and Hellman [11] proved that the results also hold in a particular case of continuous-time channel. A Gaussian wiretap channel is wiretap channel where the noise is additive white and Gaussian, such that the channel is power limited (P) and the noise processes are independent and have components that are i.i.d. $\mathcal{N}(0, \sigma_1^2)$ and $\mathcal{N}(0, \sigma_2^2)$ respectively. The achievable region of the Gaussian wiretap channel is the same as defined in eq. 6 with

$$\begin{aligned} C_M &= \frac{1}{2}\log(1 + \frac{P}{\sigma_1^2}) \\ C_W &= \frac{1}{2}\log(1 + \frac{P}{\sigma_1^2 + \sigma_2^2}) \end{aligned} \tag{8}$$

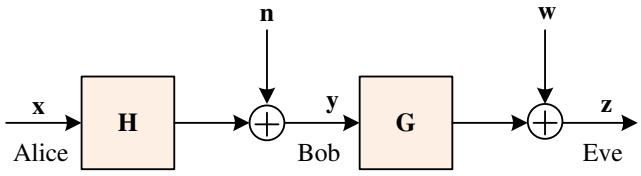


Fig. 3. Degraded broadcast channel

III. System model

Figure 2 shows a communication system, with a transmitter equipped with a transmitter, a receiver and an eavesdropper, each with multiple antennas. The user and eavesdropper channel attenuations can be represented by $N_t \times N_r$ and $N_r \times M_r$ vectors $\mathbf{H}$ and $\mathbf{G}$, where N_t is the number of transmit antennas whereas N_r and M_r are the number of receive antenna at the legitimate receiver and eavesdropper. We are considering the degraded multiple input multiple output (MIMO) channel depicted in Figure 3. We will determine the secrecy capacity of the MIMO Gaussian wiretap channel. For every N_t transmitting antenna, Alice sends symbols that have limited average power $P > 0$, i.e.,

$$\frac{1}{K}\sum_{k=0}^{K-1} E\{x^2[k]\} \leq P \tag{9}$$

Bob uses N_r receive antennas and Eve uses M_r receive antennas to recover Alice's message. Alice sends no message to Eve, so there are no common messages. The channel parameters are represented by $\mathbf{H}$ with dimention $N_r \times N_t$ and $\mathbf{G}$ with dimention $M_r \times N_r$.

The received signals at the receiver and the eavesdropper at k-th time slot are

$$\mathbf{y}[k] = \mathbf{H}\mathbf{x}[k] + \mathbf{n}[k] \tag{10}$$

$$\mathbf{z}[k] = \mathbf{G}\mathbf{y}[k] + \mathbf{w}[k] \tag{11}$$

where $\mathbf{n}$ and $\mathbf{w}$ are independent random vectors, each one of them being complex and jointly Gaussian distributed with mean 0 and non-singular covariance matrices Σ_1 and Σ_2 respectively. $\mathbf{H}$ and $\mathbf{G}$ are known and fixed. When the main and eavesdropper channel experiences fading, $\mathbf{H}$ and $\mathbf{G}$ are assumed to be a vector of i.i.d. zero mean unit-variance complex circularly symmetric Gaussian random variables.

Our goal is to show that the channel described by eq. (10) can be represented by a summation of several scalar channels and that such representation is in fact a parallel Gaussian wiretap channel. Eq. (10) can be the addition of N_t parallel single input multiple output (SIMO) channel and can be written as

$$\begin{aligned}\mathbf{y}[k] &= \sum_{i=1}^{N_t} \mathbf{y}_i[k] \\ &= \sum_{i=1}^{N_t} (\mathbf{h}_i \mathbf{x}_i[k] + \mathbf{n}_i[k])\end{aligned} \tag{12}$$

After putting the value of $\mathbf{y}[k]$ in the eq. (11) we get

$$\mathbf{z}[k] = \mathbf{G}\sum_{i=1}^{N_t} (\mathbf{h}_i \mathbf{y}_i[k] + \mathbf{n}_i[k]) + \mathbf{w}[k] \tag{13}$$

The multiple input multiple output multiple eavesdropper (MIMOME) channel can be seen as a multiple parallel single input multiple output multiple eavesdropper (SIMOME) channel. Every SIMOME channel can be converted to a corresponding scalar channel [6]. The ith main channel and the eavesdropper channel at SIMOME looks like the following

$$\mathbf{y}_i[k] = \mathbf{h}_i x_i[k] + \mathbf{n}_i[k] \tag{14}$$

Replacing $\mathbf{y}[k]$ by $\mathbf{y}_i[k]$ in eq. (11) we get

$$\begin{aligned}\mathbf{z}_i[k] &= \mathbf{G}(\mathbf{h}_i x_i[k] + \mathbf{n}_i[k]) + \mathbf{w}[k] \\ &= \mathbf{G}\mathbf{h}_i x_i[k] + \mathbf{G}\mathbf{n}_i[k] + \mathbf{w}[k]\end{aligned} \tag{15}$$

Multiply the equation (14) by $\mathbf{h}_i^\dagger \Sigma_1^{-1}$ we get

$$\mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{y}_i[k] = \mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{h}_i x_i[k] + \mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{n}_i[k] \tag{16}$$

both the sides are now 1×1 in dimension and in fact scalar. Now taking $\mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{y}_i[k] = y_i[k]$, $\mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{h}_i = h_i^2$ and $\mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{n}_i[k] = n_i[k]$ we get the equation (16) like the followings

$$y_i[k] = h_i^2 x_i[k] + n_i[k] \tag{17}$$

Again multiplying eq. (15) by $(\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger + \Sigma_2)^{-1}$ we get

$$\begin{aligned}(\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger \\ +\Sigma_2)^{-1} \mathbf{z}_i[k] &= (\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger + \Sigma_2)^{-1} \mathbf{G}\mathbf{h}_i x_i[k] \\ &\quad +(\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger + \Sigma_2)^{-1} (\mathbf{G}\mathbf{n}_i[k] \\ &\quad +\mathbf{w}[k])\end{aligned} \tag{18}$$

both the sides are now scalar as it was in the previous case. Now taking $(\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger + \Sigma_2)^{-1} \mathbf{z}_i[k] = z_i[k]$, $(\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger + \Sigma_2)^{-1} \mathbf{G}\mathbf{h}_i = g_i^2$ and $(\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1 \mathbf{G}^\dagger + \Sigma_2)^{-1} (\mathbf{G}\mathbf{n}_i[k] + \mathbf{w}[k]) = w^2[k]$ we get

$$z_i[k] = g_i^2 x_i[k] + w_i[k] \tag{19}$$

Eq. (17) and eq. (19) shows the scalar representation of this MIMOME channel. In the next section, we will derive the secrecy capacity using these results.

IV. Secrecy capacity

The equivocation rate R_e, is the conditional entropy of the transmitted message, conditioned on the received signal at the eavesdropper. The equivocation rate is a measure of the amount of information that the eavesdropper can attain about the message, and quantifies the level of secrecy in the system. The secrecy capacity, C_S, is the largest rate R achievable with perfect secrecy, i.e., $R_e = R$.

The main channel capacity for SIMOME can be written like the following

$$\begin{aligned} C_{M_i} &= \frac{1}{2}\log(1 + h_i^2 P) \\ &= \frac{1}{2}\log(1 + \mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{h}_i P) \end{aligned} \tag{20}$$

Again the eavesdropper channel capacity for SIMOME is

$$\begin{aligned} C_{W_i} &= \frac{1}{2}\log(1 + g_i^2 P) \\ &= \frac{1}{2}\log(1 + (\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1\mathbf{G}^\dagger + \Sigma_2)^{-1} \mathbf{G}\mathbf{h}_i P) \end{aligned} \tag{21}$$

So the main channel and eavesdropper channel capacity for MIMOME channel is given by

$$\begin{aligned} C_M &= \sum_{i=1}^{N_t} C_{M_i} \\ &= \sum_{i=1}^{N_t} \frac{1}{2}\log(1 + \mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{h}_i P) \end{aligned} \tag{22}$$

$$\begin{aligned} C_W &= \sum_{i=1}^{N_t} C_{W_i} \\ &= \sum_{i=1}^{N_t} \frac{1}{2}\log(1 + (\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1\mathbf{G}^\dagger \\ &\quad + \Sigma_2)^{-1} \mathbf{G}\mathbf{h}_i P) \end{aligned} \tag{23}$$

Positive secrecy capacity indicates successful secret communication and the secrecy capacity of the MIMO Gaussian wiretap channel can be written as

$$\begin{aligned} C_S &= C_M - C_W \\ &= \sum_{i=1}^{N_t} \frac{1}{2}\log(1 + \mathbf{h}_i^\dagger \Sigma_1^{-1} \mathbf{h}_i P) \\ &\quad - \sum_{i=1}^{N_t} \frac{1}{2}\log(1 + (\mathbf{G}\mathbf{h}_i)^\dagger (\mathbf{G}\Sigma_1\mathbf{G}^\dagger \\ &\quad + \Sigma_2)^{-1} \mathbf{G}\mathbf{h}_i P) \end{aligned} \tag{24}$$

V. Conclusion

A Gaussian multiple input multiple output (MIMO) channel is considered where a transmitter is communicating to a receiver in the presence of an eavesdropper. The transmitter is equipped with multiple antennas, while the receiver and the eavesdropper also contains multiple antennas. We present a technique for determining the secrecy capacity of the MIMO channel under Gaussian noise. To do so, we transform the channel into multiple single input multiple output (SIMO) Gaussian wiretap channel and then use scalar approach using standard techniques of communications theory. Introduction of fading in this MIMO channel is still an open problem which we are trying to solve . One particular direction corresponds to the case where the broadcast channel is no longer degraded, such as the one considered by I. Csiszar and J. Korner [3].

Acknowledgment

This work was supported by the u-LCRC at Kyung Hee University, Korea, under the ITRC program of MIC, Korea (IITA-2008-(C1090-0801-0002)).

References

[1] A. D. Wyner,"The wire-tap channel," *Bell Syst. Tech. J.*, 54(8):210, October 1975.

[2] S. K. Leung-Yan-Cheong and M. E. Hellman,"The Gaussian wire-tap channel," *IEEE Trans. on Information Theory,* 24(4):451456, July 1978.

[3] I. Csiszar and J. Korner,"Broadcast channels with confidential messages," *IEEE Trans. on Information Theory,* 24(3):339348, May 1978.

[4] I. E. Telatar,"Capacity of multi-antenna Gaussian channels," *European Trans. Telecommunications,* 10:585595, November 1999.

[5] R. Negi and S. Goel,"Secret communication using artificial noise,"*IEEE Vehicular Technology Conference,* Toulouse, France, May 2006.

[6] P. Parada and R. Blahut,"Secrecy capacity of SIMO and slow fading channels," *IEEE International Symposium on Information Theory,* Adelaide, Australia, September 2005.

[7] S. Shafiee and S. Ulukus,"Achievable rates in Gaussian MISO channels with secrecy constraints," *In IEEE International Symposium on Information Theory,* Nice, France, June 2007.

[8] Z. Li, W. Trappe, and R. D. Yates,"Secret communication via multi-antenna transmission," *41st Conference on Information Sciences and Systems,* Baltimore, MD, March 2007.

[9] A. Khisti, G. Wornell, A. Wiesel, and Y. Eldar,"On the Gaussian MIMO wiretap channel," *IEEE International Symposium on Information Theory,* Nice, France, June 2007.

[10] A. Khisti and G. Wornell,"Secure transmission with multiple antennas: The MISOME wiretap channel," *Submitted to IEEE Trans. on Information Theory*

[11] S. Leung-Yan-Cheong and M. E. Hellman,"The gaussian wire-tap channel," *IEEE Transactions on Information Theory,* vol. 24, no. 4, pp. 451456, 1978.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Impact of In-Band and Inter-Band Crosstalk due to Multiwavelength Optical Cross-Connect in a WDM Network

M. Jalal Uddin, and S.P. Majumder, Member, IEEE

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET)
Dhaka – 1000, Bangladesh
E-mail: jalaluddin@eee.buet.ac.bd, spmajumder@eee.buet.ac.bd

Abstract – An analytical approach is developed to evaluate the impact of in-band and inter-band crosstalk on the bit error rate (BER) performance due to a multiwavelength optical cross-connect (OXC) in a wavelength division multiplexing (WDM) network. BER performance and power penalty results are numerically evaluated at a bit rate of 10 Gb/s and the factors that degrade the network performance are identified. The results show that for a 16×16 OXC, to achieve a BER of 10^{-9} at a power penalty of approximately 1 dB with a relative component crosstalk of – 50 dB, the presence of in-band crosstalk allows maximum 4 wavelength channels in a fiber and the presence of inter-band crosstalk allows maximum 24 wavelength channels in a fiber for a wavelength separation of 0.01 nm.

I. Introduction

Wavelength division multiplexed (WDM) optical networks are attracting more and more attention because of their ability to provide increased capacity and flexibility [1]. Optical cross-connect (OXC) is an essential network element in a WDM optical network [2]. A number of OXC architectures have been proposed in [2] and [3], each of which have its own unique features, strengths, and limitations. Imperfections of the optical components used in these architectures give rise to optical crosstalk [4], [5]. Crosstalk is classified as in-band (homodyne or intra-band) and inter-band (heterodyne) crosstalk. In-band crosstalk has the same wavelength as the signal and degrades the transmission performance seriously [6]–[10]. Because of the identical wavelength, this crosstalk is difficult to be eliminated by filtering and the crosstalk beats with the signal and generates beat noise components at the receiver output [6]. In-band crosstalk can be divided into coherent crosstalk, whose phase is correlated with the signal considered and incoherent crosstalk, whose phase is not correlated with the signal. Coherent and incoherent crosstalk is studied separately in [10]. The effect of inter-band crosstalk has been neglected in [6]–[10] and in our work, it is found that, the crosstalk contribution due to inter-band crosstalk is significant for ultra dense WDM system, causes much higher noise, and degrades the network performance severely.

In this paper, the impact of in-band and inter-band crosstalk is analyzed due to a multiwavelength OXC in an intensity-modulated direct-detection receiver (IM-DD) WDM system. In section II, the system model and a mux/demux based multiwavelength OXC topology are presented. The presence of in-band and inter-band optical crosstalk is also described in this section. Afterwards, analytical expression of in-band crosstalk, inter-band crosstalk and bit error rate are derived in section III. Finally, the system performance is evaluated numerically in terms of bit error rate and power penalty in section IV.

II. System Model and OXC topology

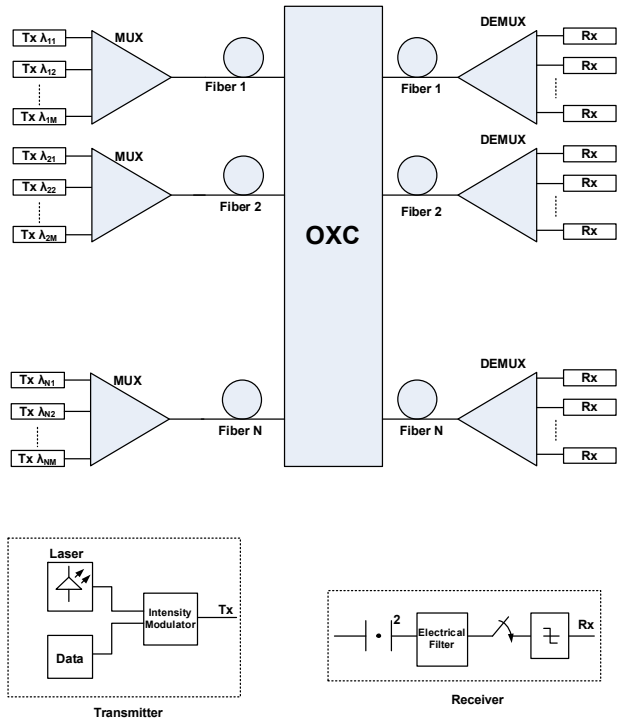


Fig. 1: Model of a WDM system with N×N unidirectional OXC.

Fig. 1 shows the system model of a WDM network using an N×N unidirectional OXC. Each input multiplexer of Fig. 1 combines M wavelengths coming from the M

transmitters, and is connected through the optical fiber to the input port of the OXC, the detail view of which is given in Fig. 2. The OXC uses N number of (M×1) multiplexers and (1×M) demultiplexers and M number of (N×N) optical switches. The demultiplexer of OXC separates M wavelengths. The optical switch takes N number of same wavelength signals coming from all the N input fibers and routes each wavelength to any of the N output fibers according to the destination address. The multiplexer of OXC again combines M number of wavelengths and sends them to a single fiber. The output demultiplexer of Fig. 1 separates the M wavelengths and sends them to the individual user terminal.

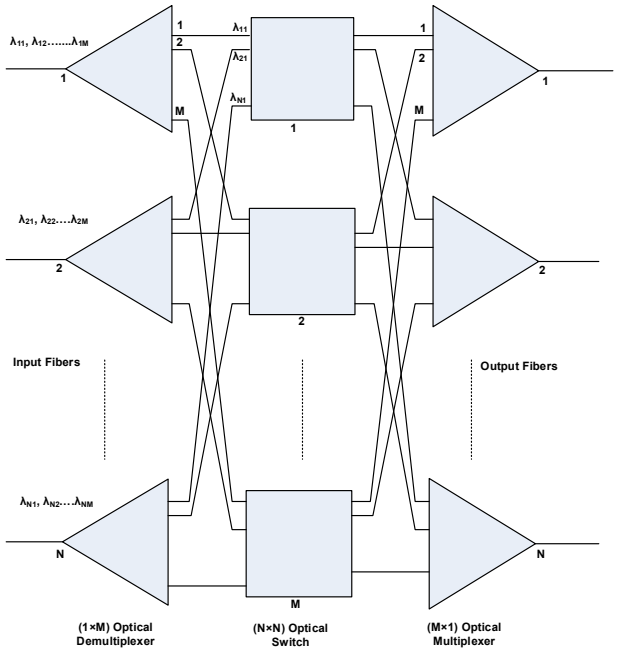


Fig. 2: Conventional OXC using multiplexer, demultiplexer and optical switch. This OXC has N input and output fibers and each fiber carries M number of wavelengths.

Crosstalk is the general term given to the effect of other channels on the desired signal. In OXC, crosstalk arises due to the imperfect operation of the multiplexers, demultiplexers and switches. The dumultiplexer ideally separates the incoming wavelengths to different output paths. In reality, however, a portion of the signal at wavelength, say λ_i leaks into the adjacent channel λ_{i+1} because of the non ideal separation within the demultiplexer. When the wavelengths are again combined into a single fiber by the multiplexer, a small portion of the λ_i that leaked into the λ_{i+1} channel, will also back into the common fiber at the output. Although both signals contain the same data, they are not in phase with each other, due to different delays encountered by them. This causes in-band crosstalk. The crosstalk penalty is highest when the crosstalk signal is exactly out of phase with the desired signal. Crosstalk arises in switches due to the non ideal isolation of one switch port from the other. In this case, the signal contains different data causing the inter-band crosstalk when combined in the multiplexer into the same output fiber. Inter-band crosstalk also arises due to the demultiplexer that selects one channel and imperfectly rejects the others.

The OXC shown in Fig. 2 consists of a total of N optical demultiplexers, M optical switches, and N multiplexers. Each of the input fibers to optical demultiplexers contains M different wavelengths. The optical demultiplexers spatially separate the incoming wavelengths into M paths. Each of these paths passes through an optical switch before they are combined with the outputs from the other M–1 optical switches.

Assuming the OXC is fully loaded, each signal passing through the OXC will be interfered by M+N–2 in-band contributions, N–1 of which are leaked by the optical switch, and the other M–1 are leaked by the demultiplexer/multiplexer pair. For facilitating the description, we now consider the signal with wavelength 1 in input fiber 1, noted as λ_{11} as the main signal. The main signal λ_{11} will be interfered by N–1 crosstalk contributions leaked from the N–1 signals with wavelength 1 (λ_{11}, λ_{21}, λ_{31}.....λ_{N1}) in the other N–1 input fibers, when passing through the optical switch 1, because of the non ideal crosstalk specification of optical switches. The N–1 crosstalk contributions can be treated as generated by different lasers and they are phase incorrelated with λ_{11}, and with each other. Similarly, when each signal with wavelength 1 is demultiplexed into one path, there will be a fraction of it in each other M–1 outputs of the corresponding demultiplexer because of the non ideal crosstalk specification of optical demultiplexers. After passing through the optical switches, the main signal is multiplexed with M–1 signals with different wavelengths. At the same time, the M–1 crosstalk contributions of wavelength 1 in the M–1 paths are combined with the main signal though isolated again by the optical multiplexer. These M–1 crosstalk contributions can be leaked from signal with wavelength 1 in all the N input fibers. Some of them can be leaked from λ_{11}, i.e., the main signal itself. Number of contributions leaked from λ_{11} is the coherent part of the crosstalk. But here we assume that all the M–1 crosstalk components are incoherent. At the output multiplexer, M wavelengths are multiplexed together into the same fiber. Here adjacent M–1 wavelengths cause inter-band crosstalk. This inter-band crosstalk can be eliminated by appropriate design of the receiver filter. But some portion of this M–1 wavelengths fall into the receiver bandwidth due to the frequency characteristics of digital bit streams that cannot be eliminated.

III. Analytical Expression

Let us define the signal and crosstalk light as,

$$S_i(t) = \sqrt{2}E_i \sin(2\pi f_o t + \phi_i(t)) \qquad (1)$$

Where E_i is the rms value of the electric field for the i th signal. The subscript i = 1 denotes the main signal and i = 2 to M+N–1 denote the M+N–2 in-band crosstalk components. f_o is the laser oscillation frequency and $\Phi_i(t)$ expresses the phase noise of the laser. This includes the frequency drift of the laser. After the signal and M+N–2

in-band crosstalk light are received by a photo detector at the destination node, the output photocurrent is given by,

$$i_p = {E_1}^2 + \sum_{i=2}^{M+N-1} 2E_1E_i \cos(\phi_1(t) - \phi_i(t)) + \sum_{i,j=2(i>j)}^{M+N-1} 2E_iE_j \cos(\phi_i(t) - \phi_j(t)) + \sum_{i=2}^{M+N-1} {E_i}^2 \quad (2)$$

The first and last terms are the optical power of the signal and crosstalk respectively. The second term is called signal-crosstalk beat noise. The third term is the crosstalk-crosstalk beat noise. The crosstalk-crosstalk beat noise is negligible compared to the signal-crosstalk beat noise [6].

The cosine function in the second term on the right hand side of equation (2) denotes the beat of the output current due to signal-crosstalk interference. If the whole noise power is inside the receiver bandwidth, the electrical noise power for each beat component is given by $2E_1^2E_i^2$. The normalized noise power is obtained by adding the individual beat noise power and dividing the sum by the signal power. The result is,

$$\sigma_{RIN}^2 = \frac{1}{(E_1^2)^2} \sum_{i=2}^{M+N-1} 2E_1^2E_i^2 = 2R_c(M+N-2) \quad (3)$$

Where R_c is the composite component crosstalk due to multiplexer, demultiplexer and switch and is defined as the optical power ratio of each crosstalk component to the signal.

$$R_c = \frac{E_i^2}{E_1^2} \quad (4)$$

σ_{RIN}^2 is referred to as the relative intensity noise (RIN). Equation (3) is obtained for the continuous wave light. Actually, the crosstalk light as well as the signal is modulated. The signal-crosstalk beat occurs when the crosstalk channel is in the binary '1' state. If the probability of occurring binary '1' is 0.5, the noise power is reduced by half. So the RIN of the signal crosstalk beat noise is given by,

$$\sigma_{RIN}^2 = R_c(M+N-2) \quad (5)$$

The received baseband digital signal in time domain is given by well known rectangular function,

$$b(t) = \sqrt{2\Pr} rect\left(\frac{t}{T_b}\right) \quad (6)$$

Where Pr is the received power and T_b is the bit period.

In frequency domain, this digital bit becomes the well-known *sinc* function, whose spectrum exists for the entire frequency range and is given by,

$$B(f) = \sqrt{2\Pr} T_b \frac{\sin(\pi f T_b)}{(\pi f T_b)} \quad (7)$$

The inter-band crosstalk due to this frequency characteristic is found by summing the optical power due to the M–1 adjacent wavelengths that falls within the bandwidth of the receiver and is given by,

$$\sigma_{\text{int}er}^2 = \sum_{i=1(f_i \neq f_s)}^{M} \int_{f_s - B/2}^{f_s + B/2} \left(\sqrt{2\Pr} T_b\right)^2 \left| \frac{\sin(\pi(f - f_i)T_b)}{(\pi(f - f_i)T_b)} \right|^2 df \quad (8)$$

where, f_s is the signal frequency, f_i is the frequency of the crosstalk signal, B is the receiver bandwidth and f is the variable frequency.

The general formula for the bit error rate (BER) for intensity-modulated direct-detection (IM-DD) receiver is given by [11],

$$BER = \frac{1}{4} erfc\left(\frac{i_p}{2\sqrt{2}\sigma_1}\right) + \frac{1}{4} erfc\left(\frac{i_p}{2\sqrt{2}\sigma_0}\right) \quad (9)$$

where, i_p is the photocurrent when the signal is in binary '1' state. This corresponds to optical power. σ_1^2 and σ_0^2 denotes the variance of noise on symbol 1 and 0 and it is assumed that noise has a Gaussian distribution. In the absence of crosstalk it is assumed that both σ_1^2 and σ_0^2 are equal to the thermal noise of the receiver, which is given by,

$$\sigma_{th}^2 = \frac{4kTB}{R_L} \quad (10)$$

where, k = Boltzman constant, T = room temperature in degree Kelvin, B = receiver bandwidth in hertzs and R_L = receiver load resistance in ohms.

As crosstalk occurs due to the binary '1' state of the signal, $\sigma_0 = \sigma_{th}$ and σ_1 varies depending on the presence of different crosstalk noise sources. It is assumed that, the beat noise in presence of in-band crosstalk has an approximately Gaussian probability distribution and the total noise is given by,

$$\sigma_1^2 = \sigma_{th}^2 + \sigma_{in}^2 \quad (11)$$

where, σ_{in}^2 is the variance of noise power in presence of in-band crosstalk and is given by,

$$\sigma_{in}^2 = \sigma_{RIN}^2 \times i_p^2 = R_c(M+n-2) \times i_p^2 \quad (12)$$

Replacing σ_1 from equation (11) into equation (9), we get the expression of bit error rate in presence of in-band crosstalk.

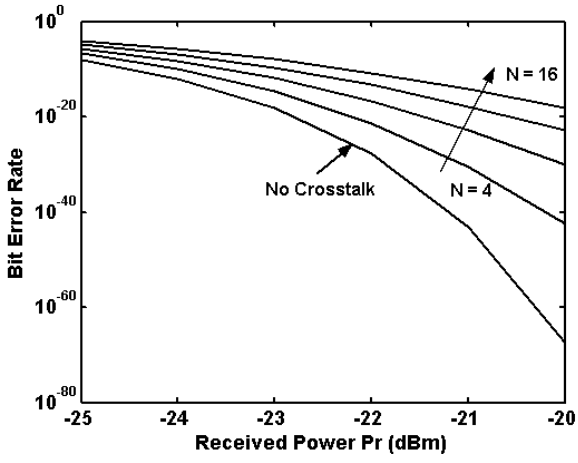
The total noise in presence of inter-band crosstalk is given by,

$$\sigma_1^2 = \sigma_{th}^2 + \sigma_{\text{int}er}^2 \quad (13)$$

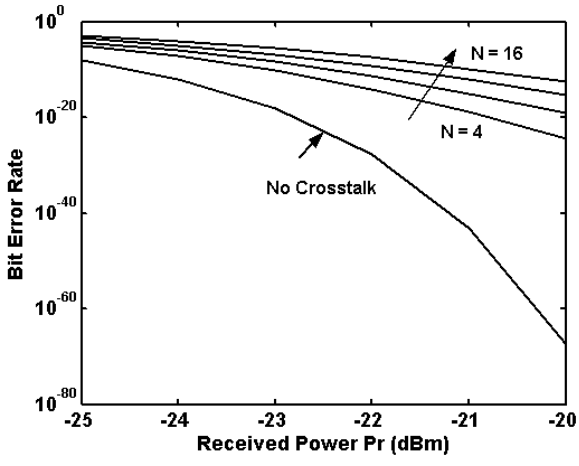
where, $\sigma_{\text{int}er}^2$ is given by equation (8). Replacing σ_1 from equation (13) into equation (9), we get the expression of bit error rate in presence of inter-band crosstalk.

IV. Results and Discussion

Following the analytical expression presented in section III, BER performance is numerically evaluated considering the effect of in-band and inter-band crosstalk for different number of fibers N, number of wavelength channels in a fiber M, component crosstalk R_c, and channel separation $\Delta\lambda$. Power penalty is calculated for each case to achieve a specific bit error rate.



(a)



(b)

Fig. 3: BER performance in presence of in-band crosstalk with varying number of fibers. (a) M = 1, (b) M = 8. Rc = – 40 dB in both cases.

Fig. 3 (a) and (b) shows the BER performance as a function of the received power in presence of in-band crosstalk, for different number of fibers with one wavelength channel in a fiber and eight wavelength channels in a fiber respectively. From the plots, it is obvious that BER decreases with the increase of received power that agrees with the result found in [6]. It is also found that BER increases with the increase in number of fibers and with the increase in number of wavelength channels in a fiber for the increased amount of crosstalk [5]. For same Pr, R_c, and M, when N=4, BER = 10^{-13}, and when N=12, BER =10^{-9}. For same Pr, R_c, and N, when M = 8, BER=10^{-13}, and when M = 12, BER = 10^{-11}.

Fig. 4 depicts the plot of power penalty in presence of in-band crosstalk as a function of the number of fibers for different number of wavelength channels in a fiber. The power penalty is calculated to achieve a BER of 10^{-9}.

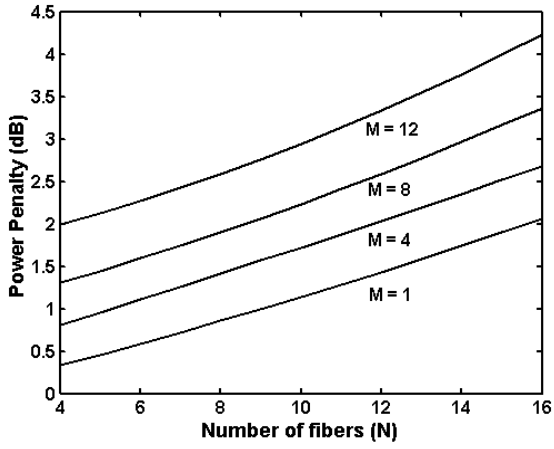
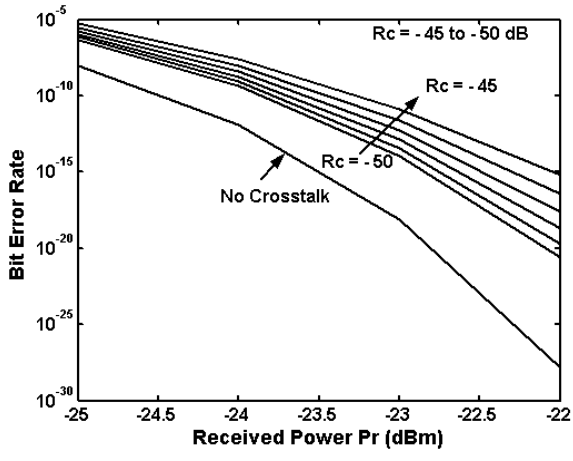
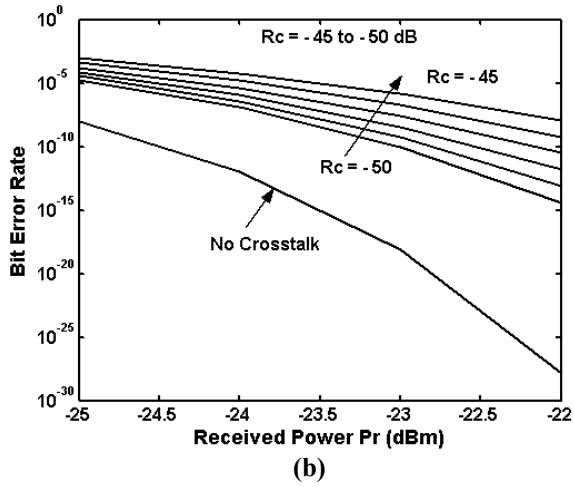


Fig. 4: Power Penalty in presence of in-band crosstalk as a function of number of fibers for different number of wavelength channels in a fiber. Power penalty is plotted to get an overall BER of 10^{-9}.



(a)



(b)

Fig. 5: BER performance in presence of in-band crosstalk for varying component crosstalk. (a) M = 8, (b) M = 20. N = 4 in both cases.

It is found that, the power penalty increases with the increase in number of fibers and also with the increase in number of wavelength channels in a fiber that agrees the

result found in [4]. For same N and R_c, when M = 8, power penalty is found to be 1.3 dB and when M = 12, power penalty is found to be 2 dB. For same M and R_c, when N = 4, power penalty is found to be 2 dB and when N = 20, power penalty is found to be 4.2 dB.

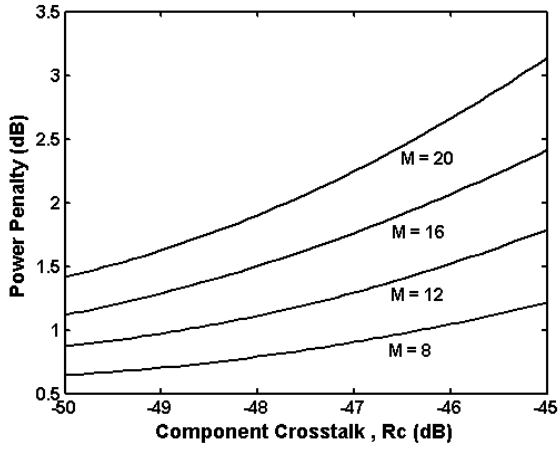


Fig. 6: Power penalty as a function of component crosstalk in presence of in-band crosstalk. Dependence of this power penalty on different number of wavelength channels in a fiber is also presented. Power penalty is plotted to get an overall BER of 10^{-9}.

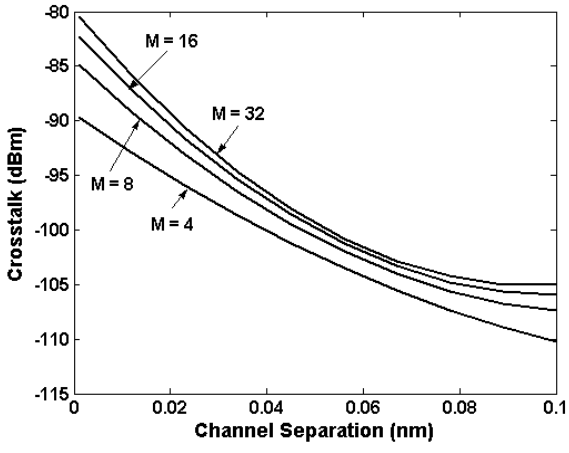
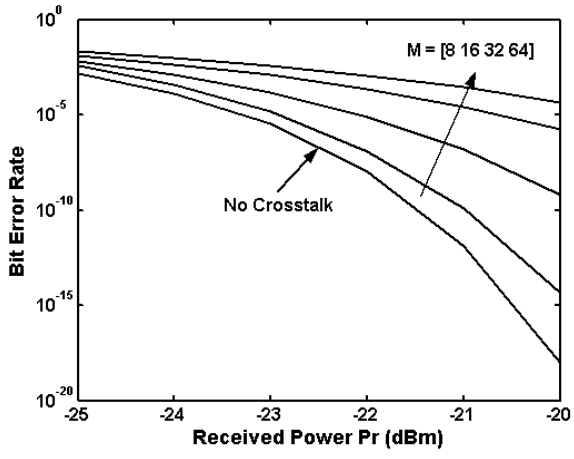


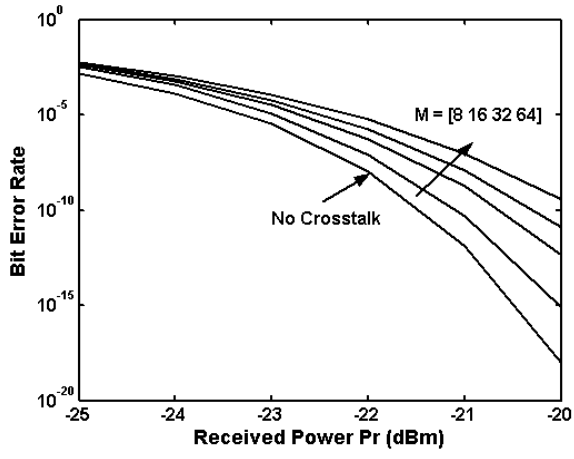
Fig. 7: Inter-band crosstalk as a function of channel separation for different number of wavelength channels in a fiber.

Fig. 5(a) and (b) shows the BER performance in presence of in-band crosstalk as a function of the received power for different component crosstalk with eight wavelength channels in a fiber and twenty wavelength channels in a fiber respectively. It is obvious from the plot that BER increases with the increase in component crosstalk that agrees with result found in [6] and with the increase in number of wavelength channels in a fiber. For same Pr, M and N, when Rc = – 40 dB, BER = 10^{-13}, and when R_c = – 45 dB, BER = 10^{-15}.

Fig. 6 shows the power penalty in presence of in-band crosstalk as a function of the component crosstalk for different number of wavelength channels in a fiber. It is found that power penalty increases with the increase in component crosstalk which aggress with result in [4] and [6] and also with the increase in number of wavelength channels in a fiber [4]. For same N and M, when R_c = – 40 dB, power penalty is found to be 1.3 dB and when R_c = – 45 dB, power penalty is found to be 1.21 dB.



(a)



(b)

Fig. 8: BER performance in presence of inter-band crosstalk with varying number of wavelength channels in a fiber and different channel separation. (a) Δλ = 0.004 nm, (b) Δλ = 0.01 nm.

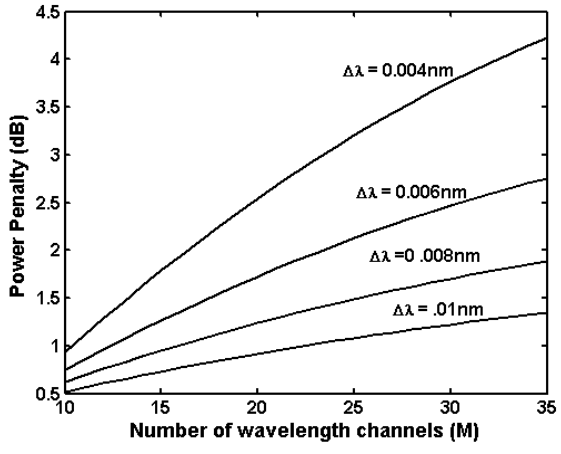


Fig. 9: Power penalty due to inter-band crosstalk as a function of the number of wavelength channels in a fiber. Dependence of this power penalty on channel separation is also presented. Power penalty is plotted for an overall BER of 10^{-9}

Fig. 7 depicts the inter-band crosstalk as a function of the channel separation for different number of wavelength channels in a fiber. It is seen that inter-band crosstalk decreases with the increase in channel separation and increases with the increase in number of wavelength channels in a fiber.

Fig. 8 (a) and (b) gives the plot of BER performance as a function of the received power in presence of inter-band crosstalk for different number of wavelength channels in a fiber with a channel separation of 0.004 nm and 0.01 nm respectively. It is found from the plot that BER increases with the decrease in channel separation and with the increase in number of wavelength channels in a fiber. For same Pr, R_c and M, when $\Delta\lambda$ = 0.004 nm, BER = 10^{-14} and when $\Delta\lambda$ = 0.01 nm, BER = 10^{-15}. For same Pr, R_c and $\Delta\lambda$, when M=8, BER=10^{-15} and when M=16, BER = 10^{-12}.

Fig. 9 shows the power penalty in presence of inter-band crosstalk as a function of the number of wavelength channels in a fiber for different channel separation. It is seen that power penalty increases with the increase in number of wavelength channels in a fiber and with the decrease in channel separation. For same Pr, R_c and M, when $\Delta\lambda$ = 0.004 nm, power penalty is found to be 1.9 dB and when $\Delta\lambda$ = 0.01 nm, power penalty is found to be 0.57 dB. For same Pr, R_c and $\Delta\lambda$, when M = 16, power penalty is found to be 1.9 dB and when M = 32, power penalty is found to be 4.2 dB.

V. Conclusion

Theoretical analysis of in-band and inter-band crosstalk is carried out and analytical expression of bit error rate is developed considering the impact of in-band and inter-band crosstalk. Network performance of an M wavelength N×N WDM network is evaluated numerically in terms of bit error rate and power penalty. It is found that bit error rate and power penalty in presence of in-band crosstalk increases with an increase in number of fibers, number of wavelength channels in a fiber and component crosstalk. It is also found that, bit error rate and power penalty in presence of inter-band crosstalk increases with the increase in number of wavelength channels in a fiber and with the decrease in channel separation. It is obvious from the analysis that for a specific network performance, in-band crosstalk limits the number of fibers and number of wavelength channels in a fiber and inter-band crosstalk limits the number of wavelength channels in a fiber.

References

[1] G. R. Hill et al., "A transport network layer based on optical network element," *J. Lightwave Technol.,* vol. 11, no. 5/6, pp. 667- 679, May/June 1993.

[2] E. Iannone, and R. Sabella, "Optical path technologies: A comparison among different cross-connect architectures," *J. Lightwave Technol.,* vol. 14, no. 10, pp. 2184 - 2194, Oct. 1996.

[3] S. Okamoto, A. Watanabe, and K. Sato, "Optical path cross-connect node architectures for photonic transport network,'' *J. Lightwave Technol.,* vol. 14, no. 6, pp. 1410 - 1422, June 1996.

[4] J. Zhou, et al., "Crosstalk in multi wavelength optical cross-connect networks," *J. Lightwave Technol.,* vol. 14, no. 6, pp. 1423 - 1435, June 1996.

[5] T. Gyselings, G. Morthier, and R. Baets, "Crosstalk analysis of multiwavelength optical cross connects," *J. Lightwave Technol.,* vol. 17, no. 8, pp. 1273 - 1283, Aug 1999.

[6] H. Takahashi, K. Oda, and H. Toba, "Impact of crosstalk in an arrayed waveguide multiplexer on N×N optical interconnection," *J. Lightwave Technol.,* vol. 14, no. 6, pp. 1120 - 1126, June 1996.

[7] S. D. Dods, J. P. R. Lacey, and R. S.Tucker, "Homodyne crosstalk in WDM ring and bus networks," *IEEE Photon.Technol. Lett.,* vol. 10, no. 3, pp. 457 - 458, Mar. 1998.

[8] K.P. Ho, C. K. Chan, F. Tong, and L. K. Chen, "Exact analysis of homodyne crosstalk induced penalty in WDM networks," *IEEE photon. Technol. Lett.,* vol. 9, no. 3, pp. 1285 - 1287, Sept. 1997.

[9] S. D. Dods, J. P. R. Lacey, and R.S. Tucker, "Performance of WDM ring and bus network in the presence of homodyne crosstalk," *J. Lightwave Technol.,* vol. 17, no. 3, pp. 388 - 396, Mar. 1999.

[10] Y. Shen, K. Lu, and W. Gu, "Coherent and incoherent crosstalk in WDM optical networks," *J. Lightwave Technol.,* vol. 17, no. 5, pp. 759 - 764, May 1999.

[11] G. Keiser, "Optical Fiber Communication," 3rd Edition, McGraw-Hill Companies, Inc., 2000.

[12] R. Ramaswami, K.N Sivaranjan, "Optical Networks," 2nd Edition, Morgan Kaufmann Publishers, 2002.

[13] T. M. Idelfonso and T. Eduward, "Crosstalk in WDM Communication Networks," The Springer International Series in Engineering and Computer Science, vol. 678, 2002.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effect of Cross Phase Modulation (XPM) on the Bit Error Rate Performance of an Optical CDMA (OCDMA) System

Shahriar Ferdous, Mohammed Shahriar Zaman, Khandaker Fahim Imran, S.P. Majumder
Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology (BUET),
Dhaka-1000, Bangladesh.
E-mail: shahriar935@yahoo.com, z_shahriar@yahoo.com, imran719@gmail.com,
spmajumder@eee.buet.ac.bd.

Abstract— An analytical investigation is carried out to evaluate the impact of cross phase modulation due to non-linear Kerr effect in a single mode fiber on the transmission performance of an optical CDMA system. Expression for the Intensity Modulation (IM) induced by PM-IM (Phase Modulation-Intensity Modulation) conversion due to GVD (Group Velocity Dispersion) is derived and the mean and variance of MAI (Multiple Access Interference) at the output of SIK (Sequence Inversed Keyed) correlator receiver is determined. Performance results are evaluated in terms of BER (Bit Error Rate) performances of OCDMA system with variable probe power, fiber length, number of users, bit rate and code length considering 31 chip and 127 chip m-sequence codes. Results show that the system performance deteriorates significantly due to the IM caused by GVD and PM.

I. Introduction

In Optical Code Division Multiple Access (OCDMA) systems, power fluctuations of one optical wave propagating in an optical fiber can modulate the phase of other co-propagating waves through cross-phase modulation (XPM) [1]. In intensity modulation-direct detection (IM-DD) OCDMA systems, XPM can limit the distance and the capacity because the group velocity dispersion (GVD) converts the XPM induced phase modulation (PM) to IM [2].

In this paper, the dependence of intensity modulation induced by XPM and its effect on BER for different values of probe power, fiber length, number of users and bit rate is investigated theoretically. A simple expression for the XPM-induced IM is derived and results are evaluated by numerical computations.

The paper is structured as follows. Section I is the Introduction of the paper. Section II describes the system physical model. In Section III a theoretical analysis of XPM in a single-segment fiber link is presented. The results of the simulation are discussed in Section IV. In Section V the main conclusion is outlined.

II. System Description

In the present research work, asynchronous sequence inversed keyed (SIK) modulation demodulation technique is considered for optical fiber CDMA network. As shwon in Fig. 1, two channels, one is the channel under investigation known as probe channel and the other is the interfering channel known as the pump channel, are co-propagating inside the single segment optical fiber [3]. In the transmitter, data of an user is modulated either by a unipolar signature or by its complement, depending on whether it is a "1" or "0" respectively. In this scheme, both the probe signal and the pump signal is modulated by the chip sequence. An optically switched correlator receiver based on the principle of unipolar bipolar correlation has been used. Unipolar bipolar correlation allows conventional bipolar signature sequences to be used in SIK DS CDMA optical fiber network using noncoherent transmission and direct detection.

In the receiver a bipolar reference sequence is correlated directly with the channel unipolar signature sequence in order to recover the original data. This unipolar bipolar correlation is practically realized in an all optical correlator , by separating the bipolar reference sequence into two complementary unipolar reference sequences which provide the unipolar switching functions to dispread the optical channel signal.

III. Theoretical Analysis

A. General Theory of XPM

XPM occurs in systems having at least two channels. The group velocity dispersion (GVD) converts the XPM induced phase modulation (PM) to IM. At the transmitter the electrical signals modulate the optical output. Consider, channel 1 is probe signal and channel 2 is the pump signal, both of them are modulated. The intensity modulated input pump signal has an angular frequency ω.

The probe signal at the fiber output after propagating through an optical fiber at fiber length L is found by the equation (1) [4]

$$A_1(L,t) = \sqrt{|A_1(0,t)|^2.P_{xpm}(L,t)}.e^{DL}.e^{(j\phi_1(L,t))} \quad (1)$$

where $|A_1(L,t)|^2$ is the intensity of probe channel signal at fiber output after propagating over fiber length L and $P_{xpm}(L,t)$ is the XPM induced Intensity Modulation (IM) at fiber length L. The term $\phi_1(L,t)$ represents the non-liner phase-shift of probe signal and it is given as (2) [4]

$$\phi_1(L,t) = \phi_{spm}(L,t) + \phi_{xpm}(L,t) \quad (2)$$

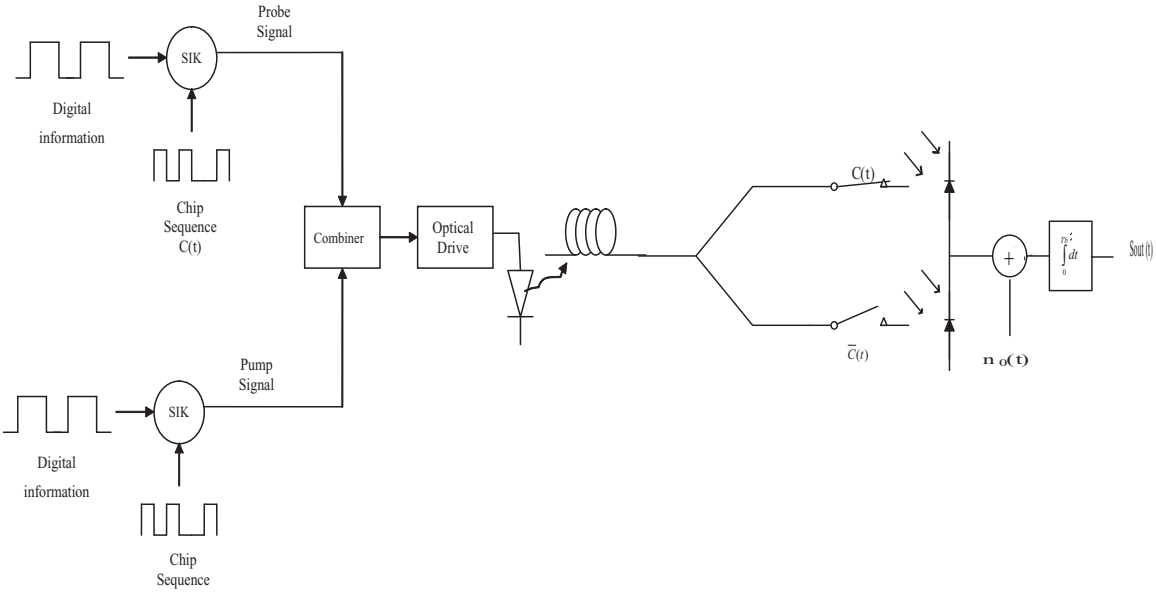


Fig. 1. Schematic block diagram of an asynchronous optical CDMA transmission system with Sequence Inversed Keyed (SIK)

where $\phi_{spm}(L,t)$ and $\phi_{xpm}(L,t)$ represents the phase-shift induced by SPM (Self Phase Modulation) and XPM (Cross Phase Modulation) respectively.

The SPM induced phase shift is defined as (3)

$$\phi_{spm}(L,t) = \gamma_1 P_{10} L_{eff} \tag{3}$$

where $P_{10} = |A_1(0,t)|^2$ is the optical power of the probe signal at fiber input (when L=0) and L_{eff} is the effective fiber length given by (4)

$$L_{eff} = (1 - e^{(1-\alpha L)})/\alpha \tag{4}$$

The nonlinear coupling coefficient is defined as (5)

$$\gamma_1 = \frac{2n_2\pi}{\lambda_1 A_{eff}} \tag{5}$$

where n_2 is the nonlinear refractive index of the fiber. A_{eff} is the effective core area of the fiber expressed as (6):

$$A_{eff} = \pi r^2 \tag{6}$$

where r is the radius of the fiber.

Considering a small section of fiber length where fiber nonlinearity and dispersion are assumed to act independently, the XPM induced phase shift in probe channel when optical pulse in pump channel has traveled a distance z from fiber input can be expressed by equation (7) [5].

$$\phi_{xpm,1}(z,\omega) = -2\gamma_1 P_2(z,\omega) \tag{7}$$

where

$$P_2(z,\omega) = P_2(0,\omega) cos(qz) e^{(-az)} e^{(-j\omega z/V_{g2})} \tag{8}$$

where $a = \alpha - j\omega d_{12}$, q=$\frac{\omega^2 D\lambda^2}{4\pi c}$, α is the attenuation coefficient, D is the dispersion coefficient and c is the light speed [2]. In a nonzero dispersion region $d_{12} = D.\Delta\lambda_{12}$ where $\Delta_{12} = \lambda_1 - \lambda_2$ is the wavelength separation between channel 1 and 2. In case of OCDMA $\lambda_1 = \lambda_2 = \lambda$. So $d_{12} = \Delta_{12} = 0$.

Equation (8) represents the pump power fluctuation after propagating at distance z from the fiber input that modulates the phase of the probe signal through XPM. The total XPM induced phase shift of probe signal after propagating at distance L is given by equation (9) [5].

$$\begin{aligned} \phi_{xpm,1}(L,\omega) &= \int_0^L \phi_{xpm}(z,\omega)\partial z \\ &= |H(\omega)||P_2(z,\omega)|. \\ &\quad cos(\frac{\omega L}{v_{g1}} + \angle P_2(z,\omega) + \angle H(\omega)) \end{aligned} \tag{9}$$

The total XPM induced phase shift in time domain can be obtained by taking the real part of inverse Fourier Transform of equation (9)

$$\phi_{xpm}(L,t) = Re(\frac{1}{\pi}\int_0^\infty \phi_{xpm}(L,\omega) e^{(-j\omega t)} \partial t) \tag{10}$$

Equation (10) represents the XPM induced phase shift in probe signal at fiber output after propagating at distance L.

This XPM induced phase shift will be converted to XPM induced intensity modulation (IM) by GVD effect. In order to determine the strength of XPM induced IM in probe channel, the expression of power fluctuations on probe signal at fiber output induced by XPM at infinitesimal distance z from fiber input is given by equation(11)[2]:

$$\begin{aligned} P_{xpm}(\omega) &= -2P_1(z) e^{(-\alpha(L-z))}. \\ &\quad e^{(-j\omega(L-z)/V_{g1})} sin(b(L-Z)). \\ &\quad \phi_{xpm,1}(z,\omega) \end{aligned} \tag{11}$$

Here $P_1(z) = P_1(0)exp(-\alpha z)$ is the average power of probe channel at distance z, $P_1(0)$ is the average power of probe channel at fiber input and b is defined as b =$\frac{\omega^2 D\lambda_1^2}{4\pi c}$. The total XPM induced IM at probe channel at fiber output is the sum of the XPM induced IM at infinitesimal section z along the fiber with length L is given by equation(13)[2]:

The total XPM induced IM in the time domain at probe channel at fiber output is obtained by inverse Fourier transform of equation (12)

$$P_{xpm}(L,t) = \left| \frac{1}{\pi} \int_0^{\infty} P_{xpm}(\omega) e^{(-j\omega t)} \partial t \right| \tag{12}$$

$$P_{xpm}(\omega) = 2\gamma_1 P_1(0) P_2(\omega) e^{(-aL)} e^{(-j\omega L/V_{g1})}. \tag{13}$$

$$\begin{aligned} &\{\frac{1}{a^2+(b+q)^2}[a.sin(bL) - (b+q)cos(bL) \\ &+[a.sin(qL) + (b+q)cos(qL)]exp(-\alpha L)] \\ &+\frac{1}{a^2+(b-q)^2}[a.sin(bL) - (b-q)cos(bL) \\ &+[-a.sin(qL) + (b-q)cos(qL)]exp(-\alpha L)]\} \end{aligned}$$

B. Bit Error Rate (BER) Derivation

At the output of the fiber, the Optical CDMA signal is received by an optical correlator receiver with balanced photodetectors. The output of the photo detector is passed through an integrator [3]. During propagation through the fiber, the optical pulse undergoes intensity fluctuation due to XPM. If $sout_i(t)$ represents the normalized output pulse ($A_i(t)$) shape due to XPM for the i-th user, then the mean of the output of the correlator matched to the i-th user, is given by equation (14)[3]

$$U = \frac{RP_R}{4T_b} \int_0^{T_b} \sum_{l=0}^{N-1} sout_i(t - lT_c)dt \tag{14}$$

The variance of interference due to Multiple Access Interference (MAI) is given by equation (15)[3]

$$\sigma^2 = U^2 . \frac{2(K-1)}{3N} \tag{15}$$

where K is the number of simultaneous users (the number of pump channels here), N is the number of chips in one bit duration = $\frac{T_b}{T_c}$, T_b is duration of one bit, T_c is duration of one chip, R is the responsivity of pin photodiode, P_R is the the received optical power.

The signal to noise ratio at the correlator output can be obtained as (16) [3]

$$SNR = \frac{U^2}{\sigma^2 + \sigma_n^2} \tag{16}$$

The noise considered in this system are the receiver shot noise and thermal noise. Considering the thermal noise to be predominant, it is assumed that the noise for transmitted "1" and "0" are of same magnitude. The total noise variance is obtained by summing the variances of the individual noise components (17):

$$\sigma_n^2 = \sigma_{shot}^2 + \sigma_{thermal}^2 \tag{17}$$

The variance of the shot noise is given by (18) [3]:

$$\sigma_{shot}^2 = \frac{2qRKP_r}{4T_b} \tag{18}$$

The variance of the thermal noise is (19) [3]:

$$\sigma_{thermal}^2 = (4kT)B_r R_L \tag{19}$$

where k= Boltzman constant, T= receiver temperature (K), T_b= one bit duration, B_r= receiver bandwidth, R_L= load resistance of the receiver, K= number of simultaneous users, R= receiver sensitivity and q= electron charge.

The Bit Error Rate (BER) for the OCDMA transmission system is the given by equation (20)[3]

$$BER = 0.5erfc(\frac{\sqrt{SNR}}{\sqrt{2}}) \tag{20}$$

IV. Results and Discussions

The Bit Error Rate is analyzed against received probe power for different values of fiber length, bit rate and number of users.

Following the analytical approach, we evaluate the XPM induced power in the probe channel due to pump channel. The parameters used in calculation are: the laser of probe channel has wavelength of λ_1 = 1550 nm and laser of pump channel also has wavelength of λ_2 = 1550 nm. So the walk off parameter $d_{12} = \lambda_1 - \lambda_2$ is zero. Non liner refractive index $n_2 = 3x10^{-20}m^2/W$, probe channel group refractive index of pure silica is n=1.5 and group velocity of probe signal is = $2x10^8$ m/s, fiber core effective area is A_{eff} = 67.433 μm^2, fiber loss is α = 0.2 dB/km, fiber dispersion coefficient is D = 16.4 ps/nm-km, receiver temperature is T = 300 K, load resistance of the receiver is R_L = 50 Ω and photodetector responsivity is R = 0.8.

The reference chip length is 127 chip m-sequence, input probe power is P_{10} = 0 dBm and system bit rate is R_b = 10 Gbps.

Fig. 2 shows the plot of BER Vs received probe power as a function of different fiber lengths at 127 chip m-sequence, while the length of the fiber is varied from 600 km to 800 km at intervals of 100km.

From Fig. 2 it is found that to achieve a BER of 10^{-9}, the required received probe power is -7.5 dBm, -6.5 dBm and -6 dBm to transmit a distance of 600, 700 and 800 km respectively.

Fig. 3 shows the plot of BER Vs received probe power as a function of bit rate. Bit rate is taken as 5, 10 and 15 Gbps. For a specific received probe power, BER deteriorates with the increase in bit rate.

For example, to achieve a BER of 10^{-9} received prove increases form -7 dBm to -6.5 dBm as the bit rate is increased from 5 Gbps to 15 Gbps.

The results reveal (Fig. 2 and Fig. 3) that the XPM effect is significant at higher fiber length and higher bit rate and the system suffers power penalty in probe power in order to cope up with the XPM induced intensity fluctuation.

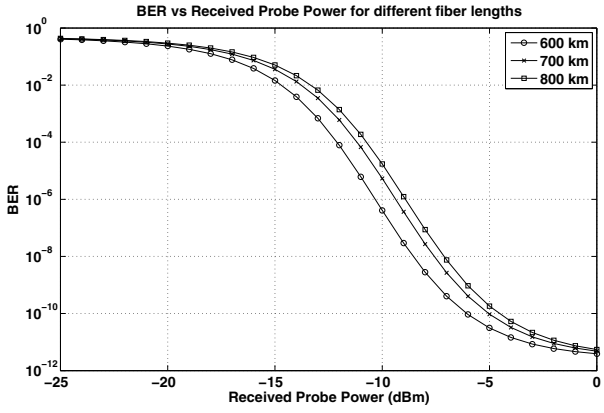


Fig. 2. BER Vs received probe power for 127 chip m-sequence with different values of fiber length (600, 700, and 800 km)

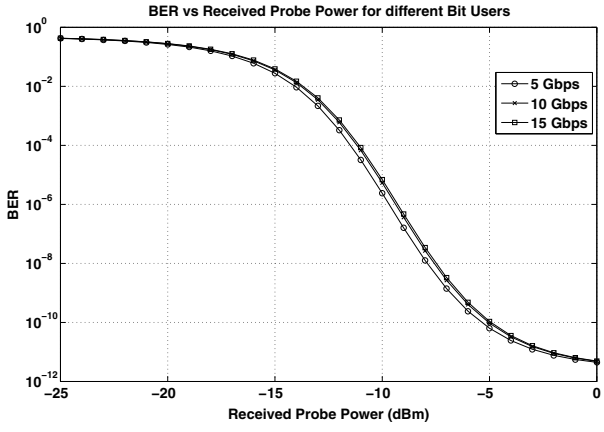


Fig. 3. BER Vs received probe power for 127 chip m-sequence with different values of bit rate (5, 10, and 15 Gbps)

Fig. 4 shows the plot of BER Vs received probe power for different of number of users. Number of user is taken from 6 to 22 at intervals of 4. For a particular received probe power, BER degrades with the increase in number of users.

Moreover, it is found that it is not possible to achieve a BER of 10^{-9} if the number of users is greater than 10.

Fig. 5 shows the plot of BER Vs number of users for different probe powers. The fiber length is 700 km and 127 chip m-sequence is used.

This curve (Fig. 5) reveals that to maintain a constant BER of 10^{-9}, the received probe power must be greater than -15 dBm for any number of user. For a received probe power of -10 dBm the maximum number of users can be 5. To increase the number of users to 7 the received probe power has to be greater than -5 dBm.

Fig. 6 shows the plot of BER Vs number of users for a fiber length of 700 km for 127 chip m-sequence.

From the plot it is found that to maintain a BER of 10^{-9}, the maximum number of users can be 7 (here the input probe power is 0dBm).

Fig. 7 shows plot of BER Vs received probe power for code length of 127 chip m-sequence and fiber length of 700 km and 5 number of users.

Fig. 8 shows plot of BER Vs received probe power for

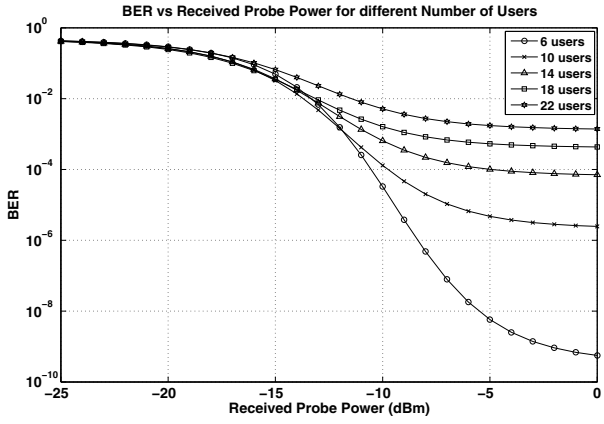


Fig. 4. BER Vs received probe power for 127 chip m-sequence with different number of users (6, 10, 14, 18 and 22)

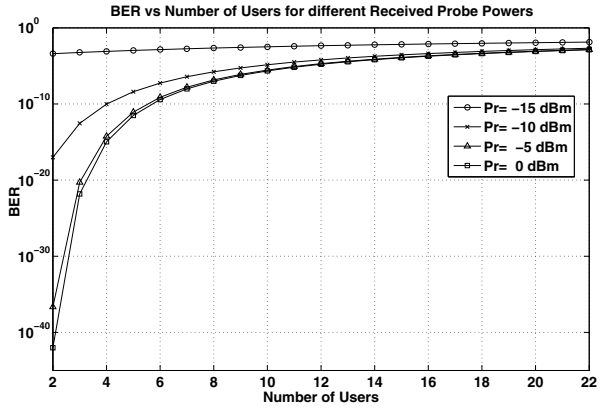


Fig. 5. BER Vs number of users for 127 chip m-sequence with different values of received probe power (-15, -10, -5 and 0 dBm)

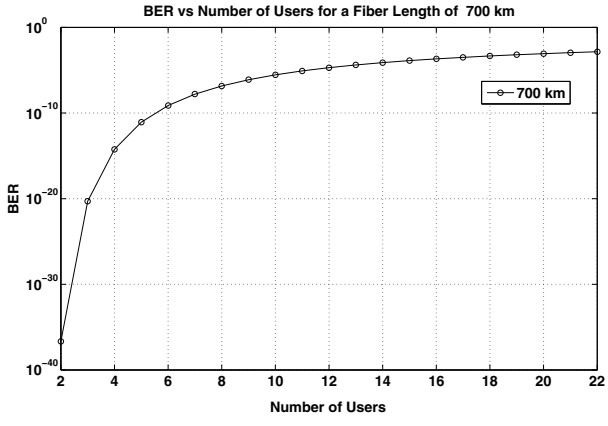


Fig. 6. BER Vs number of users for 127 chip m-sequence with a fiber length of 700km

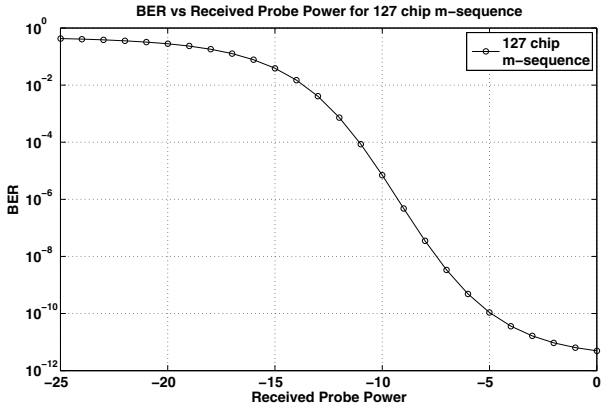


Fig. 7. BER vs received probe power for code of 127 chip m-sequence for a fiber length of 700 km & 5 users

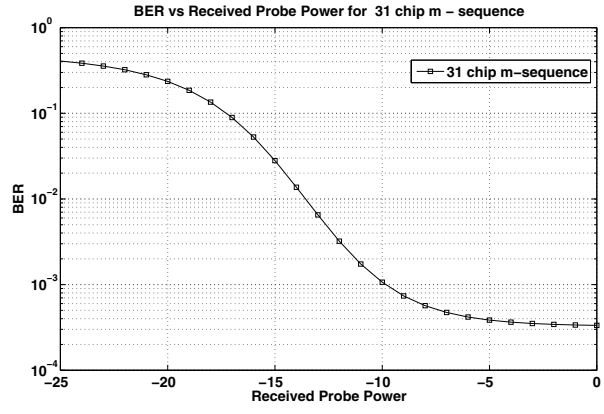


Fig. 8. BER vs received probe power for code of 31 chip m-sequence for a fiber length of 700 km & 5 users

code length of 31 chip m-sequence and fiber length of 700 km and 5 number of users.

It is found that if input probe power is 0 dBm in order to achieve a BER of 10^{-9} or better, received probe power should be greater than -7 dBm for 127 chip m-sequence code. And it is not possible to attain BER of 10^{-9} for 31 chip m-sequence code (if input probe power is 0 dBm).

V. Conclusion

An analytical approach is presented to evaluate the impact of XPM induced phase and intensity fluctuation on the bit error rate performance of an optical CDMA system. It is found that the system suffers significantly in BER performance due to multiple access interference in presence of XPM and GVD. The XPM effect is severe at longer fiber length and higher bit rate. But there is an improvement in the BER performance with increase in the code length.

References

[1] G. P. Agrawal, "Nonlinear Fiber Optics", 2nd ed. San Diego, CA: Academic, 1995.

[2] Adolfo V. T. Cartaxo, Member, IEEE, "Cross-Phase Modulation in intensity Modulation-Direct Detection WDM Systems with Multiple Optical Amplifiers and Dispersion Compensators", *Journal of Lightwave Technology*, vol. 17, No. 2, February, 1999.

[3] S. P. Majumder, Afreen Azhari, Abbou Fuad Muhammad, "Impact of Fiber Chromatic Dispersion on the BER performance of an Optical CDMA IM/DD Transmission System", *Photonics Technology Letters, IEEE*, vol. 17, Issue 6, June 2005.

[4] Ezmir M. R., S. P. Majumder, A. F. Muhammad, "Eye Penalty Due to Cross-Phase Modulation (XPM) in a Single Segment WDM IM/DD Transmission System", *Advanced topics in optoelectronics, microelectronics, and nanotechnologies: Conference, Bucharest , Rumania*, vol. 5227, pp. 382-389, 2003.

[5] Ting-Kuang Chiang, Noboyuki Kagi, Michel E Marchie, Leonid G. Kazovsky, "Cross-Phase Modulation in Fiber Links with Multiple Optical Amplifiers and Dispersion Compensators", *Joutnal of Lightwave technology*, vol. 14, No. 3, March 1999.

[6] John M. Senior, "Optical Fiber Communication", 2^{nd} edition, Prentice-Hall of India, New Delhi, 2004.

[7] Djafer K. Mynbaev, Lowell L. Scheiner, "Fiber-Optic Communications Technology", Pearson Education, 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

RF MEMS Tunable Filter: Design, Simulation and Fabrication Process

Md. Fokhrul Islam, M. A. Mohd. Ali, B. Y. Majlis and Nowshad Amin

Department of Electrical, Electronic and System Engineering
Faculty of Engineering
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, MALAYSIA
Email: engr_tutul96@yahoo.com

Abstract - This paper deals with a tunable bandpass filter topology which controls both the central frequency and bandwidth. This tunable filter results from the association of MEMS (microelectromechanical system) fixed-fixed beam, used as variable capacitors, with an original passive topology. The design proposed in this paper is the metallic bridge over a coplanar microstrip couple line suspended in both sides by two metallic posts. This MEMS filter is designed for operation at X-band satisfying the implementation requirements for radar and modern multi-band communication systems. The filter occupies a chip area of 12 × 3 mm^2 and achieved an insertion loss of only 0.5 dB and return loss of less than -15 dB throughout the operation band, thereby tuning the filter center frequency by 11% between 8.76 to 9.8 GHz. The S parameters and electromechanical displacement simulations are also presented.

I. Introduction

The growing number of multi-standard and multi-application telecommunication systems has led to the development of new tunable filter topologies. Nevertheless, neither the central frequency nor the bandwidth is independently and simultaneously controlled by the existing tunable filters [1]. Tunable filters are integral components in a variety of radar and modern multi-band communication systems. The conventional tunable filters typically utilize YIG resonators, active resonators or varactors as the tuning element. However, these varactor based tuning filters have low Q values due to high series resistance of diodes. The development in microelectromechanical system (MEMS) technology allows new and innovative design of the tunable bandpass filters [2, 3].

In this paper, the concept in [4] is extended to design Radio Frequency (RF) MEMS tunable filters for wireless application. By leveraging existing state-of-the-art of IC fabrication technologies, MEMS technology exhibits many advantages indigenous to IC technologies such as cost, size and weight reduction. These advantageous characteristics have positioned MEMS as a wining technology in many application areas, including accelerometers, pressure sensors, micro-optics, and ink-jet nozzles. New developments in satellite communications as well as advances in the area of millimeter-wave multimedia services require high performance components, and RF MEMS can fulfill that need by providing critical reductions in power consumption and signal loss, thereby extending battery life or reducing weight. Parallel-coupled Bandpass Filters (BPFs) are more favorable to be used for planar microstrip filters in modern microwave and wireless communication system, due to its weightless, low cost and easy integration [5]. In this context, a novel reconfigurable parallel-coupled BPF in microstrip platform with tunable bandwidth at the passband is presented. The proposed filter differs from the reported MEMS filters in that the bandwidth tunability is obtained by shunt capacitance variation across the resonating stub rather than the conventional length switching. Performance of the presented tunable filter is characterized by using the Electro-Magnetic (EM) simulations.

II. Basic Filter Topology

In order to design the tunable BPF, a 4th order parallel coupled BPF has been designed on a microstrip platform using coupled half-wave resonators to give a maximally flat response. Microstrip technology has been used for simplicity and ease of fabrication [6]. Although miniaturized end coupled bandpass filters are widely used for tunable applications, owing to the loose coupling of their resonators they account for high insertion losses and poor performance. This bandpass filter is designed by following the design procedure based on the even- and odd-mode impedances of the coupled lines [7], and is further optimized using IE3D.

The filter is designed on silicon substrate (ε_r=11.7) of height 600μm using Polygon-Based Layout Editor MGRID. With this, the filter requires even- and odd-mode characteristic impedances (Zoe, Zoo) of 142Ω and 42Ω, respectively, for the first coupled line section, which translates to a line width of 75μm and line gap of 50μm on a 600μm silicon substrate. The next coupled line section requires Zoe and Zoo of 114Ω and 34Ω, respectively, yielding a line width of 150μm and line gap of 50μm. The last two coupled line sections are symmetrical to the first two, thus they have the same dimensions as stated earlier. All the quarter-wave coupled lines have lengths of 2650μm at 10GHz. The

dimensions of the I/O ports are corresponding to a 50Ω microstrip line which could be consider as subminiature version A (SMA) adapter. Fig. 1 and Table 1 show the top view and various design parameters along with the amount of impedances of the filter structure respectively.

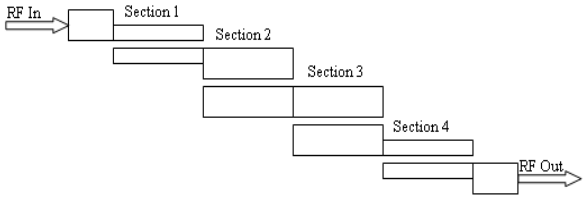


Fig. 1 Top view of the simulated BPF structure

Table 1 Dimensions of the filter structure with the impedances

Parameters	Section 1	Section 2	Section 3	Section 4
Width (μm)	75	150	150	75
Spacing (μm)	50	50	50	50
Length (μm)	2650	2650	2650	2650
Even-mode impedances (Ω)	142	114	114	142
Odd-mode impedances (Ω)	42	34	34	42
Port impedances (Ω)	50	-	-	50

III. Tuning Principle

The concept of tunability is incorporated into the basic filter design by a set of low loss RF MEMS variable capacitors. The cross section of a metal membrane variable capacitor is shown in Fig. 2. The RF MEMS capacitor consists of a lower electrode which fabricated as a filter circuit and a thin aluminum membrane suspended over the electrode. The membrane is connected directly to grounds on either side of the electrode. The air gap between the two conductors determines the MEMS capacitor off-capacitance. With no applied actuation potential, the residual tensile stress of the membrane keeps it suspended above the RF path as shown in Fig. 2a.

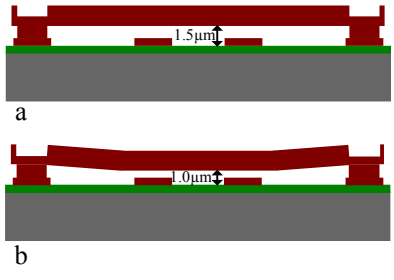


Fig. 2 Cross-section of an RF MEMS capacitor in the (a) unactuated and (b) actuated positions

When using this variable capacitor as a filter circuit, a dc-bias voltage V_p is applied to the membrane conductor, while an ac excitation signal $v_i = V_i \cos \omega_i t$ is applied to the underlying electrode. In this configuration, the difference voltage $(v_i - V_p)$ is effectively applied across the underlying electrode to suspended membrane, generating a force between the stationary electrode and the moveable beam given by [8].

$$F_d = \frac{\partial E}{\partial x} = \frac{1}{2}(v_i - V_p)^2 \frac{\partial C}{\partial x}$$

$$F_d = \frac{\partial C}{\partial x}\left(\frac{V_p^2}{2} + \frac{V_i^2}{4}\right) - V_p \frac{\partial C}{\partial x} V_i \cos \omega_i t + \frac{\partial C}{\partial x}\frac{V_i^2}{4} \cos 2\omega_i t \quad (1)$$

The first term in equation (1) represents an off-resonance dc force that statically bends the beam. The second term constitutes a force at the frequency of the input signal, amplified by the dc-bias voltage V_p which creating a variable capacitance between the electrode and suspended beam. The third term in (6) represents a term capable of driving the beam into vibration, if V_p is very large compared with V_i, this term is greatly suppressed. Fig. 2b demonstrates an RF MEMS variable capacitor in the actuated state. In this state, the characteristics of the capacitor determine the MEMS on-capacitance. By virtue of changing the position of the membrane with an applied DC voltage, the capacitance of this RF MEMS device can be changed over a significant capacitance range.

IV. Design and Simulation

A typical design flow for an RF MEMS device is shown in Fig. 3. From mechanical and electrical specifications, the device is designed iteratively until both sets of specifications are satisfied. Within the iteration loops, the software may be used to explore, and correct for, the impact of a number of potential influences on device performance/yield, such as package stress and strain state, manufacturing process or temperature variability, and hermiticity. The design process ends with an S-parameter file and/or a behavioral reduced order model for subsequent utilization in a circuit/system simulator.

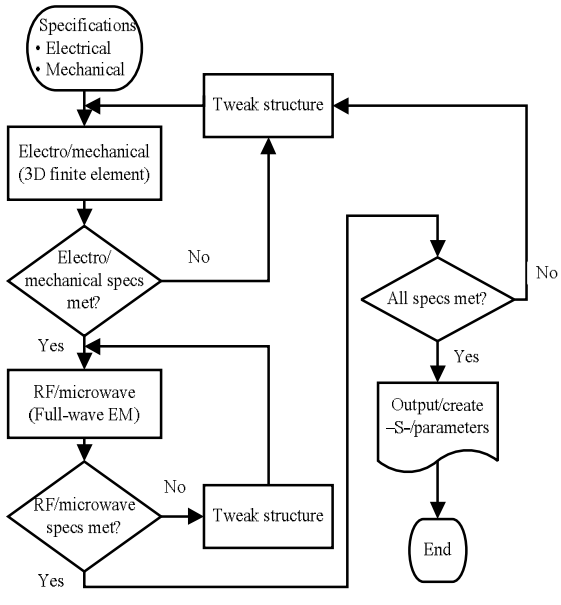


Fig. 3 RF MEMS device design flow

A. Electromechanical Simulation

When designing a MEMS device, a good electromechanical simulation tool is the commercial software ABAQUS, that uses finite element method. It provides interaction between the electrostatic force analysis and structural analysis. The Fig. 4 Shows a 3D deformation analysis of the MEMS fixed-fixed beam which is applicable for variable capacitor. This electromechanical system is simulated in Abaqus with the electrostatic force applied to the bottom surface of the beam as a ditributed load. The 3D Abaqus simulation provides good mechanical accuracy by autometically including the effects of geometric nonlinearity (or stress stiffening), compliant stepups and contact. The Abaqus model uses plane stress elements for the beam itself. For the accuracy of analysis, the ABAQUS element type C3D8I is used to model the beam. Effects of fringing fields and finite plate thickness are included in our electrostatic force model. To simplify the analytical simulation of a parrallel-plate variable capacitor with such beams, an uncoupled model is considered to find the force-displacement relation for the beam.

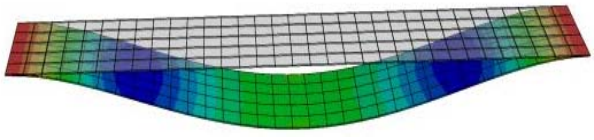


Fig. 4 Deformation structure of fixed-fixed beam

Fig. 5 presents the force-displacement characteristics of a fixed-fixed beam used to obtain the variable capacitor. As can be seen, the applied force of 2 μN is strong enough to attract the beam to a maximum displacemnt of 0.5 μm which is 1/3 of the original airgap.

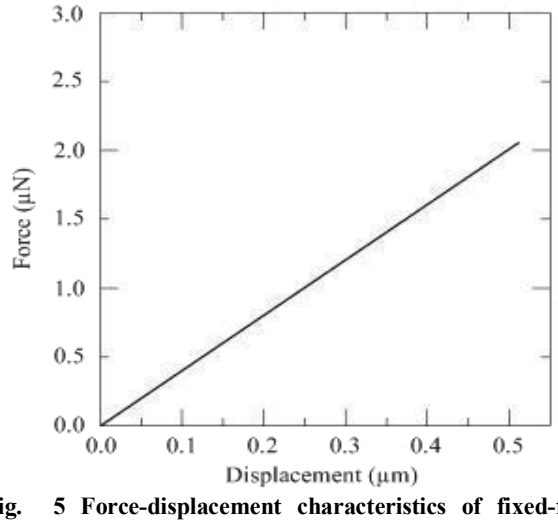


Fig. 5 Force-displacement characteristics of fixed-fixed beam

B. Electromagnetic Simulation

The filter was simulated in the layout window of IE3D software [9], including the silicon nitride (Si_3N_4) layer. The insertion loss and return loss parameters of the tunable filter are shown in the Fig. 6. Two states of the MEMS bridge were simulated: In the UP state, when the bridges are at a height of 1.5 μm the filter has a center frequency of 9.8 GHz with a 3dB bandwidth of about 45%. In the DOWN state, when the bridges are at the minimum height of 1.0 μm the center frequency shifts to a value of about 8.76 GHz with a 3dB bandwidth of about 69%.

The bandwidth tuning is almost complete over the entire band of operation. The simulated filter occupies a chip area of 12 × 3 mm^2 and achieved an insertion loss of only 0.5 dB and return loss of less than -15 dB throughout the operation band. A small frequency shift is encountered when compared to the initial center frequency due to the phase delay introduced by the air bridges and can be rectified by adding a small incremental length to the original length of the resonator. The 3dB bandwidth of the filter tunes towards the left side of the response from 45% to 69% for corresponding variations in the height of the bridge will be controlled by the actuation voltage. This filter is widely used today in radar, satellite and terrestrial communications, and electronic countermeasure applications, both militarily and commercially.

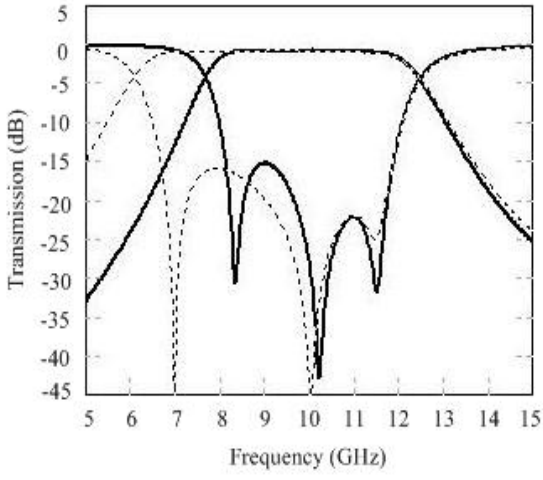


Fig. 6 Transmission characteristics of the tunable filter

V. Fabrication Process

There are three basic building blocks in MEMS technology, which are the ability to deposit thin films of material on a substrate, to apply a patterned mask on top of the films by photolithographic imaging, and to etch the films selectively to the mask. A MEMS process is usually a structured sequence of these operations to form actual devices as shown in Fig. 7. In the Fig. 8 presents a complete fabrication process flow of a MEMS tunable bandpass filter. The filters are constructed on a 2×3 cm^2 silicon substrate, with 600 μm thickness and dielectric constant $\varepsilon_r = 11.9$.

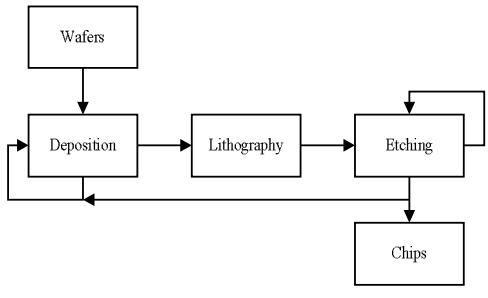


Fig. 7 MEMS fabrication process flow

The fabrication process of three mask levels is illustrated in Fig. 8 and described bellow:

- Deposition of 2000 Å silicon nitride on a silicon substrate;
- Deposition of 0.5 µm aluminum by resistive evaporation; followed by spining of thin layer of AZ1500 photoresist, patterning of the basic filter structures using UV photolithographic process and aluminum chemical etch;
- Depositon of 1.5 µm sacrifacial layer of SOG (SiO_2), used to define g_0 gap dimension and then, the patterning of the windows related to the bridge columns is made. A hard bake is needed to avoid damages to the sacrifacial layer;
- Deposition of 1.5 µm aluminum bridge layer by sputtering and patterning the bridge using UV photolithographic process and acetone lift off;
- Finally remove the sacrifacial layer of SOG using buffer oxide etching (BOE) to release the overall structure.

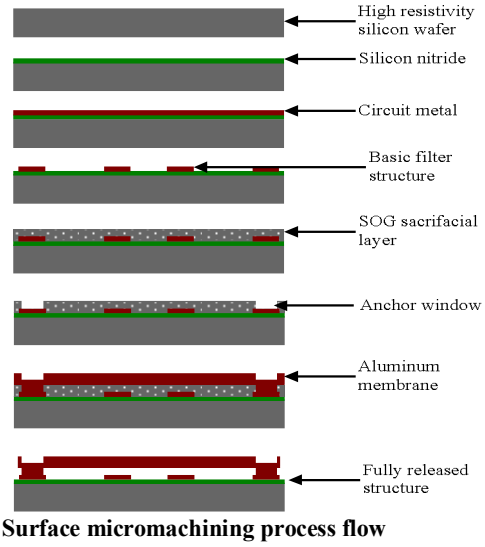


Fig. 8 Surface micromachining process flow

A. Preliminary Prototype

Surface micromachining techniques were utilized to fabricate the RF MEMS filters shown in Fig. 9 below. Among currently used MEMS process technologies, surface micromachining is widely used due to its similarity to thin-film technology used for integrated circuits. Surface micromachining is already proven to be a technology that is comercially viable since it has supported high-volume manufacture of MEMS devices. During the development of the process many changes were made to optimize it and some results did not fulfill the expectations, however every change demonstrates the viability of the fabrication.

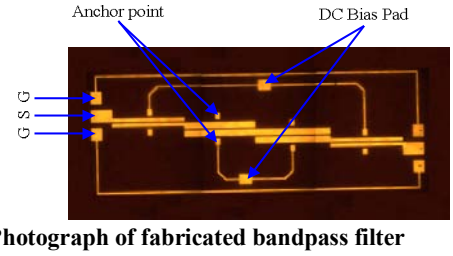


Fig. 9 Photograph of fabricated bandpass filter

VI. Conclusion

This paper described an original tunable bandpass filter topology. This tunable filter is obtained by associating a microstrip bandpass filter and MEMS variable capacitors. Such an association allows tuning simultaneously and independently both central frequency and bandwidth. A design technique has been introduced which allows these devices to be eventually used to make high performance tunable filters using only capacitor tuning. These tunable filters will have application in a variety of tunable filter, oscillator, and receiver applications. A simple process to fabricate MEMS tunable filter for multiband applications is presented. The electromechanical simulation shows accordance to the calculated pull-down force and the simulated S parameters in the UP/DOWN Stages, showed the RF switching function realized by the switch. Further works will be present a set of measurements to fully characterize the filter, including: insertion loss, isolation, response time, among other physical properties.

Acknowledgement

The authors would like to thank the Ministry of Science, Technology and Innovation (MOSTI) of Malaysia for supporting this work under the eScienceFund 03-01-02-SF0254.

References

[1] E. Foum et al., “Bandwidth and central frequency control on tunable bandpass filter by using MEMS cantilevers” IEEE MTT-S Digest, 2003, pp. 523-526, 2003.

[2] Y. Liu, A. Borgioli, A.S. Nagra, Robert A. York, “Distributed MEMS transmission lines for tunable filter applications”, Int. J. RF Microwave Comput. Aided Eng., Special Issue 11, pp. 254-260, 2001.

[3] A. Tamijani, L. Dussopt, G.M. Rebeiz, “Miniature and tunable filters using MEMS capacitors”, IEEE Trans. Microwave Theory Tech. vol. 51, pp. 1878-1885, 2003.

[4] Md. Fokhrul Islam, M. A. Mohd. Ali and B. Y. Majlis, “Parallel coupled microstrip bandpass filter for x-band applications”, DIUJST Vol. 2, Iss. 2, pp. 27-31 July 2007.

[5] WU H.-W., WENG M.-H., SU Y.-K., YANG R.-Y. and HUNG C.-Y, “Spurious suppression of a parallel coupled microstrip bandpass filter with simple ring EBG cells on the middle layer”, IEICE TRANS. ELECTRON., vol.E89-C, np.4, pp. 568-570, April 2006.

[6] T. C. Edwards and M.B. Steer, Foundation of Interconnect and Microstrip Design, John Wiley & Sons Ltd., 2000.

[7] D. M. Pozar, Microwave engineering, John Wiley & Sons Inc., 2005.

[8] M. Lobur, T. Sviridova and K. Baybakov, “RF MEMS: filter model”, in Proc. TCSET’2004, Lviv-Slavsko, Ukraine, pp. 92-93, 24-28 February 2004.

[9] Zeland Software, IE3D Full Wave Electromagnetic Simulator V12.14, 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

VLS Growth of Doped Si-Microprobe Arrays Using Varying PH_3 Flow with a Fixed Flow of Si_2H_6 at Low Temperature

Md. Shofiqul Islam[1,*], Kazuaki Sawada[2], and Makoto Ishida[2]

[1] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[2] Department of Electrical and Electronic Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan
* Corresponding Emails: shofiqul@eee.buet.ac.bd shofiq1992@yahoo.com

Abstract— Vapor-liquid-solid (VLS) growth, using Si_2H_6 as the gas source of Si, can be used to realize intrinsic Si microprobe arrays, which could be doped by diffusion process (at 1100ºC) after VLS growth. But in this work we have demonstrated that by incorporating *in-situ* doping using the gas source of Si_2H_6 and PH_3 with VLS growth process, doped n-Si microprobes can be realized directly at a temperature (around 700ºC) lower than that required at diffusion process. Here we report the realization of n-Si microprobe arrays at low temperature by using Au-catalyzed selective VLS growth using varying flow of PH_3 with a fixed flow of Si_2H_6. The effects of PH_3 flow rates on the physical and electrical properties of these n-Si microprobes, under the condition of a fixed amount of Si_2H_6 flow, have been investigated in detail.

I. Introduction

Because of the increasing demand of low-dimensional Si microstructures for various applications, the researchers are showing interest to realize needle-like Si microprobes and seek out their potential applications. Literature [1] shows that probe (needle)-like Si single crystal can be grown in <111> direction perpendicular on Si (111) substrate by vapor-liquid-solid (VLS) growth method. This method starts with the formation of dots of a metallic catalyst (normally Au) on Si substrate. When the sample with Au dots is heated in a vacuum chamber, Au particles mix with Si atoms from the substrate and thus form Au-Si alloy liquid droplets on the Si surface. Now if the gas source of Si is introduced into the chamber, the alloy droplet absorbs Si atoms from the gas source and becomes supersaturated with Si and then Si atoms begin to precipitate at the interface of the alloy droplet and Si surface. As long as the supply of Si source remains available, the precipitation continues and thus probe-like Si crystal grows perpendicular to the Si surface.

Some researchers already reported some applications of VLS grown Si probes [2-3]. For example, Asai et al. [2] applied VLS-grown Si probes as the pins for developing high-resolution probe-card for integrated circuit (IC) testing. Previously, our group studied the feasibility of applying needle-like intrinsic Si microprobes as insertion electrodes for sensing the signal from neurons [3].

From the literature it is found that almost all of the previous researchers reported intrinsic type Si microprobes grown by VLS using gas source of Si only. But these intrinsic Si microprobes exhibit high resistivity; for example, we employed VLS growth using Si_2H_6 and realized intrinsic Si microprobes whose resistivity was found as 10^4 Ω-cm. These high resistivity poses barrier to use these probes for some applications; for example, Si-microprobe electrode arrays developed by our group [3] require highly conductive probes in order to record the small signal from the neurons. Therefore, it demands the doping of Si microprobes. Again in the applications mentioned earlier, VLS grown Si microprobes were used as passive elements. In order to utilize these microprobes for an active device fabrication it also requires the doping of these Si microprobes.

In order to meet up the demand of doping, intrinsic Si microprobes, grown by VLS using Si_2H_6, were doped by phosphorous (P) diffusion at 1100ºC [4]. But if we want to integrate these doped probes or probe-related devices with on-chip circuitry, this high-temperature diffusion would be detrimental to on-chip devices. On the other hand lower-temperature-diffusion cannot dope the probes sufficiently; at 900ºC the depth of doping was found to be 0.5 μm from the surface of the probe sidewall [5]. Thus temperature issue puts limitation on realizing highly conductive probes with on-chip circuitry by conventional way of doping by thermal diffusion.

Therefore, in order to obtain the doped Si microprobes at lower temperature, we introduced an alternate approach by incorporating *in situ* doping into VLS growth system and thus we could obtain doped Si microprobes at low temperature (around 700ºC). In this work we incorporated *in situ* doping into VLS growth system using the varying flow of PH_3 with a fixed flow of Si_2H_6 and thus we could obtain doped n-Si microprobe arrays at around 700ºC. Here we investigated the influence of PH_3 flow rate variation on the properties of these n-Si microprobes.

II. Experimental

Here n-Si(111) wafer has been used as the substrate in our experiment to grow n-Si microprobes by VLS method. Fig. 1 illustrates the fabrication process. At first, a layer of SiO_2 with 870 nm in thickness was formed over n-Si(111) substrate as shown in Fig. 1a by wet oxidation at 1000ºC . Then the photolithography and etching with buffered hydrofluoric acid (BHF) were carried out to create arrays of circular windows through SiO_2 layer as in Fig. 1b. A thin film of Au (Fig. 1c) with 170 nm in thickness was deposited over this patterned structure by using evaporation technique. Au film from the resist-surface was removed by using a lift-off process, however circular Au dots remained at pre-determined microprobe sites where the Si surface was revealed through SiO_2 window as shown in Fig. 1d. Then the sample, having these circular Au dots, was inserted into a high-vacuum gas-source molecular-beam-epitaxy (GS-MBE) chamber.

The chamber was equipped with 100% Si_2H_6 as Si source and 1% PH_3 (diluted in 99%H_2) to facilitate *in-situ* doping of the probes grown by VLS mechanism. The system includes mass flow meters for controlling the flow rates of PH_3 and Si_2H_6 to vary the dopant -to-silicon ratio in the inlet gas system.

The sample was annealed at a temperature around 700ºC to form Au-Si alloy droplets inside the SiO_2 windows as shown in Fig. 1e. Then, the mixed gas of PH_3 and Si_2H_6 with desired flow rates were supplied to the growth chamber and hence n-Si microprobes with desired doping were grown on n-Si(111) substrate by VLS mechanism as shown in Fig. 1f.

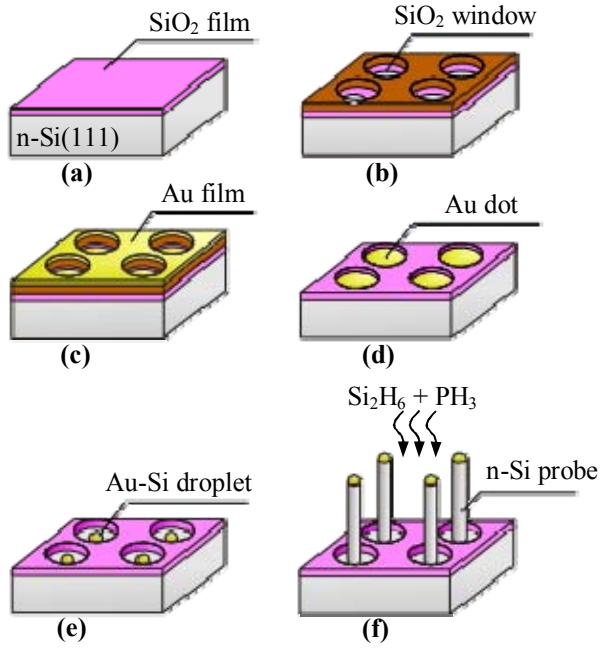


Fig. 1. Fabrication process of n-Si microprobe array: (a) SiO_2 layer formation; (b) circular window through SiO_2 made by photolithography; (c) Au film deposition; (d) lift-off Au from resist-site; (e) annealing to form Au-Si alloy droplet; (f) introduction of mixed gas of PH_3 and Si_2H_6 into the growth chamber and then n-Si microprobes grow by VLS mechanism.

III. Results and Discussions

Fig. 2 shows the SEM image of a typical 6×7 array of n-type Si microprobes having the length of 54 μm and diameter of 3.6 μm, which were grown by *in-situ* doping VLS growth at 690°C at the growth pressure of 5.3×10^{-3} Pa using PH_3 flow at a rate of 1.00 sccm with Si_2H_6 flow of 1.70 sccm for 90 min.

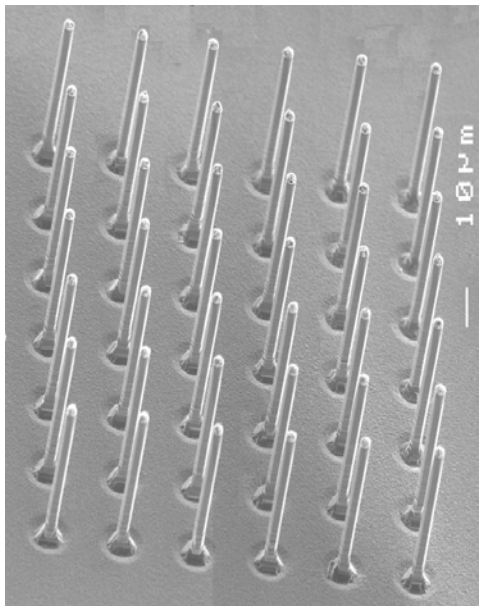


Fig. 2. SEM image of n-Si microprobe array with probe-length of 54 μm and probe-diameter of 3.6 μm, which were grown by *in-situ* doping VLS with PH_3 = 1.00 sccm and Si_2H_6 = 1.70 sccm.

About the physical properties, it is observed that for a certain set of growth conditions, probes grown at different sized Au dots are of equal lengths i.e. the probe-length is found independent of Au dot size. However probe-diameter was found to vary with the diameter of Au dot pattern (circular); Fig. 3 shows such a typical variation. Similarly, probe-diameter also varies with Au film thickness. Thus by forming Au dots with desired size and thickness, probe-diameter could be controlled in the range of 1~5 μm. Whereas, by placing SiO_2 windows and hence the Au dots at the desired sites, Si microprobes can be selectively grown at desired positions.

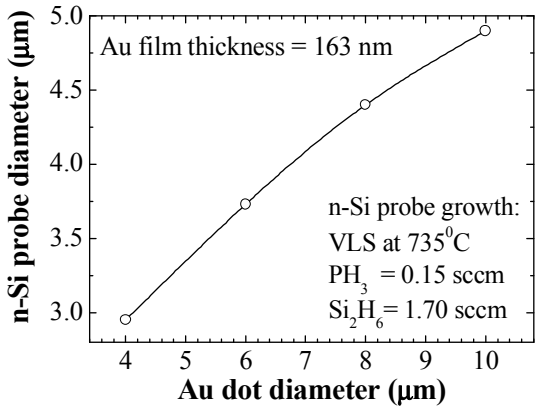


Fig. 3. Variation of the diameter of n-Si microprobe with the diameter of Au-dot-pattern.

In order to investigate the effects of PH_3 flow rates on the properties of n-Si microprobes, VLS growths were carried out with different PH_3 flow rates at different growths with a fixed amount of Si_2H_6 flow at all growths. In this work PH_3 flow rate was varied from 0.15 to 1.70 sccm while Si_2H_6 flow was maintained at a fixed level of 1.70 sccm.

The effect of PH_3 flow on the growth rate has been studied. It is found that the growth rate of n-Si microprobe decreases with the increasing flow of PH_3 as shown in Fig. 4. Some literature [6-7] report that the introduction of PH_3 to Si_2H_6 during the epitaxial growth of Si by GS-MBE or ultrahigh vacuum chemical vapor deposition (UHV-CVD), results in the reduction of the Si growth rate. P atoms bind H atoms on the growing surface and thus block the available reaction sites for the adsorption of further precursor Si_2H_6, thereby decreasing the growth rate. Similar to the trend of Si growth in GS-MBE or UHV-CVD with respect to the effect of PH_3, the growth rate of VLS grown n-Si microprobe also decreases with the increase in PH_3 flow.

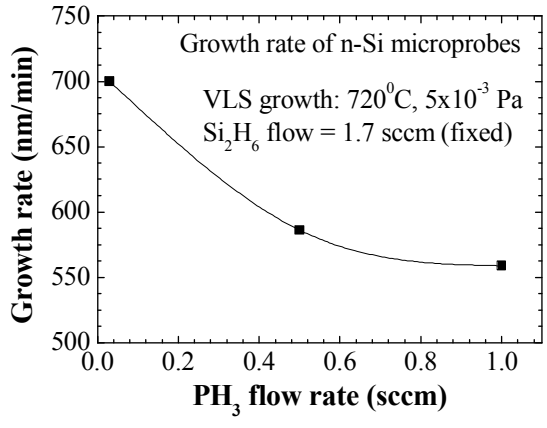


Fig. 4. Growth rate of VLS grown n-Si microprobe as a function of PH_3 flow rates used in VLS growth.

During VLS growth of n-Si microprobe arrays, a layer of n-type poly-Si also formed at the sites other than probe, since the starting of the growth in this region was over the surface of SiO_2 mask. Typical view of the surface of that n-poly Si is shown in Fig. 5. It is still to investigate the feasibility of using this by-product poly-Si for post-VLS device fabrication with Si microprobe arrays in one-chip.

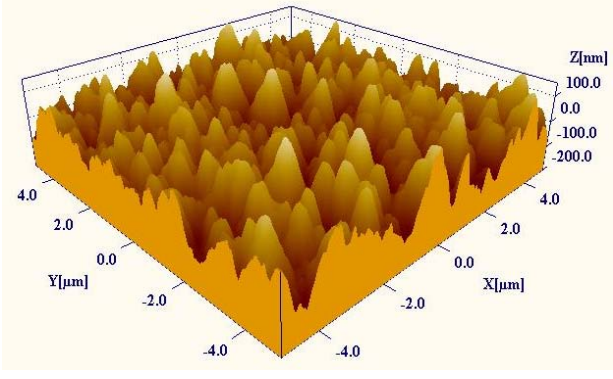


Fig. 5. The surface morphology of n-poly Si grown on SiO_2 mask during VLS growth of n-Si microprobe using PH_3 = 0.15 sccm and Si_2H_6 = 1.70 sccm at 735°C.

The growth rate of this n-poly Si also exhibits the decreasing nature with PH_3 flow as shown in Fig. 6 similar to that of n-Si probes. Although Si microprobe grows by VLS mechanism, poly-Si grows by vapor-solid (VS) mechanism. From the Fig. 4 and Fig. 6 it is found that Si probe grows at a rate 15-20 times higher than that of poly-Si. That is, Si crystal by VLS mechanism grows at a faster rate than that by VS growth. This is because VLS growth is found to happen at lower activation energy than that of VS method.

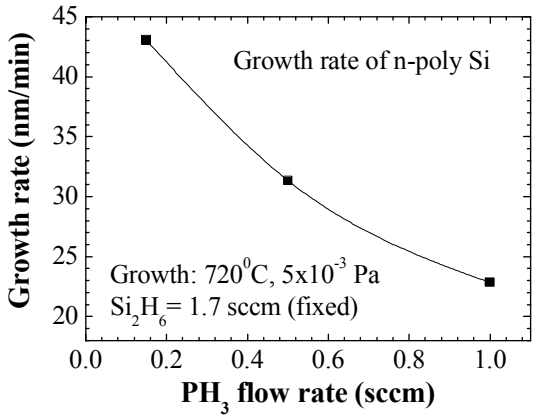


Fig. 6. Effect of PH_3 flow rates on the growth rate of n-poly Si grown as a by-product during VLS growth of n-Si microprobes.

Current-voltage (I-V) characteristic of n-Si microprobe, grown by VLS on n-Si (111) substrate, was measured by using one contact with tungsten (W) micro-needle at the tip and another contact at the base of the probe. Micromanipulator systems were used for handling the W-needles. The probes were found conductive in both directions.

The resistivity of these n-Si microprobes was found to decrease with the increase of PH_3 flow as shown in Fig. 7, which shows that resistivity of n-Si microprobes can be controlled in the range from 7×10^{-2} to 6×10^{-3} Ω-cm with the variation of PH_3 flow from 0.15 to 1.70 sccm with a fixed 1.70 sccm of Si_2H_6. Similarly, the impurity concentrations in n-Si microprobes could be controlled in the range of $10^{17} \sim 10^{19}$ cm^{-3} as shown in Fig. 8 using the aforementioned range of flow of PH_3.

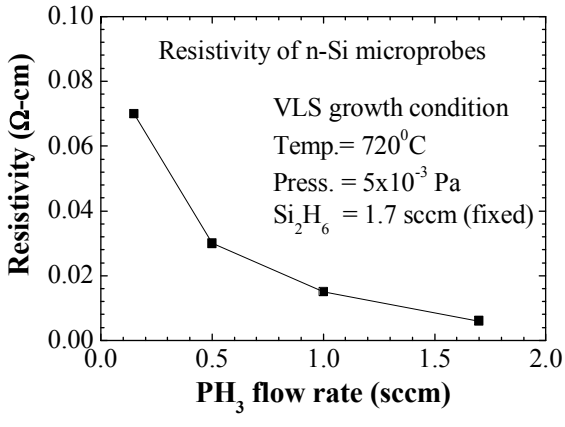


Fig. 7. Dependence of resistivity of VLS grown n-Si microprobe on the PH_3 flow rates used in VLS growth.

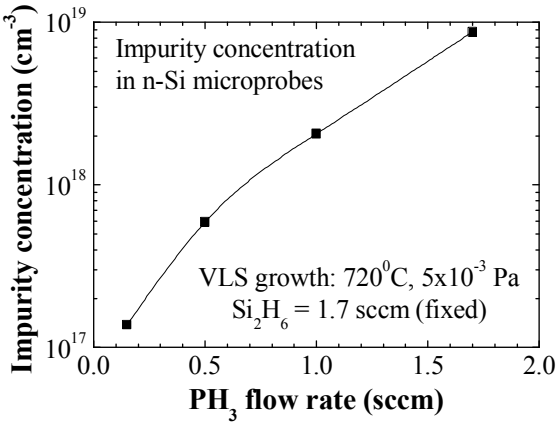


Fig. 8. Variation of impurity concentration in VLS grown n-Si microprobes as a function of PH_3 flow rates.

Electron mobility (μ_n) in these VLS grown n-Si microprobes was estimated using the theoretical relation, $\mu_n = \sigma/qn$ (where, σ = conductivity; q = electron charge; n = impurity concentration). The electron mobility is found to change within the range of 650 to 100 $cm^2\ V^{-1}\ s^{-1}$ with the variation of PH_3 flow from 0.15 to 1.70 sccm as shown in Fig. 9. Electron mobility in VLS grown n-Si microprobes are also found consistent with the reported [8] electron mobility in n-Si by other methods, for the doping levels investigated in our work.

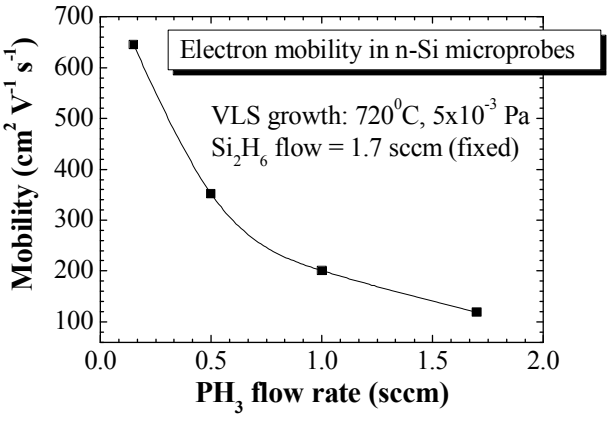


Fig. 9. Electron mobility in VLS grown n-Si microprobes versus PH_3 flow rates used in VLS growth.

IV. Applications

From the previous reports [2-3] on the applications of VLS grown Si microprobes, it is found that these applications require the realization of highly conductive Si microprobes with on-chip circuitry for processing the recorded information. However, VLS grown intrinsic Si probes exhibit high resistance, which can be reduced by doping the probes with thermal diffusion at 1100ºC [4], which is too high for on-chip circuitry. But in this paper we observed that highly conductive doped n-Si microprobe array can be fabricated by VLS at a lower temperature (around 700ºC) by incorporating *in-situ* doping with VLS growth system. Hence it would be compatible to process highly conductive doped Si microprobes with on-chip circuitry and thus promotes the realization of smart chip for various applications. Also, this detailed study of n-Si microprobes, carried out in this paper, would be useful while fabricating the arrays of vertical devices such as diodes, transistors etc. with n-Si microprobes. These vertical structured devices might be suggested to apply as insertion electrode into the living tissue for sensing the temperature distribution by using temperature dependence nature of device characteristic. They may also be used as pressure-sensor, photo-detectors, photosensitive emitter array devices etc.

V. Conclusions

The effect of PH_3 flow rates on the properties of n-Si microprobe arrays, realized by Au-catalysed VLS mechanism at low temperature (around 700ºC) using *in-situ* doping with Si_2H_6 and PH_3, has been demonstrated. These microprobes can be selectively grown with the diameter of 1~5 μm. Si microprobes by VLS mechanism are found to grow at a faster rate than Si growth by other processes. The growth rate of these n-Si microprobe was found to decrease with PH_3 flow. The resistivity of n-Si microprobe was observed to change in the range of 7×10^{-2} to 6×10^{-3} Ω-cm and impurity concentration in the range of $10^{17} \sim 10^{19}$ cm^{-3} by varying PH_3 flow from 0.15 to 1.70 sccm. Electron mobility changes within the range of 650 to 100 $cm^2\ V^{-1}\ s^{-1}$ for doping levels investigated. Due to the advantage of low temperature processing by *in-situ* doping VLS, it would be compatible to process highly conductive doped n-Si microprobe arrays with on-chip circuitry to develop smart chips for sensor applications. Also this study would be useful while fabricating the vertical devices (e.g. diodes, transistors etc.) with n-Si microprobes aiming to apply them in sensor applications.

References

[1] R.S Wagner, and W.C. Ellis, "Vapor-liquid-solid mechanism of single crystal growth," Appl. Phys. Lett., vol. 4, pp. 89-90, 1964.

[2] S. Asai, K. Kato, N. Nakazaki, and Y. Nakajima, "Probe card with probe pins grown by the vapor-liquid-solid (VLS) method," IEEE Trans. Component, Packaging, and Manufacturing Tech.-Part A, vol. 19, no. 2, pp. 258-267, 1996.

[3] T. Kawano, H. Takao, K. Sawada, and M. Ishida, "Multichannel 5×5-Site 3-dimensional Si microprobe electrode array for neural activity recording system", Jpn. J. Appl. Phys. Vol. 42, pp. 2473-2477, 2003.

[4] T. Kawano, Y. Kato, M. Futagawa, H. Takao, K. Sawada, and M. Ishida, "Fabrication and properties of ultrasmall Si wire arrays with circuits by vapor-liquid-solid growth," Sensors and Actuators, vol. A 97-98, pp. 709-715, 2002.

[5] T. Kawano, Y. Kato, R. Tani, H. Takao, K. Sawada, and M. Ishida, "Selective vapor-liquid-solid epitaxial growth of micro-Si probe electrode arrays with on-chip MOSFETs on Si (111) substrate," IEEE Trans. Electron Devices, vol. 51, pp. 415-420, 2004.

[6] M. Racanelli, and D. W. Greve, "*In situ* doping of Si and $Si_{1-x}Ge_x$ in ultrahigh vacuum chemical vapor deposition," J. Vac. Sci. & Technol. B, vol. 9, pp. 2017-2021, 1991.

[7] F. Gao, D. D. Huang, J. P. Li, Y. X. Lin, M. Y. Kong, D. Z. Sun, J. M. Li and L. Y. Lin, "Influence of phosphine flow rate on Si growth rate in gas source molecular beam epitaxy," J. Cryst. Growth, vol. 220, pp. 461-465, 2000.

[8] S. M. Sze, "Semiconductor devices, physics and technology", John Wiley and Sons, New York, 1985, p. 325.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

MEMS Switch for Designing a Multi-band Reconfigurable Antenna

A.H.M. Zahirul Alam, Md. Rafiqul Islam, Sheroz Khan, Soheli Farhana, Norsuzlin Bt. Mohd Sahar and Norasyikin Bt Zamani

Department of Electrical and Computer Engineering, Faculty of Engineering, International Islamic University Malaysia, P.O. Box 10, 50728 Kuala Lumpur, Malaysia
E-mail: zahirulalam@iiu.edu.my

Abstract - In this work, two adjacent patches and MEMS switches are proposed a reconfigurable antenna that is capable of operating at several frequencies. Optimization of wing patches is done to obtain more than three resonant frequencies of the antenna by selecting the MEMS cantilever bridge materials. The study shows that the material of MEMS switches has influenced the performance of the antenna. SiN as MEMS bridge material makes the antenna to operate at 16.76GHz, 23.56 GHz and 27.7 GHz in the "OFF" and operate at 20.9 GHz and 21.91 GHz in the "ON" states of MEMS switches. A comparative study has done for Alumina, SiN, GaAs and Teflon as MEMS bridge materials. The design is performed by using 3D electromagnetic simulator HFSS considering ideal MEMS switches.

I. Introduction

Reconfigurable multi-band antennas are attractive for many military and commercial applications where it is required to have a single common aperture antenna that can be dynamically reconfigured to transmit (or receive) on multiple frequency bands. Such common-aperture antennas lead to considerable savings in size, weight and cost. They find applications in space-based radar, communication satellites, electronic intelligence, and aircraft navigations besides many other communications and sensing applications. A number of different reconfigurable antennas, such as planar and 3-D have been developed. Some of them are developed for radar applications [1]-[2] and other planar antennas are designed for wireless devices [3]-[4]. Reconfigurable slot antennas are designed for UHF [5]. Reconfigurable patch antennas are also designed to operate in both L and X bands [6].

Radio frequency microelectrical mechanical systems (RF MEMS) is an emerging technology that promises the potential of revolutionizing RF and microwave systems implementation for the next generation of telecommunication applications. Its low-power, excellent RF performance, large tuning range, and integration capability are the key characteristics, enabling system implementation with potential improvements in size, cost, and increased functionality. They are normally built on high-resistivity silicon wafers, gallium arsenide (GaAs) wafers, and quartz substrates using semiconductor microfabrication technology with a typical four-to six-mask level processing. [7]–[11]. Typical example of MEMS based antenna are reported in [12]-[13]. In this paper, we proposed a multi-band reconfigurable antenna using RF MEMS switches that can be fabricated with easy process steps. The effect of material used for MEMS switches is simulated and analyzed under the conditions when both the switches are either "OFF" or "ON".

II. Reconfigurable MEMS Antenna Design

The schematic diagram of the proposed reconfigurable antenna is shown in Fig. 1. It consists of three patches placed on a 18×10 mm^2 Rogers substrate of thickness 0.32mm. The centre patch of dimension 4×3 mm^2 and two side patches defined as wing patches, separated by gap of 0.5mm. The co-axial feeding point is placed at the centre patch. The length of the wing patch is chosen as L_w=5mm and the width is defined as "W_w". Cantilever type MEMS switches are considered for the design, which consist of cantilever bridges and MEMS contacts. The length of bridges is 10 mm long, 1mm wide, 20μm thick

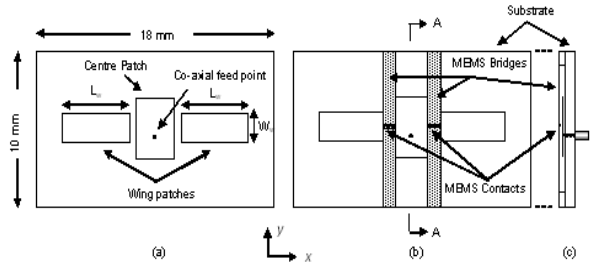


Fig. 1. Schematic diagram of MEMS reconfigurable antenna (a) without MEMS Bridge, (b) with MEMS Bridges, and (c) cross sectional view at A-A.

and placed above the patch gaps. The MEMS are placed on the centre of bridge with 1mm long and 100µm wide contacts. The distance between the patch and the MEMS is 5µm, which the MEMS are considered "OFF". The patch antenna design is supported with a model built using a high frequency structure simulator (HFSS) based on finite elements modeling (FEM). A tool with 3D modeling capabilities is necessary due to the fact that, for small ground planes, the antenna behavior depends on the ground size.

The two critical steps in designing the patch antenna are the definition of the patch dimensions and the feeding configuration. The patch dimensions have direct influence on the operating frequency and on the antenna gain. The difficulty is how to predict accurately the patch dimensions, which is related to the fringing fields together with the small size of the ground plane used.

The antenna feeding should be designed carefully since it must provide a correct impedance matching. At high-signal frequencies it is necessary to design a feeding line with specific characteristic impedance. Also, that line must be connected to a point of the antenna where the input impedance is the same as that of the feed-line characteristic impedance.

A considerable amount of effort and research is done to accurately predict the electrical behavior of RF MEMS switches [12-13]. One of the most compelling reasons for using MEMS in antenna application is that they have shown to approximate, to a very good degree, ideal switches. This allows us to forego the nearly impossible task (at present) of a complete field simulation of an entire antenna structure, including all the minute details of MEMS switches. Therefore, the reconfigurable antenna is designed by performed using a less complex approach with simplified equivalent switches in place of the actual MEMS structure.

A. Antenna Modelling

The centre patch antenna is first designed based on the equations from the Transmission Line Model (TLM) approximation [14], which states that the operating frequency of patch antenna is given by :

$$f_r = \frac{1}{2(L+\Delta L)\sqrt{\varepsilon_{reff}}\sqrt{\mu_o \varepsilon_o}} \tag{1}$$

Where, L is the length of the antenna, ε_o and μ_o are the free space dielectric permittivity and permeability respectively, ε_{eff} is the effective dielectric permittivity given as

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left[1 + 12\frac{h}{W}\right]^{-1/2} \tag{2}$$

Where, ε_r and h are the relative dielectric permittivity and thickness of the substrate and W is the width of the patch. Because of the fringing effects, the antenna looks larger than its physical dimensions. The parameter ΔL in Eqn. 1 takes this effect into account and can be computed by:

$$\Delta L = 0.412h\frac{\left(\varepsilon_{eff} + 0.3\right)\left(W/h + 0.264\right)}{\left(\varepsilon_{eff} - 0.258\right)\left(W/h + 0.8\right)} \tag{3}$$

In order to facilitate its characterization, the antenna is designed to have matching input impedance (50 Ω) as mentioned earlier. The input impedance of the antenna can be adapted by choosing the right position for the feeding point. The impedance is maximum at the patch border and decreases as it moves in from the border, given by:

$$Z_{in} = Z_{\max}\cos^2\left(y\frac{\pi}{L}\right) \tag{4}$$

Where Z_{max} is the impedance at y=0. Using the values given by the TLM approximation, a model for the antenna is built using HFSS. The model is used to trim the antenna dimensions for the desired frequency. The two wing patches are placed on the two sides of the centre patch to obtain reconfigurability.

III. Results and Discussions

The reconfigurable antenna performance depends on the separation between the centre patch and the wing patch, wings and MEMS dimensions as well as the material chosen for the MEMS cantilever bridges. Optimization has been performed by using HFSS simulator. The up and down states of both the MEMS switches are defined as "OFF" and "ON" states. The "ON" states shortened the wing and the centre patches.

A. Effects of MEMS contact Width

The effects of MEMS contact of width on the return loss is shown in Fig. 2, considering Alumina as a bridge material. The figure shows that the antenna performance reduces with the increase in width of the MEMS switches whether the MEMS are in "OFF" or "ON" states. A width of 0.2mm is chosen as the optimum width of MEMS switches considering them in "OFF" and "ON" states. It is also observed that the wide width of the MEMS switches has no effect on variation of the resonance frequency whether the MEMS are either in "OFF" or "ON" states. This indicates that the wide MEMS switches provide continuous path between centre and wing patches.

The antenna resonance frequency and the return loss with different MEMS contact width are shown in Fig. 3 when the MEMS switches are "OFF" and "ON" states. There are three resonant frequencies for the MEMS with width of 0.2mm and less and two resonant frequencies for width larger than 0.2mm when the MEMS contacts are in the "OFF" states. It is also observed that the return loss of the antenna is higher for MEMS width larger than 0.2mm, except for the width of 0.4mm when the MEMS contacts

are in the “OFF” states. When the MEMS contacts are in the “ON” states, there is only single resonant frequency except widths of 0.1mm and 0.9mm. However, low return loss can be obtained when the MEMS contact width is less than 0.4mm. Therefore, width of MEMS switch is chosen at 0.2mm for optimum performance of the antenna as mentioned earlier. It is also mentioned that the centre resonant frequency when the MEMS contacts are in the “OFF” states does not deviate appreciably from the resonant frequency when the MEMS contacts are in the “ON” states.

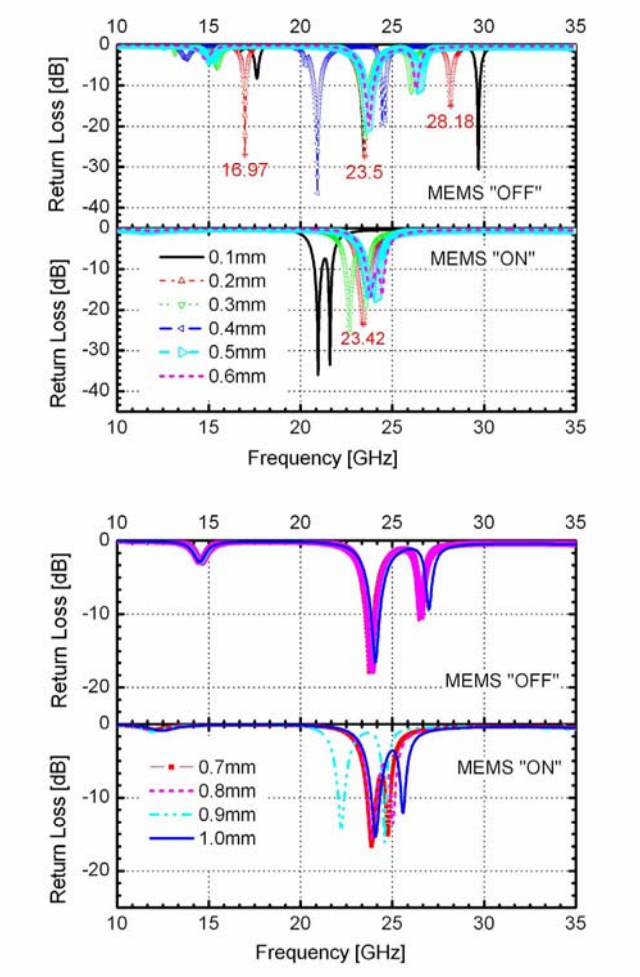


Fig. 2. Return loss for different MEMS contact width for “OFF” and “ON” states.

B. Effects of Wing Patch Width

The simulation has done for varying the wing patch width by maintaining the length L_W of the wing at 5mm and considering Alumina as a MEMS bridge material. The length is varied from 1mm to 4mm in 1mm step, and the results are presented in Fig. 4. The figure shows that the antenna loses its performance when the wing patch width is 4mm, the same width as the centre patch width. It also shows that return losses decrease with the increase in width of the wing patch. The multi-band antenna can be obtained by reducing the width of the wing patches. It is observed that there are three resonant frequencies, namely 16.97 GHZ, 23.5 GHz and -28.18 GHz with the return losses of 26.86 dB, -27.4 dB and -14.9 dB, respectively for a wing width of 2mm, when the MEMS are in the “OFF” states. The single resonant frequency of 23.42 GHz with return loss of -23.47 dB has been observed for the same wing dimensions when the MEMS are in the “ON” states. It is said that the antenna behaves like a single patch antenna when the MEMS are in the “ON” state.

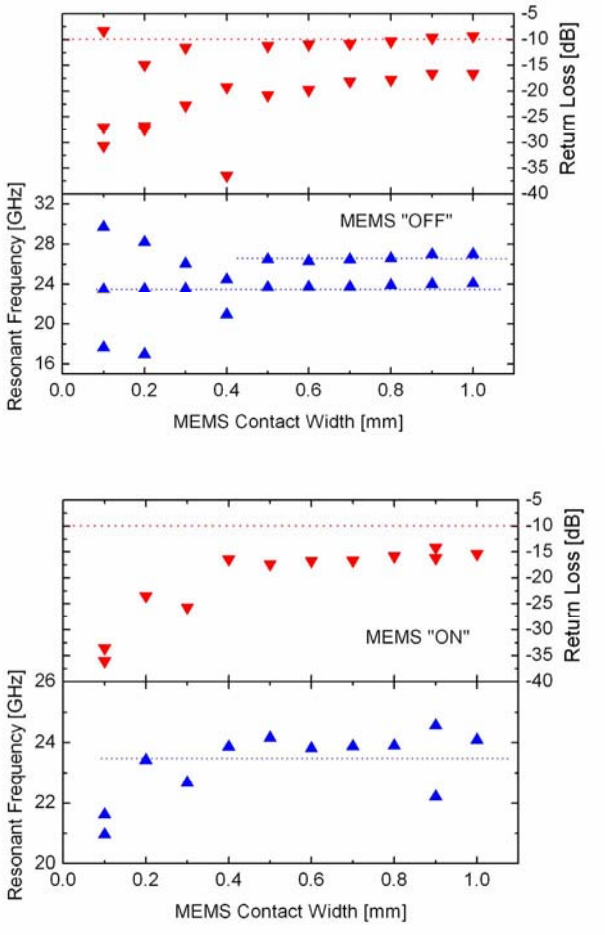


Fig. 3. Return loss and resonant frequency for different MEMS contact width for “OFF” and “ON” states.

C. Effects of MEMS bridge materials

The simulation is done by choosing Alumina, GaAs, SiN and Teflon as a MEMS cantilever materials by varying the wing patch width maintaining the MEMS contact width at 0.1mm. The simulation results of Fig. 3 show that wider the MEMS contact width, larger the values of return losses for Alumina as MEMS cantilever bridge. These phenomena are also applicable to other MEMS bridge materials with wing widths of 3mm and 4mm as shown in Fig. 5.

The effects of MEMS cantilever bridge material on the antenna return losses are shown in Fig. 6 for four different MEMS bridge materials when the MEMS contacts are in the “OFF” and “ON” states, by maintaining optimized wing width. The resonant frequencies and the return losses for different materials are presented in Table 1.

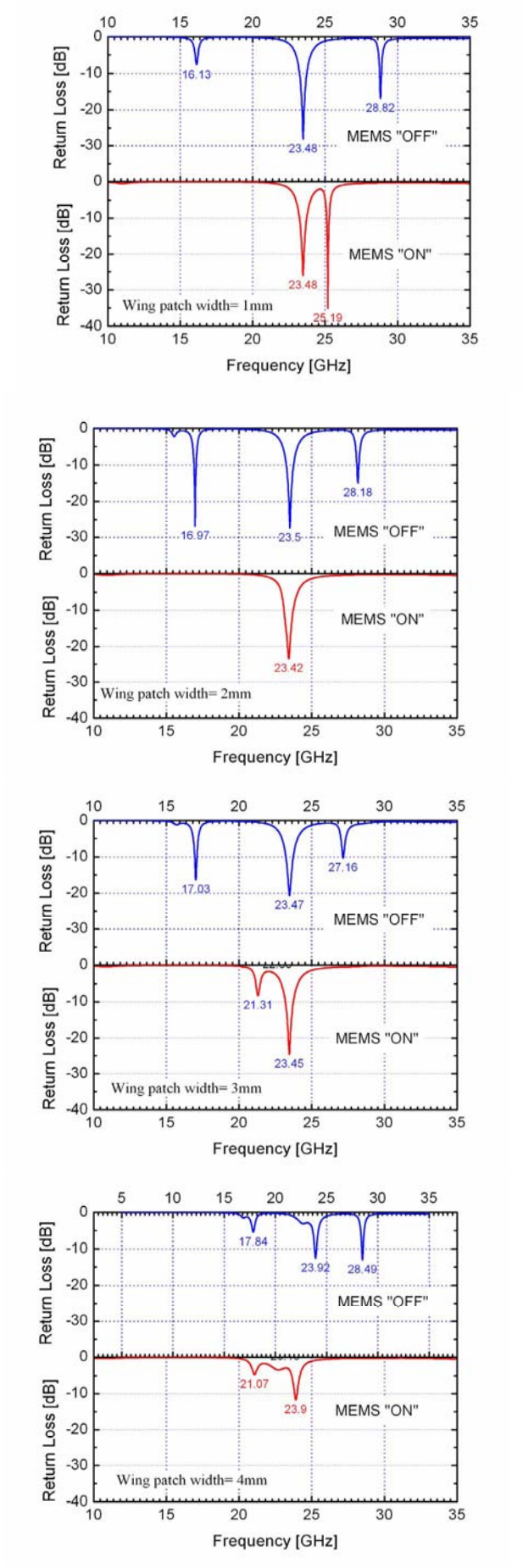


Fig. 4. Return losses for four antennas with different wing patch width.

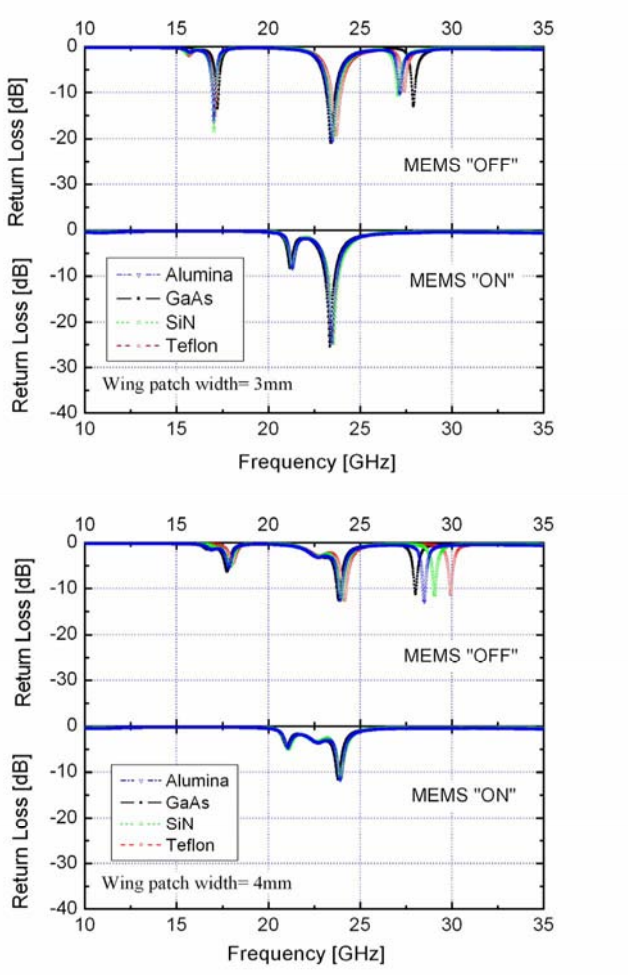


Fig. 5. Return losses for four antennas with different wing patch width for various MEMS cantilever bridge material.

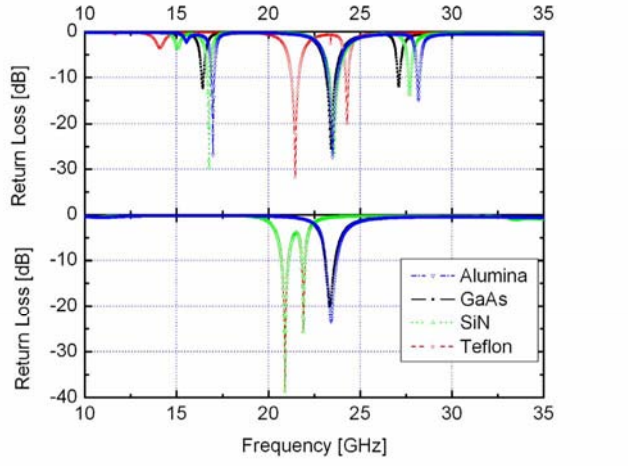


Fig. 6. Return losses for four antennas with different MEMS cantilever bridge material.

It is observed that there are three resonant frequencies for Alumina, GaAs and SiN materials when the MEMS are in the "OFF" states whereas Teflon provides two resonant frequencies. When the MEMS contacts are in the "ON" states, Alumina and GaAs provide single resonant frequency, whereas SiN and Teflon materials provide two resonant frequencies. It is also observed that in "ON" states of MEMS contact has no effect of shifting the

resonant frequencies when the materials are either SiN or Teflon. Therefore, multi-band antenna can be made by choosing Alumina as MEMS cantilever bridge material for operating three different frequencies (16.97 GHz, 23.5/23.42 GHz and 28.18 GHz). By choosing SiN as MEMS cantilever bridge material, it has been observed that the antenna is able to operate at different frequencies (16.97 GHz, 20.9 GHz, 21.91 GHz, 23.56 GHz and 27.7 GHz).

Table 1. Resonant frequencies and return losses for different MEMS cantilever material.

MEMS material (ε_r)	"OFF" state resonant frequency and return loss			"ON" state resonant frequency and return loss	
	f_{R}, S_{11}(dB)	f_{R}, S_{11}(dB)	f_{R}, S_{11}(dB)	f_{R1}, S_{11}(dB)	f_{R1}, S_{11}(dB)
Alumina (9.4)	16.97, -26.86	23.5, -27.4	28.18, -14.9	23.42, -23.47	-
GaAs (12.9)	16.41, -12.24	23.42, -25.45	27.11, -11.91	23.36, -20.06	-
SiN (9)	16.76, -29.56	23.56, -26.31	27.7, -13.69	20.9, -38.69	21.91, -25.72
Teflon (2.1)	21.46, -31.74	24.28, -19.99	-	20.9, -38.69	21.91, -25.72

The simulation has done by increasing the distance between patch and MEMS contact from 5µm to 10µm. The return loss versus the resonant frequency with different MEMS bridge materials is shown in Fig. 7 when the MEMS contacts are in the "OFF" states. It is observed that there are three resonant frequencies for all the materials. The lower, centre and higher resonant frequencies occur almost at the same frequency irrespective of the materials used.

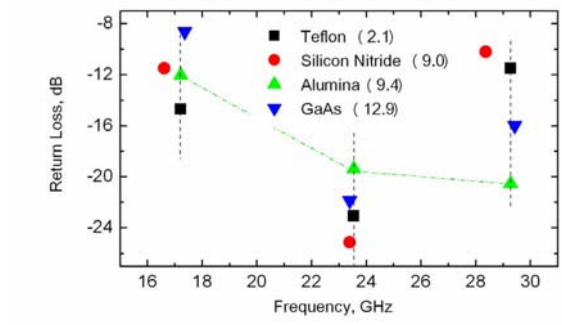


Fig. 7. Return loss of the antenna for different types of MEMS bridge materials when the switches are at "OFF" state.

The return losses and the resonant frequencies of the antenna for different types of MEMS bridge materials are shown in Fig. 8. Here the MEMS contacts are in the "ON" states for 10µm gap between contacts and patches, which shows that there is only single resonant frequency irrespective of the MEMS bridge materials, this is differ from SiN and Teflon materials for a separation of 5µm. It is mentioned that larger separation between MEMS switches and patches need higher turn-on voltage to contact patches.

Therefore, it can be concluded that Alumina, as the MEMS cantilever bridge materials, provides optimum antenna performance for operating at three different frequency bands. However, there is no need of MEMS switch for Alumina materials since there is no appreciable shift of resonant frequency when the MEMS contacts are in the "ON" states or in the "OFF" states for one particular frequency band.. However, SiN as MEMS cantilever bridge material provides five resonant frequencies, three in the "OFF" states and two in the "ON" states. Therefore, reconfigurable antenna can be made by using SiN as MEMS cantilever bridge materials. It is also noted that Teflon material can also be used for reconfigurable antenna to operate at four frequency bands.

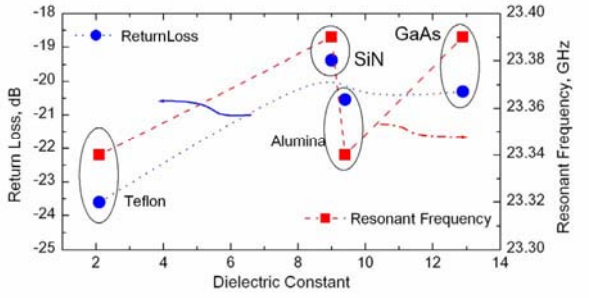


Fig. 8. Return loss of the antenna for different types of MEMS bridge materials when the switches are at "ON" state.

The radiation pattern of the resonant frequencies is shown in Figs. 9 for Alumina when the MEMS contacts are in the "OFF" and "ON" states for 5µm separation.

IV. Conclusions

Multi-band antenna has been designed by placing and optimizing the dimension of two adjacent patches. Additional resonant frequencies have been obtained by incorporating MEMS switches. It is concluded that the MEMS cantilever beam material plays an important role for providing antenna to operate at multi-band frequencies. A triple-band antenna operating at 16.97 GHz, 23.5 GHz and 28.18 GHz is possible by selecting Alumina as MEMS bridge material. A penta-band reconfigurable antenna is possible by constructing MEMS bridge using SiN material. The antenna can be operated at 16.76GHz, 23.56 GHz and 27.7 GHz when the MEMS switches are in the "OFF" states and can be operated at 20.9 GHz and 21.91 GHz when the MEMS switches are in the "ON" states. The proposed reconfigurable antenna can be implemented with easy fabrication process steps by the Sandwich method of fabrication. This work has shown the possibility of using the antenna for multi-band communication applications.

Acknowledgment

This work is funded by the Fundamental Research Grant Scheme by Ministry of Higher Education Malaysia through the Research Centre, International Islamic University Malaysia, Grant No. IIUM/504/RES/G/14/ 3/05/ FRGS 0106-28.

References

[1] K.Tomiyasu, "Conceptual reconfigurable antenna for 35 GHz high-resolution spaceborne synthetic aperture radar," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 39, Issue: 3 pp. 1069 – 1074, 2003.

[2] J.T. Aberle, Sung-Hoon Oh, D.T. Auckland, and S.D. Rogers, "Reconfigurable antennas for wireless devices," *IEEE Antennas and Propagation Magazine*, Vol.45, Issue: 6, pp. 148 – 154. 2003.

[3] G.H. Huff, J. Feng, Shenghui Zhang, G. Cung, and J.T. Bernhard "Directional reconfigurable antennas on laptop computers: Simulation, measurement and evaluation of candidate integration positions," *IEEE Transactions on Antennas and Propagation*, Vol. 52, Issue:12, pp. 3220 – 3227, 2004.

[4] W.H. Weedon, W.J. Payne, and G.M. Rebeiz, "MEMS-switched reconfigurable antennas," *IEEE Antennas and Propagation Society International Symposium*, Vol. 3, pp. 654 – 657, 2001.

[5] D. Peroulis, K. Sarabandi and Linda P.B. Katehi, "Design of reconfigurable slot antennas," *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 2, pp. 645 – 654, 2005.

[6] C. E. Tong, and R. Blundel, "An annular slot antenna on a dielectric half-space," *IEEE Transactions on Antennas and Propagation*, Vol. 2, no.7, pp. 967 – 974, July 1994.

[7] J. J. Yao and M. F. Chang, "A surface micromachined miniature switch for telecommunications applications with signal frequencies from DC up to 4 GHz," *Proc. Transducers*, pp. 384–387, 1995.

[8] D. Hyman, J. Lam, B. Warneke, A. Schmitz, T. Y. Hsu, J. Brown, J.Schaffner, A. Walston, R. Y. Loo, M. Mehregany, and J. Lee, "Surface-micromachined RF MEMS switches on GaAs substrates," *Int. J. RF Microwave Computer-Aided Eng.*, vol. 9, pp. 348–61, 1999.

[9] D. Hah, E. Yoon, and S. Hong, "A low-voltage actuated micromachined microwave switch using torsion springs and leverage," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 2340–2345, 2005.

[10] J. Y. Park, G. H. Kim, K.W. Chung, and J. U. Bu, "Monolithically integrated micromachined RF MEMS capacitive switches," *Sens. Actuators A, Phys.*, vol. A89, no. 1-2, pp. 88–94, 2001.

[11] S. P. Pacheco, L. P. B. Katehi, and C. T. C. Nguyen, "Design of low actuation voltage RF MEMS switch," *MTT-S Int. Microwave Symp.*, vol. 1, pp. 165–168, 2000.

[12] B.A. Cetiner, H. Jafarkhani, J.Y. Qian, H.J. Yoo, A. Grau, and F. DeFlaviis, "Multifunctional reconfigurable MEMS integrated antennas for adaptive MIMOsystems," *IEEE Commun. Mag.*, vol. 42, no.12, pp.62-70, 2004.

[13] G. H. Huff and J. T. Bernhard, "Integration of Packaged RF MEMS Switches With Radiation Pattern Reconfigurable Square Spiral Microstrip Antennas," *IEEE Transactions on Antennas and Propagation*, Vol. 54, Issue: 2, pp. 464 – 469, 2006.

[14] R. Grag, P. Bhartia, I. Bahl and A. Ittipiboon, *Microstrip Antenna Design Handbook*, Artech House, 2001.

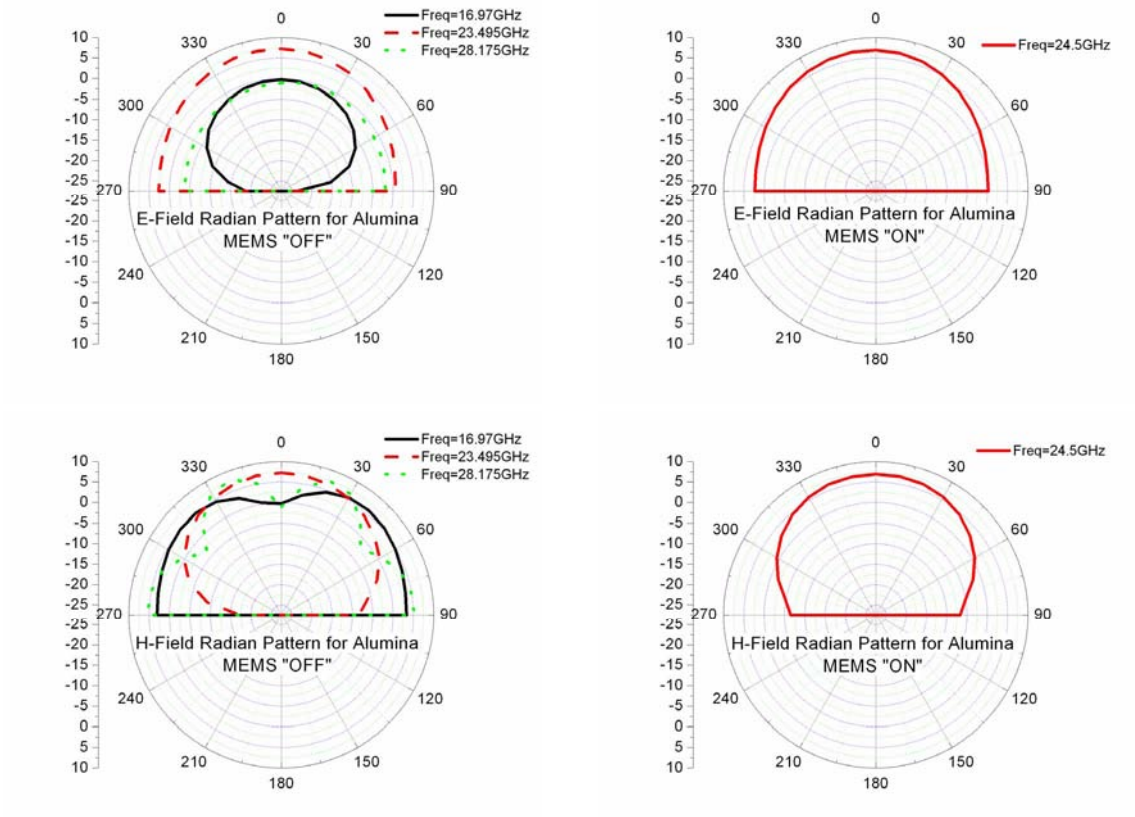


Fig. 9. Radiation pattern for the total gain at E- and H-planes for different resonant frequencies when the MEMS contacts are in "OFF" and "ON" states for Alumina as MEMS materials.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Analysis of Electric Fields Developed Inside Microchannels of Microfluidic Devices

Bashir I. Morshed
Department of Electronics
Carleton University
Ottawa, ON K1S5B6, Canada

Maitham Shams
Department of Electronics
Carleton University
Ottawa, ON K1S5B6, Canada

Tofy Mussivand
Professor of Surgery and Engineering
University of Ottawa Heart Institute
Ottawa, ON K1Y4W7, Canada

***Abstract*—An analysis of electric fields inside microchannels of microfluidic (MF) devices is reported in this paper. Microchannel-based MF devices using electric fields are of recent research interest for various purposes, including on-chip manipulation of biological elements (cell, DNA, proteins and other macromolecules). To determine electric field strengths inside the microchannels of such devices, the uniform electric field equation (eg. $E = \delta V/\delta l$) is commonly used. However, this can lead to a significant estimation error, especially for smaller dimensions of the microchannel, which is the future trend. Finite element method (FEM) analysis should be performed for such smaller dimensions. However, this method is time consuming and computationally expensive, particularly during design phases, as there exist many unbound parameters. In this paper, analytical expressions to determine electric fields and other parameters of interest are developed for a simplistic model, and then compared the FEM simulation results to that of MF devices for a range of microchannel dimensions. The results show that significant estimation errors can occur. For example, more than 10% overestimation of electric field results in microchannel lengths smaller than 2.5 mm. The analysis and graphs can aid MF device designers during design phases.**

I. Introduction

Microfluidic (MF) devices, such as the lab-on-a-chip (LOC), the micro-total-analysis-system (μTAS), the biomedical microelectromechanical systems (bioMEMS) etc., use microchannels, among other micro-structures, to process biological elements like cells, deoxyribonucleic acid (DNA), proteins, etc. [1]. A commonly used microchannel structure in such devices to develop electric fields inside these microchannels constitutes of excitation voltage applied through electrodes inserted inside reservoirs on both ends of the microchannels [2]–[7]. Applications of such electric field-based microchannels are electrophoresis separation of DNA and proteins, electric field-based cell lysis, drug delivery through electroporation, electro-osmotic flow etc.

Although the electric field inside the microchannel can be described using the *Poisson equation* [1], this is rarely used in practice because it is difficult to apply for any specific MF device. To determine electric field strengths inside microchannels, finite element method (FEM) analysis is sometimes used [5], [8]. FEM analysis, however, is not suitable for the design phase, as the design parameters are sometimes unbound. A few attempts were made to develop analytical expressions to calculate electric fields. Exact expressions were developed for certain MF devices, but these are complex and design specific [8]–[10]. The uniform electric field equation (eg. $E = \delta V/\delta l$) is most commonly used to estimate the electric field strengths [2]–[6]. However, this might result in significant estimation errors, especially for smaller dimensions of microchannel - which is the future trend - as demonstrated in this work.

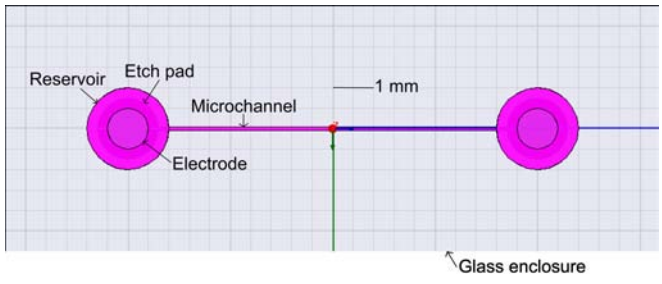


Fig. 1: Top view of a section of the MF device showing a microchannel to be analyzed for developed electric fields.

II. The Microfluidic Device to be Analyzed

The MF device to be analyzed was fabricated using the Protolyne Fabrication Process (Micralyne Inc., Edmonton, Alberta). This foundry allows researchers to obtain the semi-custom MF device fabricated within a short amount of time while being relatively inexpensive. The device consists of two glass-slides, fused together. Eight predefined access holes (i.e. reservoirs) of 2 mm diameter are through-drilled in the top slide. Eight etch-pads of 1.5 mm diameter are etched on the top surface of the bottom slide aligned with the reservoirs. To form microchannels, trenches are etched on the top surface of the bottom slide according to the designer's requirements. These trenches form microchannels when both glass slides are fused together. The length of a microchannel is defined by the intersections of a trench (forming microchannel) and the reservoirs at the ends, while the etch depth is set to 20 μm. Most microfluidic devices share the same topology as this MF device to be analyzed [2], [3], [5], [6].

Fig. 1 shows a section of the design of a microchannel of the MF device joining two access holes (reservoirs). During experiments to develop an electric field inside the microchannel, electrodes are inserted inside these reservoirs. The electric field inside the microchannel is developed by filling

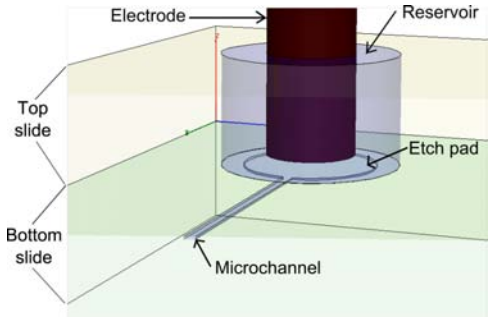


Fig. 2: A 3-D view of a portion of the microchannel of the MF device is shown schematically. The top and bottom glass-slides that compose the glass enclosure are indicated.

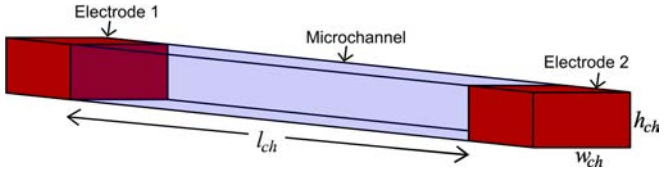


Fig. 3: A simplistic model of a rectangular box shaped microchannel filled with buffer fluid having electrodes on both sides. The whole structure is enclosed within a glass substrate.

the microchannel and the reservoirs with an ionic buffer fluid, such as Dulbeccos Phosphate Buffered Saline (D-PBS), and exciting the electrodes with an applied voltage. This setup is depicted in Fig. 2 using a 3-D view drawing of a portion of the microchannel.

III. Simplistic Model of the Microchannel

To begin electric field analysis of the MF device, a simplistic model of a microchannel structure is considered as shown in Fig. 3. In this simplistic model, the microchannel is a rectangular box with length, width and height of l_{ch}, w_{ch} and h_{ch}, respectively. Both electrodes contact the complete cross-sectional areas on both ends of the microchannel. The microchannel is assumed to be filled with D-PBS buffer fluid, with relative permittivity (ϵ_r) and conductivity (σ_{ch}) of 80 and 1.6 Ω/m, respectively [5]. The whole structure is enclosed within a glass substrate, whose conductivity and permittivity are negligible when compared to those of D-PBS. An electric field, E_{ch}, is developed inside the microchannel as an excitation voltage, V_{app}, is applied across the electrodes. The voltage across the microchannel, V_{ch}, is the same as V_{app} for this simplistic model.

IV. Analytical Expressions for the Simplistic Model

For the simplistic model of the microchannel (Fig. 3), electric flux inside the microchannel can be assumed to be uniform. Thus, E_{ch} can be expressed using the uniform electric field expression between two parallel electrodes [11], as

$$E_{ch} = \frac{V_{ch}}{l_{ch}}. \tag{1}$$

Due to the current flow through the buffer fluid, the (ohmic) power dissipation, P_d, inside the microchannel can be expressed as [12],

$$P_d = \frac{V_{ch}^2}{R_{ch}} \tag{2}$$

where R_{ch} is the electrical resistance of the microchannel. Using the resistivity law [12], R_{ch} for the rectangular box can be expressed as

$$R_{ch} = \rho_{ch} l_{ch} / A_{ch} \tag{3}$$

where ρ_{ch} ($= 1/\sigma_{ch}$) is the resistivity of the buffer fluid inside the microchannel and A_{ch} ($= w_{ch} \times h_{ch}$) is the cross-sectional area of the microchannel. So, from expression (2),

$$P_d = \frac{V_{ch}^2}{\rho_{ch}} (\frac{A_{ch}}{l_{ch}}) = \frac{E_{ch}^2}{\rho_{ch}} \mathbf{V}_{ch} \tag{4}$$

where $\mathbf{V}_{ch}$ ($= A_{ch} \times l_{ch}$) is the volume of the microchannel. Hence, for a given electric field, the power dissipation is proportional to the volume of the microchannel.

Again, the energy density, u_{ch}, stored inside the microchannel due to the electric field can be expressed by [11],

$$u_{ch} = \frac{1}{2} \epsilon E_{ch}^2 = \frac{\epsilon}{2} \frac{V_{ch}^2}{l_{ch}^2}. \tag{5}$$

Here ϵ ($= \epsilon_0 \epsilon_r$) is the permittivity of the buffer fluid, where ϵ_0 is the permittivity of the free space and ϵ_r is the relative permittivity of the buffer fluid. The total energy stored, U_{ch}, within the microchannel can be obtained by integrating u_{ch} over $\mathbf{V}_{ch}$ [11]. For the rectangular box, U_{ch} can be expressed by the simple expression,

$$U_{ch} = \int_{\mathbf{V}_{ch}} u_{ch} d\mathbf{V} = \frac{1}{2} \epsilon E_{ch}^2 \mathbf{V}_{ch} = \frac{\epsilon}{2} V_{ch}^2 (\frac{A_{ch}}{l_{ch}}). \tag{6}$$

Again, for a given electric field, the total energy stored is proportional to the volume of the microchannel. From expressions (6) and (4), U_{ch} is proportional to P_d with a proportionality constant of $\epsilon \rho_{ch}/2$.

These analytical expressions (1,4-6) for the simplistic model are plotted in Fig. 4 to 7 (as solid lines) for a range of V_{app}. In Fig. 4 and 6, plots of electric field strengths and energy densities are plotted against the lengths of the microchannels (l_{ch}) using expressions (1) and (5), respectively. Fig. 5 represents power dissipation curves against the area-over-length ratio (A_{ch}/l_{ch}) using expression (4). Fig. 7 shows the relationship of the total energy stored inside the microchannel against A_{ch}/l_{ch} using expression (6).

V. FEM Simulation of the Simplistic Model

The simplistic microchannel model (Fig. 3) was simulated by using an FEM analysis tool (Maxwell3D simulator from Ansoft Corp). A script file was written to simulate 175 different sets of dimensions (l_{ch}, w_{ch} and h_{ch}). The range of l_{ch}, w_{ch} and h_{ch} simulated were 100 to 10000 μm, 1 to 1000 μm, and 10 to 1000 μm, respectively. Each structure was simulated for excitation voltages ranging from 1 V to 1000 V. The material property for the buffer fluid inside

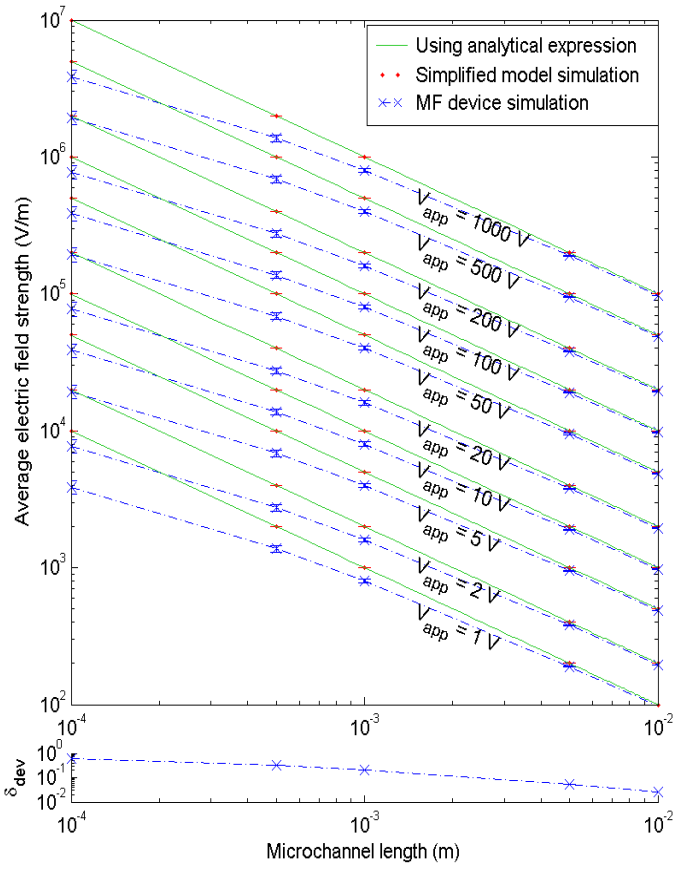


Fig. 4: Plot representing relationship between the developed electric field strengths (E_{ch}) with the lengths of the microchannels (l_{ch}) for various excitation voltages (V_{app}). Normalized deviations (δ_{dev}) are plotted at the bottom. All simulation results include error bars representing standard errors.

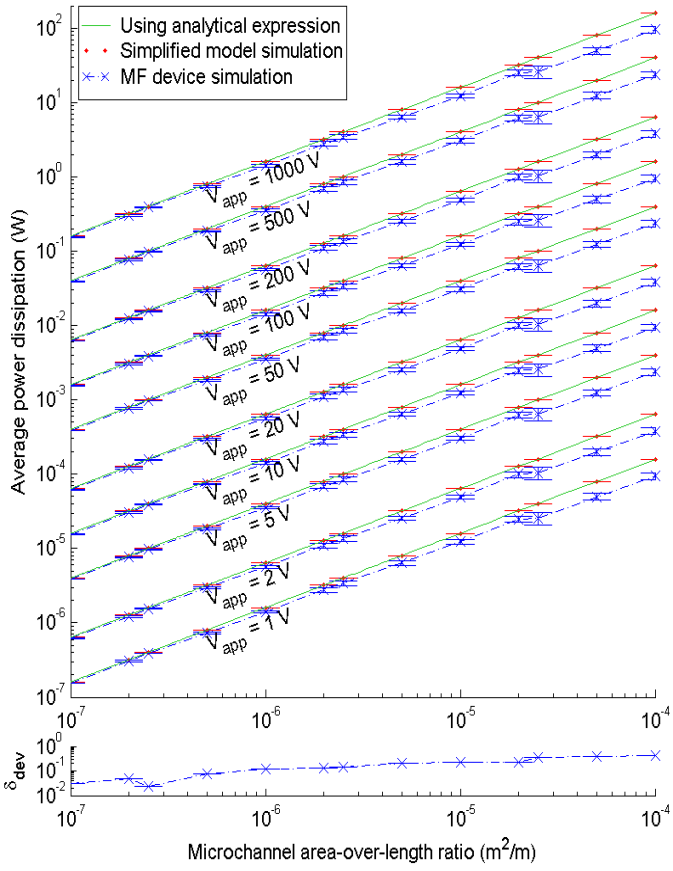


Fig. 5: Plot depicting power dissipations (P_d) for various area-over-length ratios (A_{ch}/l_{ch}) of the microchannels for various excitation voltages (V_{app}). Here, ρ_{ch} = 0.625 Ω-m. Normalized deviations (δ_{dev}) are plotted at the bottom. All simulation results include error bars representing standard errors.

the microchannel was set similar to D-PBS ($\epsilon_r = 80$ and $\sigma_{ch} = 1.6\Omega$/m). The number of iterations was set to 20, as higher iterations provide insignificant change. The resultant data were collected, analyzed, and plotted using Matlab in Fig. 4 to 7. The mean of these data are represented by dots (.), and the standard errors are denoted by the error-bars on both sides of the mean.

These simulation results very closely relate to the analytical expressions. The maximum values of the normalized standard errors are tabulated in Table I. The normalized standard errors are calculated by dividing the standard errors by the corresponding means. The very small values of these normalized standard errors indicate good agreement of the analytical expressions with the FEM simulations.

VI. FEM Simulation of the Microfluidic Device

The MF device model (Fig. 1 and 2) was simulated using another script file with the same FEM analysis tool. Dimensions of the microchannel were varied (exact same ranges as the simplistic model), while the dimensions of the reservoirs, electrodes, and etch holes were kept constant. The electrodes of a 1 mm diameter were positioned in the centers of the reservoirs, 0.1 mm above the bottom plate. The microchannel and reservoirs were assumed to be filled with the same material having properties similar to D-PBS, and the same numbers of iterations were performed. The resultant data were collected, analyzed, and plotted using Matlab and are shown in Fig. 4 to 7. The mean of the data are denoted by crosses ($\times$), and the standard errors are denoted by the error-bars on both sides of the means. The mean values are connected with dot-dashed lines.

The maximum values of the normalized standard errors for this MF device (Table I) are higher compared to those of the simplistic model. These maximum values of the normalized standard errors were below 0.3. This indicates that the y-axis parameters are influenced by additional parameters other than the x-axis parameters. For example, E_{ch} of the MF device is dependent on other parameters (such as w_{ch}, h_{ch} etc.) in addition to l_{ch}.

The mean data of the MF device simulation results deviate from those of the simplistic model, especially for small l_{ch} (i.e. large A_{ch}/l_{ch}). This type of deviation should be taken under consideration for electric field related applications, especially

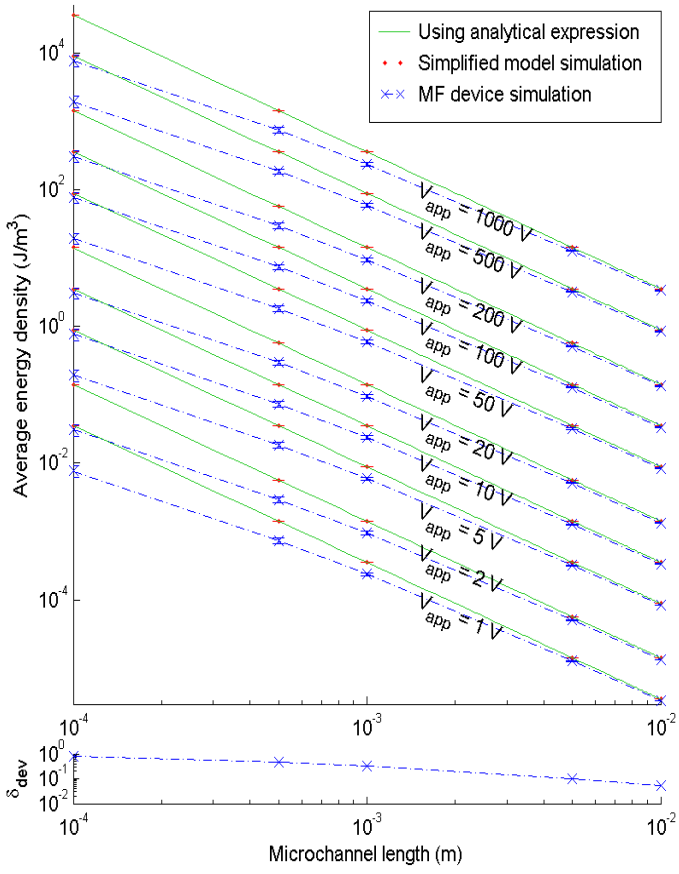


Fig. 6: Plot showing dependance of the energy densities (u_{ch}) inside the microchannels with the microchannel lengths (l_{ch}) for various excitation voltages (V_{app}). Here, ϵ_r = 80. Normalized deviations (δ_{dev}) are shown at the bottom. All simulation results include error bars representing standard errors.

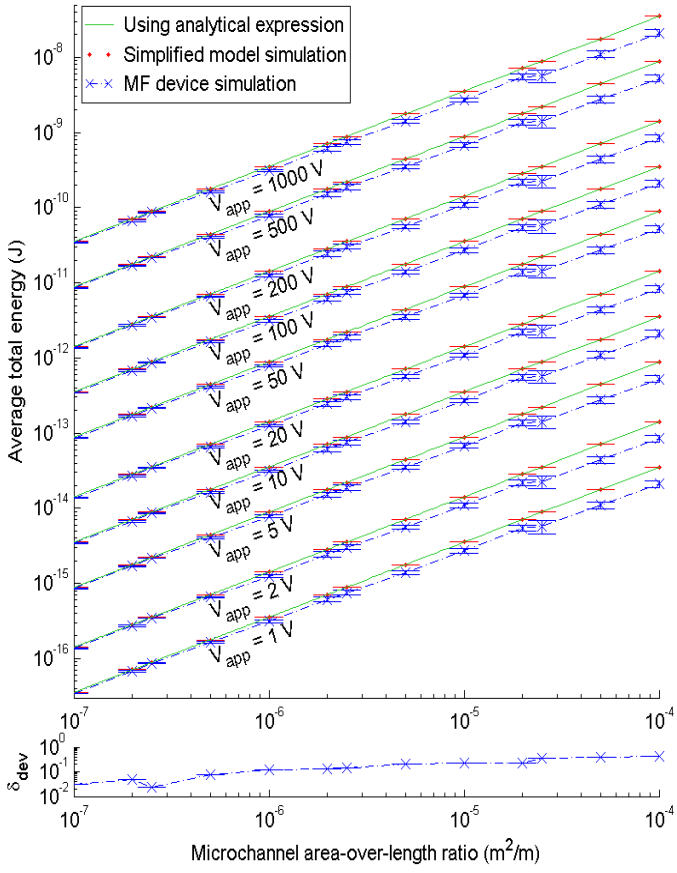


Fig. 7: Plot representing the total energies (U_{ch}) stored versus area-over-length ratios (A_{ch}/l_{ch}) of the microchannels for various excitation voltages (V_{app}). Here, ϵ_r = 80. Normalized deviations (δ_{dev}) are given at the bottom. All simulation results include error bars representing standard errors.

TABLE I: Maximum values of the normalized standard errors from FEM simulations

Plot	Simplistic model	MF device
E_{ch} vs l_{ch}	0	0.1145
P_d vs A_{ch}/l_{ch}	9.8×10^{-17}	0.2997
u_{ch} vs l_{ch}	1.7×10^{-7}	0.1905
U_{ch} vs A_{ch}/l_{ch}	5.65×10^{-7}	0.2997

TABLE II: Overestimation ranges of the analytical expressions when compared to the MF device simulation results

Parameter	10% overestimation	50% overestimation
E_{ch}	$l_{ch} < 2.5$ mm	$l_{ch} < 0.1$ mm
P_d	$\frac{A_{ch}}{l_{ch}} > 7\times10^{-7}$ m^2/m	$\frac{A_{ch}}{l_{ch}} > 1\times10^{-4}$ m^2/m
u_{ch}	$l_{ch} < 5$ mm	$l_{ch} < 0.4$ mm
U_{ch}	$\frac{A_{ch}}{l_{ch}} > 7\times10^{-7}$ m^2/m	$\frac{A_{ch}}{l_{ch}} > 1\times10^{-4}$ m^2/m

for the case of small lengths of microchannels [3]–[5]. To quantify these deviations, the normalized deviations (δ_{dev}) are plotted beneath each plots (Fig. 4 to 7). Here, δ_{dev} are calculated by subtracting the mean data for the MF device from the mean data for the simplistic model, then dividing by the mean data for the simplistic model. These plots are averaged for various V_{app}. The mean values of these averages are shown using crosses ($\times$), while the standard deviations are shown as the error bars in these δ_{dev} plots.

The values of δ_{dev} are always positive, indicating the analytical expressions always overestimate, due to the fact that V_{ch} is always smaller than V_{app} in MF devices. As microchannel lengths become smaller (i.e. area-over-length ratios become greater), δ_{dev} values increase. The amount of overestimation can be determined from the δ_{dev} plots. The range for 10% and 50% overestimation are tabulated in Table II. One can rationalize that these estimation errors resulted from the presence of the reservoirs in the MF device. These estimation errors can be minimized by determining V_{ch} for the MF device. To calculate V_{ch}, one has to determine the resistance introduced by the reservoirs, which is tedious, and dependent on the structure and dimensions of the reservoirs.

VII. Conclusions

To analyze the electric fields developed inside the microchannels of MF devices, analytical expressions were developed using a simplistic model. FEM simulations of this simplistic model and the MF device revealed that significant errors are introduced if the analytical expressions for the simplistic model are employed for microchannels of MF devices with very small dimensions. This analysis and the resulting graphs can aid in designing microchannels for developing electric fields, especially during the design phases. Further analysis is being conducted to develop empirical expressions for MF devices by extending analytical expressions for the simplistic model to calculate various parameters of the MF device to achieve a higher degree of accuracy.

Acknowledgement

The authors gratefully acknowledge the contributions of Canadian Microelectronics Corporation (CMC), Natural Sciences and Engineering Research Council (NSERC) of Canada, and Ottawa Heart Institute (OHI).

References

[1] Steven S. Saliterman, *Fundamentals of BioMEMS and Medical Microdevices*, Wiley-Interscience, WA, USA, 2006.

[2] J. Gao, X. Yin, and Z. Fang, "Integration of Single Cell Injection, Cell Lysis, Separation and Detection of Intercellular Constituents on a Microfluidic Chip," *Lab Chip*, vol. 4, pp. 47–52, 2004.

[3] H. Wang, A. K. Bhunia, and C. Lu, "A Microfluidic Flow-through Device for High Throughput Electrical Lysis of Bacterial Cells Basaed on Continuous DC Voltage," *Biosensors and Bioelectronics*, vol. 22, pp. 582–588, 2006.

[4] A. S. Bhagat, S. Dasgupta, R. K. Banerjee, and I. Papautsky, "Effects of Microchannel Cross-section and Applied Electric Field on Electroosmotic Mobility," in *Conf Solid-State Sensors, Actuators and Microsystems*, 2007, pp. 1853–1856.

[5] Dong Woo Lee and Young-Ho Cho, "A Continuous Electrical Cell Lysis Device Using a Low DC Voltage for a Cell Transport and Rupture," *Sensors and Actuators*, vol. B 124, pp. 84–89, 2007.

[6] L. A. Legendre, J. M. Bienvenue, M. G. Roper, J. P. Ferrance, and J. P. Landers, "A Simple, Valveless Microfluidic Sample Preparation Device for Extraction and Amplification of DNA from Nanoliter-volume Samples," *Analytical Chemistry*, vol. 78, no. 5, pp. 1444–1451, 2006.

[7] J. W. Hong, H. Hagiwara, T. Fujii, H. Machida, M. Inoue, M. Seki, and I. Endo, "Separation and Collection of a Specified DNA Fragment by Chip-based CE System," in *Micro Total Analysis Systems*, 2001, pp. 113–114.

[8] A. Jenkins, C. P. Chen, S. Spearing, L. A. Monaco, A. Steele, and G. Flores, "Design and Modelling of a Microfluidic Electro-lysis Device With Controlling Plates," in *Int. MEMS Conf.*, 2006, pp. 620–625.

[9] P. Linderholm, U. Seger, and P. Renaud, "Analytical Expression for Electrical Field Between Two Facing Strip Electrodes in Microchannel," *Electronic Letters*, vol. 42, no. 3, pp. 145–146, 2006.

[10] A. N. Chatterjee and N. R. Aluru, "Combined Circuit/Device Modeling and Simulation of Integrated Microfluidic System," *IEEE J MEMS*, vol. 14, no. 1, pp. 81–95, 2005.

[11] Fawwaz T. Ulaby, *Fundamentals of Applied Electromagnetics*, Pearson Prentice Hall, NJ, USA, 2004.

[12] J. W. Nilsson and S. A. Riedel, *Electric Circuits*, Prentice-Hall, NJ, USA, 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Electrical Characteristics of Coaxial Nanowire FETs Based on Analytical Approach

Alireza Kargar [a], *Student Member*, *IEEE*, Alireza Rezvanian [b]

[a] Department of Electrical Engineering, Shiraz University, Shiraz, Iran
[b] Department of Electrical and Computer Engineering, Azad University of Qazvin, Iran
E-mail [a]: arkargar@ieee.org

Abstract - In this paper, an analytical approach based on ballistic current transport is presented to investigate the electrical characteristics of the coaxial nanowire field effect transistor (CNW FET). The potential distribution along the nanowire is derived analytically by applying Laplace equation. In addition to assumption of ballistic transport, tunneling process and quantum state of energy are implemented to determine the amount of electron transport along the nanowire from the source to the drain terminals. To consider the tunneling phenomena, WKB approximation is used and the transmission coefficients on both sides of the channel are obtained separately. In ballistic regime, an expression for channel current in terms of the bias voltages and Schottky barrier height is derived.

I. Introduction

Scaling limits of conventional silicon micro transistors, grow the interest in nano transistors with one-dimensional channels, such as carbon nanotube transistors [1,2] and silicon nanowire transistors [3]. Application of one-dimensional analysis makes the electrostatics of nanowire devices quite different from that of the bulk silicon devices. Shrinkage of the channel length of the transistor into very short scales such as sub-50-nm has made the conventional bulk MOSFET devices unreliable. This deficiency is due to inappropriate ratio of channel length to channel cross-section which magnify the short-channel effects such as high leakage current and poor gate control [4].

In order to improve the short channel effects various attempts are made to scale down the channel cross-section and length simultaneously. Therefore, different device structures such as the double-gated FinFET [5], tri-gated [6], Π-gated [7], Ω-gated [8], nanowire body [9], and gate all around [10], have been extensively studied to restrict the short-channel effects in nanowire FETs. Among these devices gate all around (more focused on coaxial structures) nanowire FETs have recently drawn wide research attentions due to their excellent short-channel effect, and the superior current–voltage characteristics [11]. To understand the device physics and characterize the carrier transport in nanowire transistors and to assess their scaling limits, the theoretical analysis and reliable simulation are important. Recently, using both tight–binding–relations and parabolic energy bands, the characteristics of n–type SNWTs are evaluated by a semi-numerical ballistic FET model [12]. And also using three-dimensional quantum mechanical simulations of SNW FETs have been accomplished based on the parabolic effective-mass approximation [13]. However, in recent years, more attention has been drawn to the analytical and numerical models of carrier transport which are far-from-equilibrium. Analytical modeling is expected to be important in architectures of such a device, but for transport in far-from equilibrium cases no treatment exists to describe the transition from drift-diffusion to ballistic regime. Because of heavy mathematics which is needed to achieve an analytic expression for characterizing the electrical behavior of coaxial nanowire FETs, this approach has not been greatly considered so far.

In this paper, we have proposed an analytical approach to investigate the electrical characteristics of coaxial nanowire FETs in ballistic regime. To achieve a formula for channel current, the potential distribution within the nanowire is derived first. Then using WKB approximation for tunneling problem, the transmission coefficients for two terminal contacts are achieved. It is shown that the transmission probability changes with bias voltages and Schottky barrier height. Finally, by assumption of ballistic conduction, the electron current passing through the nanowire channel is derived, where the derived equation includes three relations of operation. Comparing the results of this work with those of the other works, verifies that the proposed approach can apply to predict the electrical characteristics of CNW FETs.

II. Analytical Approach

In recent years, most of the researches on nanoelectronic devices are greatly focused on theoretical investigation of planar structure [17]. Avoiding heavy numerical calculations, forces the designer to seek analytical solutions to carrier transport equations for device characteristics in terms of applied terminal voltages and Schottky barrier height (SBH). In this work,

based on an analytical approach and including the ballistic transport of carriers we tried to achieve an expression for performance of CNW FETs. Fig. 1 shows the schematic view of CNW FET, where carriers are electrons. To obtain the device current, the energy band diagram must be achieved first. For a barrier in exponential form like the Schottky barrier of a coaxial structure, the Schrödinger equation does not give an analytical solution. Therefore, to obtain the potential function within the nanowire, the Laplace equation is applied. In the coaxial gate structure, the minima of the conduction band edges at both terminals are achieved as

$$U_1(z) = qV_{GS}\left(\exp(-\frac{2z}{t_{OX}}) - 1\right) + SBHS \qquad (1)$$

$$U_2(z) = qV_{GS}\left(\exp(-\frac{2(L-z)}{t_{OX}}) - 1\right) + (SBHD - qV_{DS}) \qquad (2)$$

where $U_1(z)$ and $U_2(z)$ are the potential (energy distribution) at the source and the drain contacts respectively, V_{GS} and V_{DS} are the bias voltages, t_{OX} is the gate oxide thickness, *SBHS* and *SBHD* are the Schottky barrier heights at the source and drain contacts respectively, and L is the length of nanowire. Depending on the transmission coefficient and the terminal bias voltages, the carriers are injected from the contacts into the wire or vice versa. Under non-equilibrium condition, the existing electrons in the source reservoir tends to surmount the Schottky barrier (SB) between the nanowire and the metal- source contact that is denoted by $U_1(z)$. The Wentzel-Kramers-Brillouin (WKB) approximation is often suited the tunneling process in such problems. Application of this approach, the transmission coefficients at both source and drain terminals are derived, respectively as

$$T_1(E) = \exp(\frac{\sqrt{8m}}{\hbar}(\arctan(\frac{\sqrt{SBHS - E}}{\sqrt{V_{GS} - (SBHS - E)}}) \times (\sqrt{V_{GS} - (SBHS - E)}) - \sqrt{SBHS - E})) \qquad (3)$$

$$T_2(E) = \exp(\frac{\sqrt{8m}}{\hbar}(\arctan(\frac{\sqrt{(SBHD - V_{DS}) - E}}{\sqrt{V_{GS} - (SBHD - V_{DS} - E)}}) \times (\sqrt{V_{GS} - (SBHD - V_{DS} - E)}) - \sqrt{(SBHD - V_{DS}) - E})) \qquad (4)$$

where m is the carrier mass, $\hbar$ is the Planck's constant, and E is the total energy. As (3) shows the transmission probability from the source to the wire ($T_1(E)$) is a function of the gate-source voltage and Schottky barrier height at the source (*SBHS*). Similarly, the transmission coefficient from the wire to the drain ($T_2(E)$), is a function of three variables, including the gate-source voltage, the Schottky barrier height at the drain (*SBHD*), and the drain-source voltage. When the gate-source voltage becomes smaller than the drain-source voltage, it can be well approximated that only the transmission coefficient from the source is enough to calculate the channel current. When the gate-source voltage becomes larger than the drain-source voltage, we have to consider both

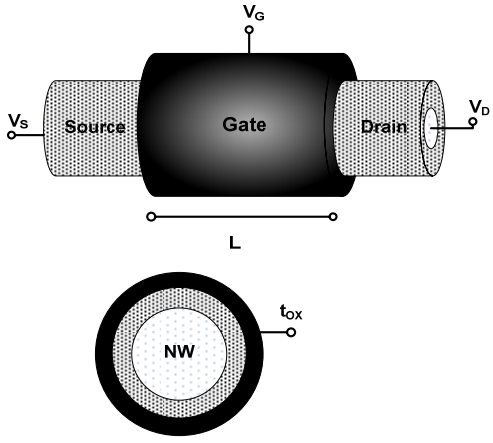


Fig. 1. Schematic view of a coaxial nanowire FET (CNWFET).

Schottky barriers at the source and the drain ends. In this case, the effective transmission coefficient, $T^*(E)$, is given by

$$T^*(E) = \frac{T_1(E)T_2(E)}{T_1(E) + T_2(E) - T_1(E)T_2(E)} \qquad (5)$$

where $T_1(E)$ and $T_2(E)$ are given in (3) and (4). At this point the current through the structure is calculated by means of the Landauer-Buttiker formula as

$$I = \frac{4q}{h}\int_{-\infty}^{+\infty} T(E)(f_S(E) - f_D(E))dE \qquad (6)$$

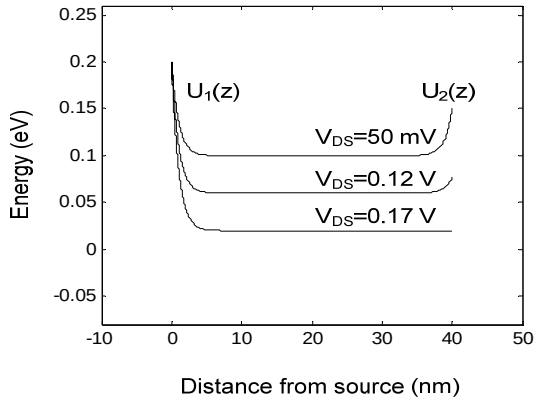
where $f_S(E)$ anzd $f_D(E)$ are the Fermi-Dirac distribution functions in source and drain regions, respectively. Now, by considering a one-dimensional ballistic transport between terminal contacts, we replace the solutions of $T_1(E)$ and $T^*(E)$ from (3) and (5) into (6). Consequently, the results at different biases are given, respectively as

$$I = \frac{2e}{\pi\hbar}(1 - \exp(\frac{-eV_{DS}}{kT})) \times \begin{cases} \int_{SBH - V_{GS}}^{SB} T_1(E)\exp(\frac{-E}{kT})dE \ , \ 0 < V_{GS} \le V_{DS} & (7\text{-a}) \\ \int_{SBH - V_{GS}}^{SBH - V_{DS}} T^*(E)\exp(\frac{-E}{kT})dE \ , \ V_{DS} < V_{GS} \le SBH & (7\text{-b}) \\ \int_{0}^{SBH - V_{DS}} T^*(E)\exp(\frac{-E}{kT})dE \ , \ V_{GS} \ge SBH & (7\text{-c}) \end{cases}$$

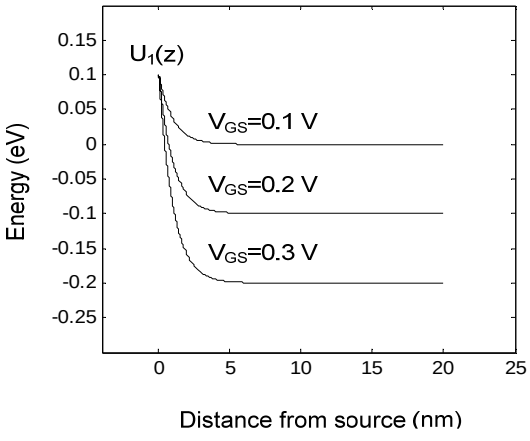
where $\hbar$ is the Plank's constant, k is the Boltzmann's constant and $T(E)$ is the transmission probability, In deriving (7) it's assumed *SBHS* = *SBHD* = *SBH*.

III. Results and Discussions

Equation (7) which is derived in the previous section can be used to demonstrate the electrical characteristics of coaxial nanowire FETs (CNWFETs). It makes possible to simulate CNW FET with different contact materials.



(a)



(b)

Fig. 2. The energy band diagram of CNWFET versus distance from source, (a) at different drain-source voltages, and (b) at different gate-source voltages.

Fig. 2(a) shows the energy band diagram within the nanowire in ballistic regime at various drain voltages, where the gate-source voltage is equal to 0.1 V and the gate oxide thickness is equal to 2 nm. Fig. 2(b) shows the energy band diagram near the source at various gate voltages and the *SBHS* of 0.1 eV.

As the energy band diagrams of Fig. 2 show, when the drain-source voltage increases, the Schottky barrier height at the drain (*SBHD*) reduces. At an appropriate drain voltage (approximately equal to *SBHD*), the barrier at the drain is almost removed and only the barrier at the source remains. In the case of $V_{DS} < SBHD$, we have two Schottky barriers at both ends of the nanowire. Therefore, in this case, for calculating the transmission probability we must apply (5). If V_{DS} changes in the range of *SBHD* by an appropriate approximation we can apply $T_1(E)$ instead of $T^*(E)$. Fig. 3 demonstrates the potential function near the source contact for different gate oxide thicknesses. In this figure the gate source bias voltage is 0.1V and Schottky barrier height (*SBHS*) is equal to 0.2 eV. As Fig. 3 shows, changing the gate oxide thickness affect the electrical potential (energy) within the nanowire so that it changes the channel current significantly. If the gate oxide thickness is in the range of the channel length, the source or drain field penetrates into the channel and

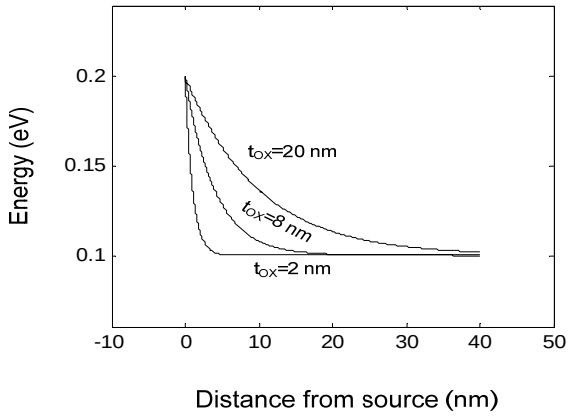


Fig. 3. The energy distribution of CNWFET near the source at different gate oxide thickness.

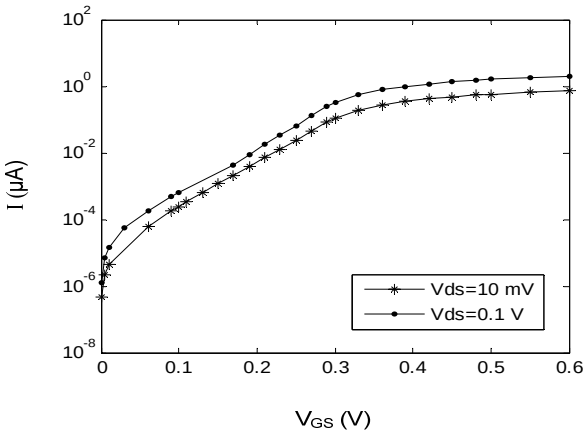


Fig. 4. The drain current versus the gate-source voltage at different drain bias voltages.

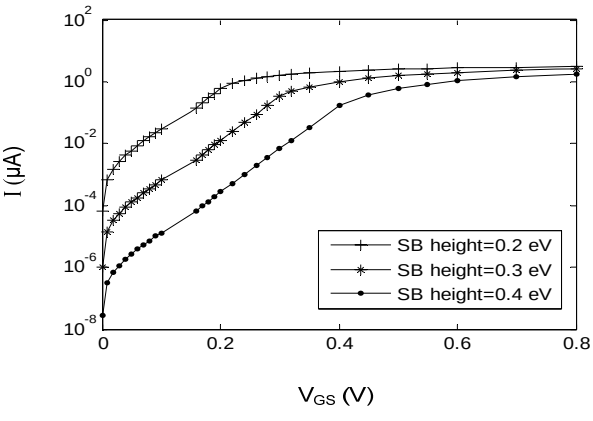


Fig. 5. The drain current versus the gate-source voltage at different Schottky barrier heights (SBH).

the transistor cannot be turned off. When the gate oxide is thin, however, the gate still has very good control over the channel current and the transistor is well turned off. Now, we apply (7) to obtain the transfer characteristic of the device. According to (7) there are three limits of changing the gate-source voltage at each drain-source voltage. Thus, for the case of $V_{GS} < V_{DS}$ the (7-a) is used. For the case of $V_{DS} < V_{GS} < SBH$ the band diagram contains

Schottky barriers at the source and drain, (7-b) gives the device current. If $V_{GS} > SBH$, (7-c) is used to determine the current. Fig. 4 shows the variations of the drain current versus the gate-source voltage at different drain voltages. Fig. 5 shows the drain current versus the gate-source voltage for different *SBH*s, where the gate oxide thickness is equal to 2 nm and the drain-source voltage is equal to 0.1 V. As the curves in Fig. 5 demonstrate, an increase in the Schottky barrier height (*SBH*) decreases the channel current and the ratio I_{on}/I_{off} is changed consequently.

IV. Conclusion

Lack of a fully analytical solution to the I-V relationship of the nanowire FETs forced the designers to apply numerical simulation to analyze these devices. However, numerical simulation of these devices encounters a heavy numerical calculation and time. Thus, deriving analytical expressions for I-V behavior of nanowire FETs has been a challenging issue in recent years. In this paper, based on WKB approximation and ballistic transport, we presented an analytical expression for current in CNWFETs. In order to derive analytical expression for I-V characteristic, tunneling process and quantum state of energy are implemented in analytical forms.

References

[1] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, "Ballistic Carbon Nanotube Field-Effect Transistors," *Nature*, vol. 7, pp. 424-654, 2003.

[2] S. Heinze , J. Tersoff , R. Martel , V. Derycke, J. Appenzeller , and Ph. Avouris, "Carbon Nanotubes as Schottky Barrier Transistors," *Phys. Rev. Lett.,* vol. 89, id. 106801, 2002.

[3] Y. Cui, Z. Zhong, D. Wang, W. Wang, and M. Lieber, "High performance silicon nanowire field effect transistors, " *Nano Lett.*, vol. 3, pp. 149-152, 2003.

[4] J. Wang, E. Polizzi, and M. Lundstrom, "A computational study of ballistic silicon nanowire transistors," *In IEDM Tech. Dig., Washington, DC,* pp. 695–598, 2003.

[5] D. Hisamoto, W.C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, et al, "FinFET—A self-aligned double-date MOSFET scalable to 20 nm," *IEEE Trans. Electron Devices*, vol. 47, pp. 2320–2325, 2002.

[6] B. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, et al, "High performance fully-depleted tri-gate CMOS transistor," *IEEE Electron Device Lett.*, vol. 24, pp. 263–265, 2003.

[7] J. Park, J.-P. Colinge, and C. H. Diaz, "Pi-gate SOI MOSFET," *IEEE Electron Device Lett.*, vol. 22, pp. 405–406, 2001.

[8] F. L. Yang, H. Y. Chen, F. C. Chen, C. C. Huang, C. Y. Chang, H. K. Chiu, et al, "25 nm CMOS omega FETs," *In IEDM Tech. Dig*., pp. 255–258, 2002.

[9] X. Duan, C. Niu, V. Sahi, J. Chen, J. W. Parce, S. Empedocles, et al, "High-performance thin-film transistors using semiconductor nanowires and nanoribbons," *Nature*, vol. 425, pp. 274–248, 2003.

[10] E. Leobandung, J. Gu, L. Guo, and S. Y. Chou, " Wire-channel and wrap-around-gate metal-oxide-semiconductor field-effect transistors with a significant reduction of short channel effects," *J. Vac. Sci. Technol. B, Microelectron. Process. Phenom*., vol. 15, pp. 2791–2794, 1997.

[11] T. Bryllert, L. E. Wernersson, L. E. Froberg, and S. Samuelson, "Vertical high-mobility wrap-gated InAs nanowire transistor," *IEEE Electron Device Lett.*, vol. 27, pp. 323–325, 2006.

[12] A. Rahman, J. Guo, S. Datta, and M. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices* 2003; 50: 1853–64.

[13] J. Wang, E. Polizzi, and M. Lundstrom, "A three-dimensional quantum simulation of silicon nanowire transistors with the effective-mass approximation," *J. Appl. Phys*., vol. 96, pp. 2192–2203, 2004.

[14] B. C. Paul, R. Tu, S. Fujita, M. Okajima, T. H. Lee, and Y. Nishi, "An Analytical Compact Circuit Model for Nanowire FET," *IEEE Trans. Electron Devices*, vol. 54, pp. 1637-1644, 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance comparison of zero-Schottky-barrier single and double walled carbon nanotube transistors

Md. Abdul Wahab[1] and Khairul Alam[2]
Department of Electrical and Electronic Engineering
[1]United International University Dhaka-1209, Bangladesh
[2]East West University, Dhaka-1212, Bangladesh
Email: awahab@eee.uiu.ac.bd

***Abstract*— Atomistic quantum simulation is performed to compare the performance of zero-Schottky-barrier single walled (SW) and double walled (DW) carbon nanotube transistors having almost equal magnitude of band gap. The DW nanotube is generated from two semiconducting SW tubes and the SW tube is the outer tube of the DW nanotube. The DW nanotube transistor has better off current, better inverse subthreshold slope, and better on/off current ratio. The SW nanotube transistor has better on current and switching performance. The better switching performance of SW nanotube transistor is the consequence of higher transconductance and on current that results from current saturation in DW nanotube transistor after source-channel flat band condition. The inverse subthreshold slope of DW nanotube transistor is 63.11 mV/dec and that of SW nanotube transistor is 65.26 mV/dec. The on-state transconductance is 21.0963 μS and 1.5023 μS, the intrinsic switching delay is 33.6415 fS and 55.0184 fS, respectively for SW and DW nanotube transistors.**

I. Introduction

Since its discovery in 1991 by the Japanese electron microscopist Sumio Iijima [1], a significant progress has been achieved to understand the fundamental properties of carbon nanotube (CNT) and to explore its applications in future nanoelectronics. Its development as a field-effect transistor (FET) has been rapid since the first demonstration of carbon nanotube field-effect transistor (CNTFET) in 1998 [2], [3]. Carbon nanotubes can be single walled, double walled, or multi walled (MW) depending on whether they consist of one, two or several graphene sheets rolled-up into a concentric cylinder. Early attempt to fabricate room temperature MW CNTFETs was not successful due to larger radii of MW CNTs (about 10 nm), and the experimental and theoretical research work has mainly focused on FETs based on SW CNTs [3]–[18]. Research on DW CNTFETs is still in early stage [19]–[23]. Synthesis of DW nanotubes has been performed [24], [25]. However, few measurement on electrical transport in DW nanotube has been carried out so far [26]. For a DW CNT with metallic inner tube and semiconducting outer tube, Wang *et al.* [21], [23] have found that free charges in the inner metallic shell screen the outer semiconducting shell from the gate effect, and this screening is directly related to the inter shell interaction. This screening is disadvantageous to the performance of DW CNTFETs [21]. Electronic transport properties of Cs-encapsulated DW CNTs synthesized via a plasma irradiation method have been studied [22]. The Cs-encapsulated DW CNTs exhibit high performance n-channel FETs in contrast to ambipolar behavior of pristine DW CNTs.

In this paper, we compare the characteristics and performance of ballistic SW and DW CNTFETs using p_z orbital basis quantum simulation. The simulation model self-consistently solves the non-equilibrium Green function (NEGF) equations for charge density and the two dimensional Poisson's equation in cylindrical coordinates for electrostatics. The DW nanotube is generated from two semiconducting, (10,0) and (19,0), zigzag nanotubes with wall separation of 0.34 nm [27]–[29] and the SW nanotube is the (19,0) zigzag tube. We make this choice because the band gap of a DW CNT is controlled by the outer tube, and therefore, the SW and the DW CNTs have almost same band gap in our study. Our simulation shows that both the SW and the DW nanotubes have very similar I-V characteristics. However, the DW CNTFET has lower values of on-state transconductance and capacitance and longer delay.

II. Simulation Model

The simulation model uses a self-consistent solution between electrostatics and charge density. For a coaxially gated CNTFET, we obtain electrostatic potential by solving two-dimensional Poisson's equation in cylindrical coordinates (r, ϕ, z)

$$\frac{\partial^2 V}{\partial r^2} + \left[\frac{1}{r} + \frac{1}{\epsilon}\frac{\partial \epsilon}{\partial r}\right]\frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial z^2} = \frac{-\rho}{\epsilon}. \qquad (1)$$

The Poisson's equation, Eq. (1), is discretized using finite difference and solved by standard Newton-Raphon method. The potential is fixed to $V_{GS} - \Phi_G/q$ at the gate electrode, to $-\Phi_S/q$ at the source electrode, and to $V_{DS} - \Phi_D/q$ at the drain electrode. Here V_{GS} and V_{DS} are the gate to source and drain to source voltages, and Φ_G, Φ_S, and Φ_D are the work functions of gate, source, and drain metallizations. Von Neumann boundary conditions are used along the exposed surface of the dielectric. There, the radial component of the electric field is set to zero.

For charge density calculation, we use nearest-neighbor π-bond model to create Hamiltonian and recursive Green's function (RGF) algorithm [30]–[32] to solve non equilibrium Green's function (NEGF) equations. The bonding energy between two atoms of inner tube and outer tube is calculated from [33]–[36]

$$\begin{aligned} -\gamma(\mathbf{R_1},\mathbf{R_2}) &= V_{pp\sigma}\exp\left(-\frac{d-c/2}{\delta}\right)\left(\frac{\mathbf{p}_1.\mathbf{d}}{d}\right)\left(\frac{\mathbf{p}_2.\mathbf{d}}{d}\right) \\ &\quad -V_{pp\pi}\exp\left(-\frac{d-a_0}{\delta}\right)[(\mathbf{p}_1.\mathbf{e})(\mathbf{p}_2.\mathbf{e}) \\ &\quad +(\mathbf{p}_1.\mathbf{f})(\mathbf{p}_2.\mathbf{f})], \end{aligned} \tag{2}$$

where $\mathbf{R_1}$ and $\mathbf{R_2}$ are the atomic positions in inner tube and outer tube respectively, $V_{pp\sigma}$ and $V_{pp\pi}$ are the energy integrals between π orbitals decomposed into the directions parallel and perpendicular, respectively, to the vector $\mathbf{d}=\mathbf{R}_1-\mathbf{R}_2$, $\mathbf{p}_1$ and $\mathbf{p}_2$ are unit vectors directed along π orbitals at $\mathbf{R}_1$ and $\mathbf{R}_2$, respectively, $d=|\mathbf{d}|$, $\mathbf{e}$ and $\mathbf{f}$ are unit vectors perpendicular to $\mathbf{d}$ and to each other, a_0 is the distance between two nearest neighbor carbons in two dimensional graphite, c is the lattice constant along the c axis in bulk graphite, and δ is the decay rate of π orbital. Within the layer, the on-site energy is ϵ_p and the carbon-carbon bond energy is $V_{pp\pi}$. The Hamiltonian parameter values are taken from [37].

With the Hamiltonian, we calculate the charge density at the L^{th} atomic layer from

$$\rho_L = (2q)\int \frac{dE}{2\pi}\mathrm{tr}\{f_S A^L_{L,L} + f_D\left[A_{L,L}-A^L_{L,L}\right]\} \tag{3}$$

where the factor of 2 comes from spin degeneracy and f_S and f_D are the source and drain Fermi functions, respectively. The full spectral function is calculated from $A_{L,L} = \mathrm{Im}\left(G_{L,L}-G^\dagger_{L,L}\right)$ and the left spectral function from $A^L_{L,L} = G_{L,1}\Gamma_{1,1}G^\dagger_{L,1}$. The Green's function is defined as $G=(E-H-\Sigma)^{-1}$ and the broadening function Γ is negative twice the anti Hermitian component of self energy $\Sigma_{1,1} = -it_{1,0}$, where $t_{1,0}$ is the coupling matrix between 0^{th} and 1^{st} atomic layers. The simulation starts with an initial guess of potential profile calculated by solving Laplace's equation and Andrson mixing scheme is used to accelerate convergence. Once the potential profile is converged, the coherent current is calculated from

$$I = \frac{2q}{h}\int dE T(E)\left(f^L - f^R\right), \tag{4}$$

where the transmission is [30]

$$T(E) = \mathrm{tr}\left[\Gamma_{1,1}\left(A_{1,1}-G_{1,1}\Gamma_{1,1}G^\dagger_{1,1}\right)\right]. \tag{5}$$

III. Numerical Results and Discussions

We simulate coaxially gated zero-Schottky-barrier SW and DW CNTFETs. The SW CNT is a (19,0) zigzag tube with diameter of 1.5 nm and band gap of 0.5202 eV and the DW CNT is generated from (10,0) and (19,0) zigzag tubes that has a band gap of 0.5163 eV. The device, shown in Fig. 1, has a gate length L_g of 5 nm, source and drain underlaps L_u of 22.5 nm, and source and drain metal extensions L_{ex} of 15 nm. The gate oxide thickness t_{ox} is 2 nm and the gate dielectric is SiO_2. Poisson solver uses an extended dielectric thickness t_{ox-ex} of 6 nm so that the fringing electric fields emanating from source, drain, and gate metals are taken care of. The gate metal has the same work function as the CNT has and the source and drain metal Fermi functions align with the conduction band of the tube that is corresponding to zero-Schottky-barrier. The conduction and valence bands of CNTs are essentially symmetric around the center of the band gap so all results apply equally to having the Fermi level aligned with the valance band edge. Ballistic transport is assumed. Ballistic transport in short CNTs ($\sim$ 60 nm) has been confirmed experimentally by measuring conductance at different temperatures [13].

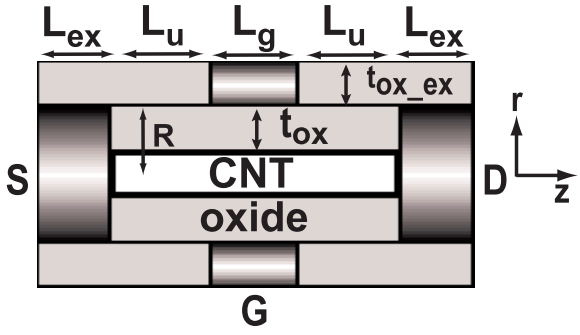


Fig. 1. Device cross section used for simulation.

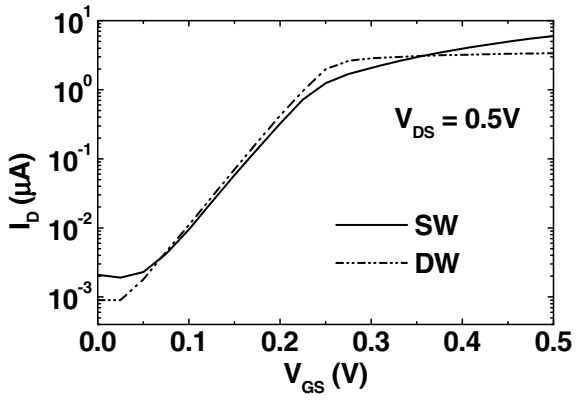


Fig. 2. Simulated log I_D versus V_{GS} characteristics of SW and DW CNTFETs.

The simulated log I_D versus V_{GS} plots for both SW and DW CNTFETs are shown in Fig. 2. The off current (V_{GS} = 0 V) is slightly better in DW CNTFET and the current in DW CNTFET gets saturated after gate bias of $\approx$ 0.25 V. Also the current increase in SW CNTFET after gate bias of 0.25 V is not much. If we define V_{GS} = 0.5 V as the on-state, then on-state current is 5.99 μA for SW and 3.38 μA for DW CNTFETs. The off-state currents are $2.1\times10^{-3}\mu$A and 8.6 $\times10^{-4}$ μA, respectively for the SW and DW CNTFETs. The on/off current ratio of SW and DW CNTFET is 2.9×10^3 and

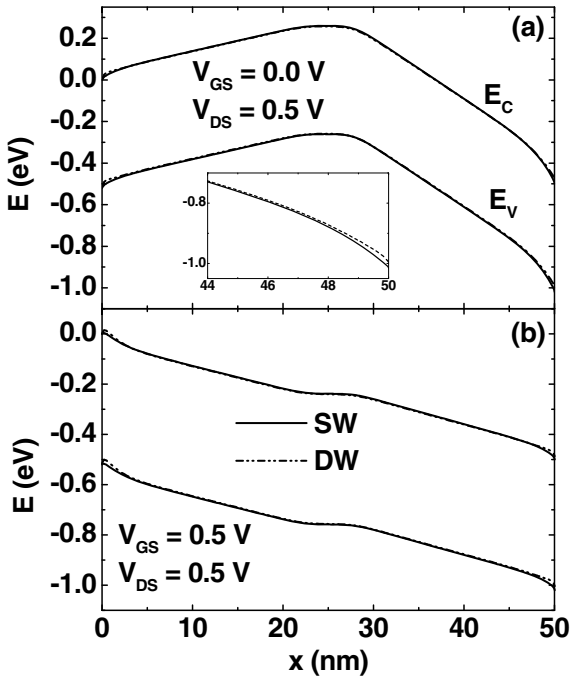


Fig. 3. Conduction and valence band profiles of SW and DW CNTFETs in (a) off-state and (b) on-state.

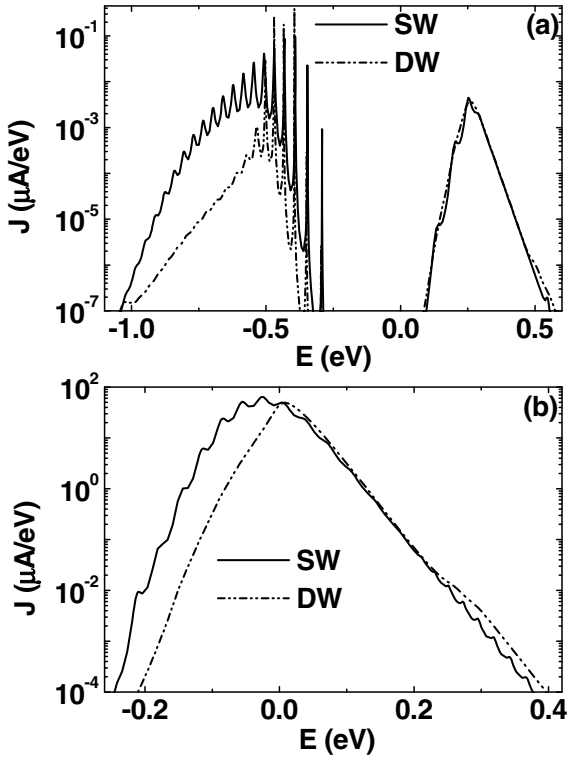


Fig. 4. Energy spectrum of current in (a) off- and (b) on-states of SW and DW CNTFETs.

3.9×10^3, respectively. The inverse subthreshold slopes defined as $S = (\partial log_{10} I_D/\partial V_{GS})^{-1}$ are 65.26 mV/dec and 63.11 mV/dec for SW and DW CNTFETs, respectively. The inverse subthreshold slope has been calculated in the linear region of the log I_D-V_{GS} characteristic curves of Fig. 2.

To understand the off-state and on-state I-V characteristics of SW and DW CNTFETs, we plot, in Fig. 3, the conduction and valence band profiles, and in Fig. 4, the energy spectrum of current, in off- and on-states for both SW and DW CNTFETs. The band profiles of DW CNTFET is the energy versus position plots at the surface of outer tube. The source Fermi level is at 0 eV and the drain Fermi level at -0.5 eV. The band profile of SW and DW SNTFETs is almost similar in both off- and on-states. However, there is a slight change in the band profile near the source and drain contacts as shown in the inset of Fig. 3(a) near the drain contact. This slight change in band profile near the contacts changes the tunneling current as shown in the energy spectrum of current in Fig. 4. In off-state, the thermal plus electron tunneling current through the conduction band (E>0 of Fig. 4(a)) of both SW and DW CNTFETs is almost same. However, the hole tunneling current through the valence band (E< 0 of Fig. 4(a)) of DW CNTFET is smaller than that of SW CNTFET, and therefore, the off-state current in DW CNTFET is smaller as shown in Fig. 2. In on-state, both the devices have thermal plus electron tunneling current. The electron tunneling current (E< 0 of Fig. 4(b)) in DW CNTFET is smaller, and therefore, the on-state current shown in Fig. 2 is smaller in DW CNTFET. After gate bias of 0.25 V, the thermal current does not change due to potential barrier saturation at $\approx$ 0 eV, and the gate bias does not modulate the tunneling current, and therefore, the current gets saturated in DW CNTFET. The slight increase of current in SW CNTFET after gate bias of 0.25 eV is due to slight change in the tunneling current through the conduction band.

Finally we compare the performance metrics, namely the transconductance g_m, the intrinsic switching delay τ_S, and the intrinsic unity current gain frequency f_T. For this the gate capacitance is calculated from

$$C_g = 2\pi R \int_0^{L_g} dz \frac{\delta D_r}{\delta V_g} + 2\pi \int_{t_{ox}}^{t_{ox-ex}} r dr \frac{\delta D_z}{\delta V_g}, \qquad (6)$$

where R is the radius of the dielectric covering t_{ox}. The first integral takes care of the fluxes emanating from the bottom surface of the gate metal and the second integral takes care of the fluxes emanating from the two sides of the gate metal facing to the source and drain. This gives the total gate capacitance $C_g = C_{gs} + C_{gd}$ which includes the effect fringing fields directly from the gate metal to the source and drain. The intrinsic switching delay is calculated from $\tau_S = C_g V_{DD}/I_{ON}$ and the intrinsic unity current gain frequency from $f_T = g_m/2\pi C_g$, where transconductance is computed from $g_m = \partial I_D/\partial V_{GS}$ at $V_{DS} = V_{DD}$. V_{DD} is 0.5 V in our study.

The gate capacitance and the transconductance versus gate bias plots for both the SW and DW CNTFETs are shown in Fig. 5. The transconductance peaks at a gate bias $\approx$ E_g/2q where there is approximately a flat band situation between the channel potential and the source Fermi level. This

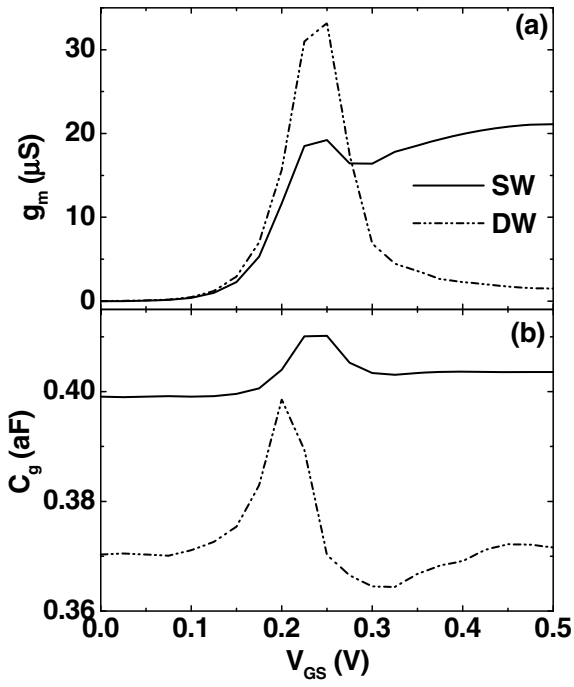


Fig. 5. (a) Transconductance and (b) gate capacitance versus gate bias plots for both the SW and DW CNTFETs. Here $_{DS}$ = 0.5 V.

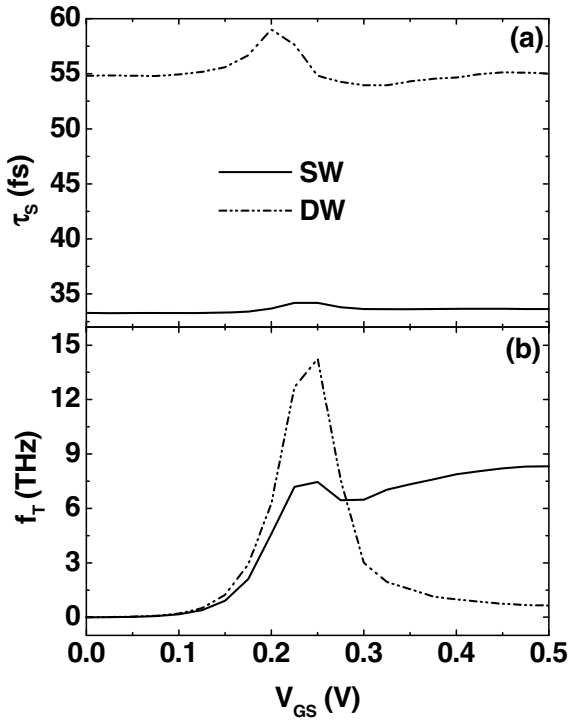


Fig. 6. (a) Intrinsic switching delay and (b) intrinsic unity current gain frequency versus gate bias plots for both the SW and DW CNTFETs. Here $_{DS}$ = 0.5 V.

phenomenon has been observed before for SW CNTFETs [38]. However, the gate capacitance peaks at a slightly lower gate bias for DW CNTFETs. The on-state value of g_m is much lower in DW CNTFET due to current saturation. Fig. 6 shows the τ_S and f_T versus gate bias plots for both the SW and DW CNTFETs. The switching delay is higher in DW CNTFETs due to lower value of C_g and I_{ON} and the f_T is lower due to lower value of g_m. The on-state current of SW CNTFET is ≈ 1.75 times larger than that of DW CNTFET. The on-state value of g_m is 21.0963 and 1.5023 μS, respectively, for SW and DW CNTFETs. While DW CNTFET shows better performance in terms of on/off current ratio and inverse subthreshold slope, the SW CNTFET has better performance in terms of on current and switching delay.

IV. CONCLUSION

Performance of zero-Schottky-barrier SW and DW nanotube transistors are compared using π-bond atomistic quantum simulation. Both the transistors have similar current-voltage characteristics, however the current in DW nanotube transistor saturates after source-channel flat band condition, and the slight increase in current of SW nanotube transistor beyond this point is due to tunnel barrier modulation with gate bias. The on current and switching performance is better in SW nanotube transistor and the DW nanotube transistor has better off current, inverse subthreshold slope, and on/off current ratio.

REFERENCES

[1] S. Iijima, "Helical microtubules of graphitic carbon," Nature, vol. 354, pp. 56–58, 1991.

[2] R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, "Single- and multiwall carbon nanotube field-effect transistors," Appl. Phys. Lett., vol. 73, pp. 2447–2449, 1998.

[3] S. J. Tans, A. R. M. Verschueren, and C. Dekker, "Room-temperature transistor based on a single carbon nanotube," Nature, vol. 393, pp. 49–52, 1998.

[4] K. Alam and R. Lake, "Performance of 2 nm gate length carbon nanotube field-effect transistors with source/drain underlaps," Appl. Phys. Lett., vol. 87, p. 073104, 2005.

[5] Y. Yoon and J. Guo, "Analysis of strain effects in ballistic carbon nanotube FETs," IEEE Trans. Elec. Dev., vol. 54, no. 6, pp. 1280–1287, 2007.

[6] S. Heinze, M. Radosavljevic, J. Tersoff, and P. Avouris, "Unexpected scaling of the performance of carbon nanotube schottky-barrier transistors," Phys. Rev. B, vol. 68, no. 2, p. 235418, 2003.

[7] R. Martel, V. Derycke, C. Lavoie, J. Appenzeller, K. K. Chan, J. Tersoff, and P. Avouris, "Ambipolar electrical transport in semiconducting single wall carbon nanotubes," Phys. Rev. Lett., vol. 87, no. 25, p. 256805, 2001.

[8] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Controlling doping and carrier injection in carbon nanotube transistors," Appl. Phys. Lett., vol. 80, no. 15, pp. 2773–2775, 2002.

[9] S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and P. Avouris, "Carbon nanotubes as schottky barrier transistors," Phys. Rev. Lett., vol. 89, no. 10, p. 106801, 2002.

[10] M. Radosavljevic, S. Heinze, J. Tersoff, and P. Avouris, "Drain voltage scaling in carbon nanotube transistors," Appl. Phys. Lett., vol. 83, no. 12, pp. 2435–2437, 2003.

[11] S. Heinze, J. Tersoff, and P. Avouris, "Electrostatic engineering of nanotube transistors for improved performance," Appl. Phys. Lett., vol. 83, p. 5038, 2003.

[12] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, "Ballistic carbon nanotube field-effect transistors," Nature, vol. 424, pp. 654–657, 2003.

[13] A. Javey, J. Guo, M. Paulsson, Q. Wang, D. Mann, M. Lundstrom, and H. Dai, "High-field quasiballistic transport in short carbon nanotubes," Phys. Rev. Lett., vol. 92, no. 10, p. 106804, 2004.

[14] A. Javey, J. Guo, D. B. Farmer, Q. Wang, D. Wang, R. G. Gordon,

M. Lundstrom, and H. Dai, "Carbon nanotube field-effect transistors with integrated ohmic contacts and high-k dielectrics," Nano Lett., vol. 4, no. 3, pp. 447–450, 2004.
[15] M. Radosavljevic, J. Appenzeller, and P. Avouris, "High performance of potassium n-doped carbon nanotube field-effect transistors," Appl. Phys. Lett., vol. 84, pp. 3693–3695, 2004.
[16] A. Javey, M.Shim, and H. Dai, "Electric properties and devices of large-diameter single-walled carbon nanotubes," Appl. Phys. Lett., vol. 80, no. 7, pp. 1064–1066, 2002.
[17] J. Guo, S. Goasguen, M. Lundstrom, and S. Datta, "Metal-insulator-semiconductor electrostatics of carbon nanotubes," Appl. Phys. Lett., vol. 81, no. 8, pp. 1486–1488, 2002.
[18] J. Appenzeller and D. J. Frank, "Frequency depedent characterization of transpoet properties in carbon nanotube transistors," Appl. Phys. Lett., vol. 84, no. 10, pp. 1771–1773, 2004.
[19] T. Shimada, T. sugai, Y. Ohno, S. Kishimoto, T. Mizutani, H. Yoshida, T. Okazaki, , and H. Shinohara, "Double-wall carbon nanotube field-effect transistors: Ambipolar transport characteristics," Appl. Phys. Lett., vol. 84, p. 2412, 2004.
[20] D. Kang, N. Park, J. Hyun, E. Bae, J. Ko, J. Kim, and W. Park, "Adsorption-induced conversion of the carbon nanotube field effect transistor from ambipolar to unipolar behavior," Appl. Phys. Lett., vol. 86, p. 093105, 2005.
[21] S. Wang, X. Liang, Q. Chen, Z. Zhang, and L. peng, "Field-effect characteristics and screening in double-walled carbon nanotube field-effect transistors," J. Phys. Chem. B, vol. 109, p. 17361, 2005.
[22] Y. Li, R. Hatakeyama, T. Kaneko, T. Izumida, T. Okada, and T. Kato, "Electronic transport properties of Cs-encapsulated double-walled carbon nanotubes," Appl. Phys. Lett., vol. 89, p. 093110, 2006.
[23] S. Wang and M. Grifoni, "Schottky-barrier double-walled carbon-nanotube field-effect transistors," Phys. Rev. B, vol. 76, p. 033413, 2007.
[24] R. R. Bacsa, C. Laurent, A. Peigney, W. S. Bacsa, T. Vaugien, and A. Rousset, "High specific surface area carbon nanotubes from catalytic chemical vapor deposition process," Chem Phys. Lett., vol. 323, pp. 566–571, 2000.
[25] B. W. Smith and D. E. Luzzi, "Formation mechanism of fullerene peapods and coaxial tubes: a path to large scale synthesis," Chem Phys. Lett., vol. 321, pp. 169–174, 2000.
[26] M. Kociak, K. Suenaga, K. Hirahara, Y. Saito, T. Nakahira, and S. Iijima, "Linking chiral indices and transport properties of double-walled carbon nanotubes," Phys. Rev. Lett., vol. 89, p. 155501, 2002.
[27] J. C. Slonczewski and P. R. Weiss, "Band structure of graphite," Phys. Rev., vol. 109, no. 2, pp. 272–279, 1958.
[28] S. Bandow, M. Takizawa, K. Hirahara, and M. Y. S. Iijima, "Raman scattering study of double-wall carbon nanotubes derived from the chains of fullerenes in single-wall carbon nanotubes," Chem. Phys. Lett., vol. 337, p. 48, 2001.
[29] J. W. McClure, "Band structure of graphite and de haas-van effect," Phys. Rev., vol. 108, p. 612, 1957.
[30] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," J. Appl. Phys., vol. 81, no. 12, pp. 7845–7869, 1997.
[31] K. Alam and R. K. Lake, "Leakage and performance of zero-schottky-barrier carbon nanotube transistors," J. Appl. Phys., vol. 98, p. 064307, 2005.
[32] S. Datta, Quantum Transport Atom to Transistor. Cambridge: Cambridge University Press, 2005.
[33] S. Uryu and T. Ando, "Electronic intertube transfer in double-wall carbon nanotubes with impurities: Tight-binding calculations," Phys. Rev. B, vol. 76, p. 155434, 2007.
[34] T. Nakanishi and T. Ando, "Conductance of crossed carbon nanotubes," J. Phys. Soc. Jpn, vol. 70, p. 1647, 2001.
[35] S. Uryu, "Electronic states and quantum transport in double-wall carbon nanotubes," Phys. Rev. B, vol. 69, p. 075402, 2004.
[36] J. C. Slater and G. F. Koster, "Simplified LCAO method for the periodic potential problem," Phys. Rev., vol. 94, no. 6, pp. 1498–1524, 1954.
[37] P. Lambin, V. Meunier, and A. Rubio, "Electronic structure of polychiral carbon nanotubes," Phys. Rev. B, vol. 62, p. 5129, 2000.
[38] K. Alam and R. Lake, "Performance metrics of a 5 nm, planar, top gate, carbon nanotube on insulator (COI) transistor," IEEE Trans. Nanotechnol., vol. 6, no. 2, pp. 186 – 190, 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

PI-Clay Nanocomposites: Synthesis and Characterization

Shaikh Md. Mominul Alam

College of Textile Technology, Tejgaon, Dhaka -1208, Bangladesh.

Email: dalim70@yahoo.com

Abstact: A series of Polyimide (PI)-Organically modified clay nanocomposites were made to enhance tensile modulus, thermal stability of rigid PI. PI was made from 3′, 4, 4′-biphenyltetracarboxylic dianhydride (BPDA), p-phenylediamine (PDA). Montmorillonite, one type of layered clay, was treated by n hexadecyltrimethylammonium bromide salt. XRD indicated that OMMT layers were exfoliated and dispersed into PI-film. Tensile measurements indicated that small amount of OMMT (up to 3%) increased tensile modulus nicely. The glass transition temperatures are higher than pristine PI. TGA showed that nanocomposites have higher decomposition temperatures in comparison with the original PI.

I. Introduction

Organic-inorganic nanocomposites usually have unique properties because of the combination of advantages of inorganic materials like rigidity, low co-efficient of thermal expansion (CTE), high thermal stability and the organic polymers like flexibility, dielectricity, processability etc. Due to the distribution at nano-meter size, organic-inorgannic composites often exhibit some special mechanical, electronic properties which extend their application to many new sectors [1-4].

Among organics, PI is well-known as high performance polymer which exhibits outstanding dielectric, mechanical properties, thermal stability and low CTE. Rigid PI from 3′, 4, 4′-biphenyltetracarboxylic dianhydride (BPDA), p-phenylediamine (PDA) are widely used in microelectronics and aerospace industries because of its unique features like high modulus, low CTE and good mechanical strength, low solvent swelling and moisture uptake and spontaneous in-plane orientation. PIs possess limitations for much higher performance applications in which inorganic materials are used, because of their intrinsic nature as organic materials [5]. Inorganic materials exhibit excellent thermal stability and high modulus. Thus, the formation of nanocomposites of PI with inorganic materials has been suggested to meet the demands of balanced properties for both organic and inorganic materials. There are different types of inorganics like layered silicate (clay), silica etc. which are used to developed PI-inorganic nanocomposites. The most commonly used clay in the preparation of polymer/clay nanocomposites is Montmorillonite (MMT) in which galleries naturally exist inorganic cations, balancing the charge of oxide layers in a hydrophilic environment. The ion-exchange of these cations with the organic ammonium salts affords hydrophobic environment inside the galleries of MMT (OMMT) [6]. The resulting organophilic galleries of OMMT will exchange the compatibility with polymers, improve the dispersion of the silicate layers into the matrix [7] and assist the penetration of monomers and/or polymers into the gallories [8]. Also, organic ammonium salts can provide functional groups that can react or interact with the monomers or polymers to improve the interfacial strength between the reinforcement and the polymer matrix [9].

In the present study, PI (BPDA/PDA)/OMMT nanocomposites were made with different ratio of inorganics (1, 3, 5, 10%). The performance mainly thermal and mechanical properties of the nanocomposite films were studied then in details.

II. Experimental

2.1. Reagents

Kunipia-F, a Na+-montmorillonite, was supplied by Kunimine Ind. Co., with cation exchange capacity (CEC) of 119 meq/100g. OMMT was prepared from MMT by ion-exchange reaction using n hexadecyltrimethylammonium bromide according to the reported method [7]. BPDA from UBE Industries Ltd., Japan and PDA from Tokyo Kasei Kogyo Co. Ltd, Japan were purified by sublimation. N-Methyl-2-Pyrrolidone (NMP) from Osaka Chemicals, Japan was dried by distillation under reduced pressure over Sodium Hydride. The 35wt% Hcl and n hexadecyltrimethylammonium bromide from Tokyo Kasei Kogyo Co. Ltd, Japan, were used as received. PAA was prepared from BPDA and PDA (Scheme 1).

2.2. Preparation of PI-clay nanocomposite

PI-OMMT nanocomposites with different weight percentage of OMMT (1, 3, 5, 10%) were prepared according to the reported method [14].

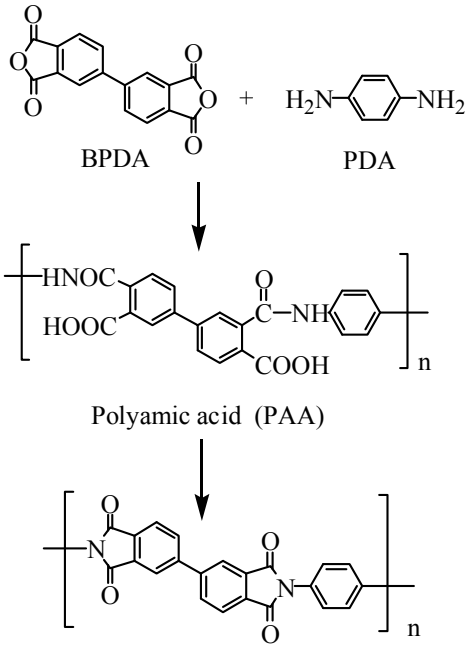
2.3. Measurements

IR spectra were obtained with Jasco Spectrophotometer model FT/IR-420. DSC was recorded using Rigaku Thermo Plus 2DSC8230 at a heating rate of 10˚C/min under nitrogen. TGA was performed with Rigaku Thermo Plus 2TG-DTA TG8120 at a heating rate of 5˚C/min under argon. Dynamic mechanical analysis (DMA) were conducted on ORIENTEC Automatic Dynamic Viscoelastomer Rheovibron model DDV-01FP at 35Hz at a heating rate of 4°C/min. XRD was measured in reflection mode using a X-ray diffractometer, Rigaku, RINT2000 using CuKα radiation. Tensile properties were recorded with Imada Seisaku-sho Model SV-3 at a cross-head speed of 1mm/min using films of 2 cm long. Transparency was checked by JASCO V-550 UV/vis spectrophotometer in where all samples thickness were about 0.03 mm.

III. Results and Discussion

3.1. Preparation of OMMT

In our study, we used ammonium salt of n-hexadecyltrimethyl bromide for ion exchanging of the Na+ ions on the MMT surfaces to render the surface hydrophobic. Thus, there will be hydrophobic environment into the clay galleries to accommodate the hydrophobic PAA. XRD measurements (Fig.1) showed that the interlayer spacing of MMT (2θ=7.22, d=1.22 nm) was increased after surface



Scheme 1. Preparation of PI

treatment with ammoniumbromide salt of n-hexadecyltrimethyl to afford OMMT (2θ=4.52, d=1.95nm)

3.2. Morphological study

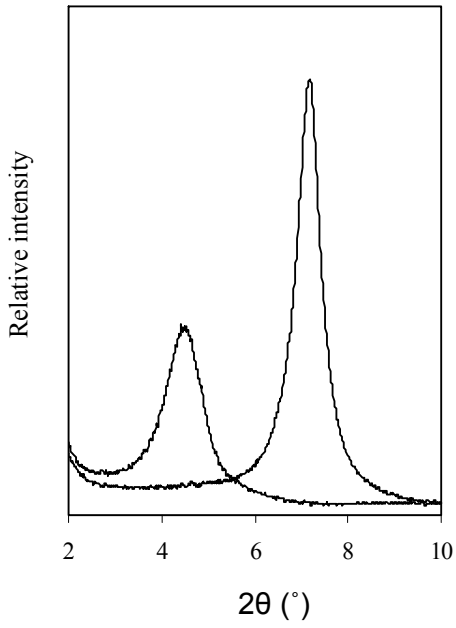


Fig.1. XRD of MMT and OMMT

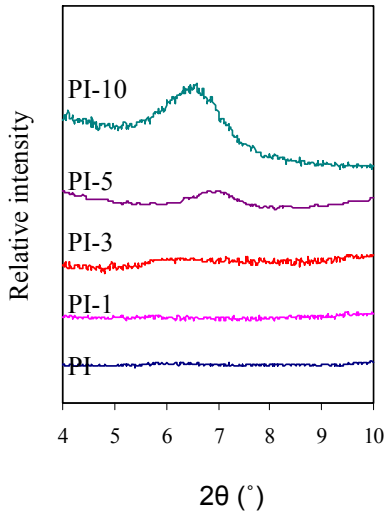


Fig.2. XRD of various PIs

XRD patterns of different PI films are shown in Fig.2. With the increase of the OMMT loading 5 and 10%, the hybrid show a slight pick corresponding to the basal spacing suggesting that a small part of OMMT was not dispersed in the molecular level and aggregates. The PI-10%OMMT related composites became darker due to aggregation of OMMT.

Transparency of all composites were checked by UV-spectrophotometer. From UV-spectras, it was found that the transparency became 80, 76.7, 66.9, 52, 44.3% in case of 0, 1, 3, 5, 10% OMMT related composites at 700nm wave no (Fig. 3). Higher amount of inorganics related PI-OMMT composites became translucent which also proved the aggregation of inorganics at higher ratio.

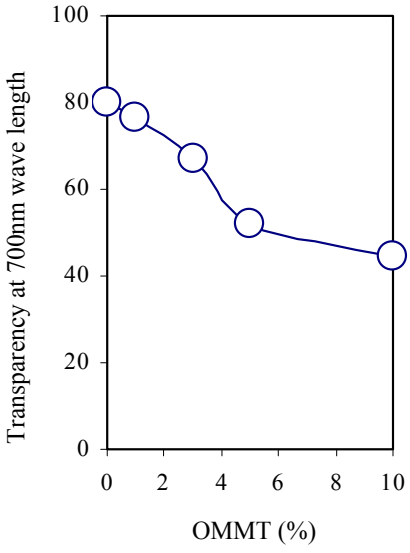


Fig.3. Effect of OMMT on Transparency

IV. Performance of PI-inorganic composites

4.1. Tensile properties of PI-inorganics composites

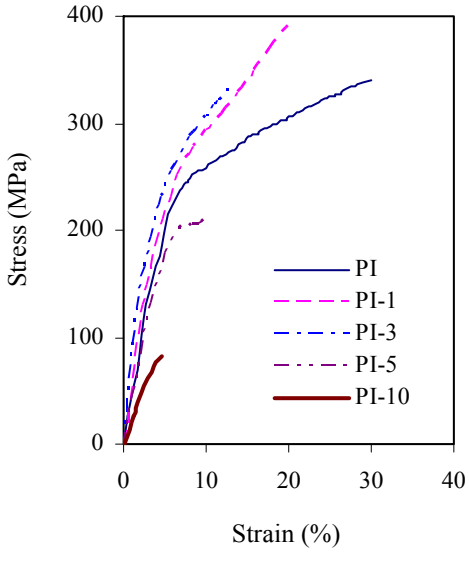


Fig.4. Tensile properties of various PIs

Inorganics has higher rigidity than organics but brittle behaviour which effect in composites properties as well. Pristine PI tensile modulus was 8.2 GPa which became 10.3, 11.5, 6.5, 3.5 GPa for PI-OMMT composites due to the reinforcement effect attained by the dispersion of clay nanolayers in PI (Fig. 4). At higher contents of clay (5, 10%), the layered silicate not exfoliated, it means less reinforcement effect which decreased the modulus of the hybrids. Tensile strength converted from 340MPa to 390, 332, 210, 165 MPa for PI-1 to 10% PI-OMMT and composites. Small amount of inorganic (1%) increased the strength of organic-inorganic hybrid which is natural for interfacial interaction. OMMT decreased the elongation at break % due to the brittle behaviour of clay.

4.2. Dynamic mechanical analysis (DMA) of PI-inorganics composites

The viscoelastic properties of the composites were studied using DMA. The storage modulus (E'), loss modulus (E") and tanδ were measured from Fig. 5 for PI-OMMT. The increase of T_gs of the nanocomposites in comparison with the neat PI can be attributed to maximizing the adhesion between the polymer and the inorganic surfaces because of the nanometer size which restricts the segmental motion near the organic and inorganic interface. The storage

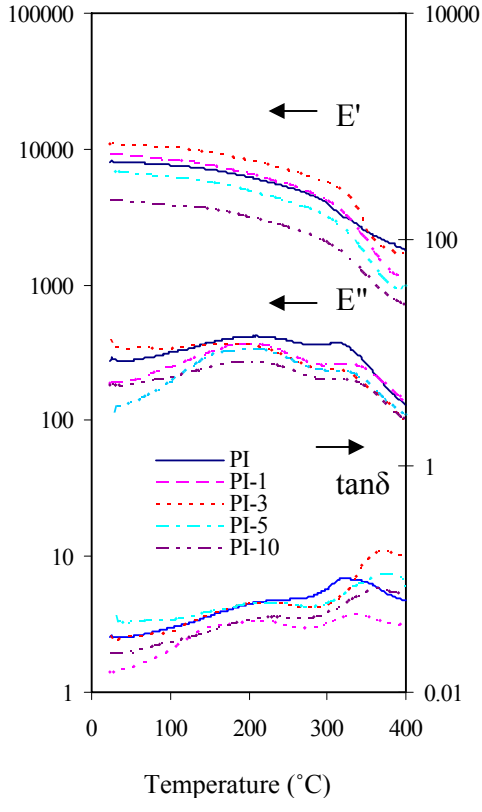


Fig.5. DMA of various PIs

modulus at room temperature arise highest at 3%OMMT and after that it started to decrease at 5, 10%OMMT.

4.3. Thermal properties of PI-inorganic nanocomposites

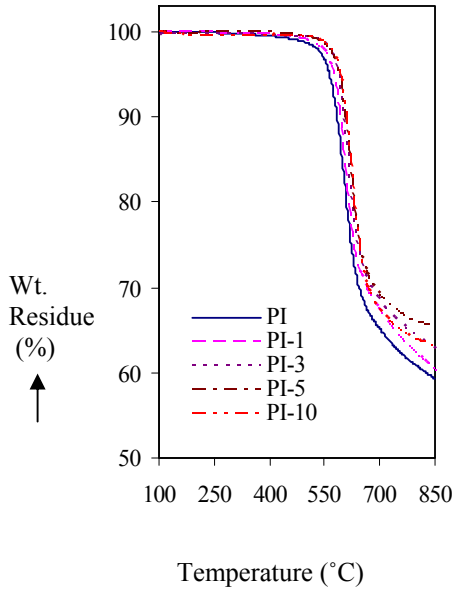


Fig. 6. TGA of various PIs

TGA were recorded for PI-inorganic nanocomposites to monitor the effect of OMMT and silica on the thermal properties. In case of PI-OMMT nanocomposites, the 5 and 10% degradation temperature (T_5, T_{10}) showed that the thermal stability of the PI-films were enhanced by the incorporation of the clay nanolayers. For example, 5% OMMT increaed the T_5 about 26˚C than the corresponding neat PI. The effect of clay almost constant at 10% loading probably because of the aggregation of additional OMMT (Fig. 6). The homogeneously dispersion of clay nanolayers hinders the permeability of volatile degradation products out from the material and showed delayed decomposition [15-17].

V. Conclusions

The properties of inorganics are distributed nicely and prominently in rigid PI in case of rigid PI-inorganic nanocomposites. Higher thermal properties of inorganics increased the thermal behaviour of the composites. The rigidity of inorganics increased the tensile modulus of composites in presensence of small amount of inorganics. T_g enhanced prominently in PI-inorganic nanocomposites as well. PI-small amount of OMMT (1, 3%) showed better performance concerning modulus, thermal stability than PI-silica composites.

Acknowledgements: The author is grateful to Prof. T. Takeichi, Department of Materials Science, Toyohashi University of Technology, Japan for his every side support of this research.

References

[1]. M. Ree, T. L. Nunes and J. S. Lin, Polymers, 35, 1148, 1994.
[2]. C. Yen, W. Chen, D. Liaw, H. Lu, Polymer, 44, 7079, 2003.
[3]. W. Qiu, Y. Luo, F. Chen, Y. Duo, H. Tan, Polymer, 44, 5821, 2003.
[4]. E. R. Leite, F. L Souza, P. R. Bueno, S. D. Lazaro, E. Longo, Chem Mater, 17, 4561, 2005.
[5]. A. Morikawa, Y. Iyoku, M. Kakimoto, Y. Imai, Polym. J., 24 , 107, 1992.
[6]. A. Akelah, P. N. In Prasad, J. E. Mark, J. F. Tung, editors. Polymers and others advanced materials: emerging technologies and business opportunities. New York: Plenum, 1995.
[7]. M. Kawasumi, N. Hasegawa, M. Kato, Y. Kojima, A. Usuki, A. Okada, Macromolecules, 30, 6333, 1997.
[8]. P. C. LeBaron, Z. Wang, T. J. Pinnavaia, Appl. Clay Sci., 15, 11, 1999.
[9]. E. P. Giannelis, In: Mann S editor. Biomimetric materials chemistry. New York: VCH, 1996.
[10]. S. Inoue, K. Morita, K. Asai, H. Okamoto, J. Appl. Sci. 92, 2211-2219, 2004.
[11]. H. Wang, W. Zhong, Q. Du, Y. Yang, H. Okamoto, S. Inoue, Polymer Bulletin 51, 63-68, 2003.
[12]. A. Al Arbash, Z. Ahmad, F. Al-Sagheer, A. A. M. Ali, J. Nanocomposites, Article ID 58648, 1-9, 2006.
[13]. Z. Shang, C. Lu, L. Gao, Polymer Int.55, 1277-1282, 2006.
[14]. T. Agag, T. Koga, T. Takeichi, Polymer, 42, 3399, 2001.
[15]. M. Inagaki, T. Takeichi, Y. Hishiyama, A. Oberlin, Chem Phys Carbon, 26, 245, 1999.
[16]. S. M. M. Alam, T. Agag, T. Kawauchi, T. Takeichi, Reactive & Functional Polymers, 67, 1218-1224, 2007.
[17]. H. Chao-Cheng, J. Guang-Way, C. Kung-Chin, H. Wei-I and Y. Jui-Ming, Polymer International, 57, 605-611, 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

The effects of doping, gate length, and gate dielectric on inverse subthreshold slope and on/off current ratio of a top gate silicon nanowire transistor

Sishir Bhowmick[1], Khairul Alam[2], and Quazi Deen Mohd Khosru[1]
Department of Electrical and Electronic Engineering
[1]Bangladesh University of Engineering and Technology, Dhaka-1000
[2]East West University, Dhaka-1212, Bangladesh
Email: sishir@eee.buet.ac.bd

***Abstract*— The effects of gate length L_g, gate dielectric constant ϵ_{ox}, gate oxide thickness t_{ox}, and sourc/drain doping concentration on inverse subthreshold slope and on/off current ratio of a top gate silicon nanowire on insulator device are studied using three dimensional quantum simulation. The variation of inverse subthreshold slope and on/off current ratio are very sensitive to gate length, gate dielectric constant, and oxide thickness and relatively less sensitive to doping concentration. Significant improvement in subthreshold slope and on/off current ratio can be achieved using high-K gate dielectric with thinner oxide and relatively longer gate. The key feature of this improvement is the better gate control of channel potential with longer L_g, higher ϵ_{ox}, and thinner t_{ox}. Due to better control of channel potential, the tunneling current through the conduction band is significantly suppressed in the subthreshold regime that improves the subthreshold slope and on/off current ratio.**

I. Introduction

Scaling the transistor sizes has made significant improvement in the cost effectiveness and performance of integrated circuit over the last few decades. The bulk CMOS technology is rapidly approaching the scaling limit and alternate materials or device structures are essential for future electronics. One dimensional nanostructures such as the carbon nanotubes and silicon nanowires are the attractive materials for future nanoelectronics because their electronic properties can be controlled in a predictable manner. Controlled growth of silicon nanowires (SiNWs) down to 3 nm diameter [1], their applications as field-effect transistors (FETs) [2]–[5], logic gates [6], and sensors [7] have been demonstrated.

In this paper, we study the effects of gate length L_g, gate dielectric constant ϵ_{ox}, gate oxide thickness t_{ox}, and source/drain doping concentration on inverse subthreshold slope and on/off current ratio. The gate length and the dielectric constant and thickness have significant effects on on/off current ratio and subthreshold slope. The key quantity is the short channel effects that becomes severe for shorter gate length devices. Relatively longer gate significantly suppresses the tunneling current through the conduction band and improves the subthreshold characteristics. High-K dielectric and thinner oxide also have better gate control that improves the on/off current ratio and subthreshold slope. Improvement of subthreshold slope and drain induced barrier lowering has been observed by shin [8], [9].

II. Simulation Model

The simulation model uses a self-consistent solution between three dimensional (3D) Poisson's equation and effective mass Schrodinger's equation. The 3D Poisson's equation in cartesian coordinates is

$$\frac{\partial}{\partial x}\left(\epsilon\frac{\partial V}{\partial x}\right)+\frac{\partial}{\partial y}\left(\epsilon\frac{\partial V}{\partial y}\right)+\frac{\partial}{\partial z}\left(\epsilon\frac{\partial V}{\partial z}\right)=-\frac{\rho}{\epsilon_o}, \quad (1)$$

where ϵ_o is the free space permittivity, ϵ is the relative dielectric constant, V is the 3D potential, and ρ is the charge density, which is non-zero in silicon nanowire only. Poisson kernel is created by discretizing Eq. (1) using finite difference. Potential is fixed at the gate metal and zero field boundary condition is applied at the source and drain ends and at the exposed surfaces of dielectric. There the normal component of electric field is set to zero. Standard Newton Rapshon method is used to solve Poisson's equation.

The Schrodinger's equation in 3D cartesian coordinates is

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial x}\left(\frac{1}{m_x}\frac{\partial\psi}{\partial x}\right)-\frac{\hbar^2}{2}\frac{\partial}{\partial y}\left(\frac{1}{m_y}\frac{\partial\psi}{\partial y}\right)-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m_z}\frac{\partial\psi}{\partial z}\right)=E\psi \quad (2)$$

where ψ is the wave function, m_x, m_y, and m_z are the effective masses in device coordinates, and $\hbar$ is the reduced Plank's constant. The nanowire is grown in $\langle 100\rangle$ direction, which is device x coordinate in our study. Ballistic transport is assumed. Recursive Green's function algorithm (RGFA) [10] is used to solve Schrodinger's equation for charge density and current calculations. The open boundary condition in transport direction (x) is included in Schrodinger's equation via self-energy matrices. Hard-wall boundary condition is used in the transverse directions (y and z). For RGFA, the layer (cross-section) Hamiltonian and layer-to-layer coupling matrices are created by discretizing Eq. (2) using finite difference. The charge density at each grid point of the L^{th} layer (cross-

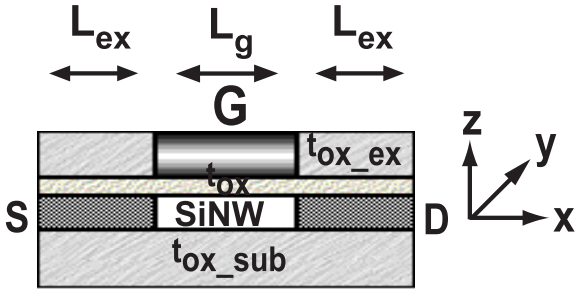


Fig. 1. Device cross section used for simulation. The device coordinates are also shown.

section) is calculated from

$$\rho_i = (4e) \int \frac{dE}{2\pi} \mathrm{diag}\left\{ f_S A^L_{i,i} + f_D \left[A_{i,i} - A^L_{i,i} \right] \right\}, \quad (3)$$

where e is the electronic charge, f_S and f_D are the source and drain Fermi functions, respectively, and the factor 4 includes spin plus valley degeneracy. The charge density is calculated for each pair of valley and summed over the pairs to obtain the total charge density. The full spectral function is calculated from the imaginary part of retarded Green's function $A_{i,i} = -2\mathrm{Im}\left(G_{i,i}\right)$, where $G = (E - H - \Sigma)^{-1}$. The left spectral function is calculated from $A^L_{i,i} = G_{i,1}\Gamma_{1,1}G^\dagger_{i,1}$, where $\Gamma_{1,1} = i\left(\Sigma_S - \Sigma_S^\dagger\right)$ is the broadening function and the self-energy is calculated from $\Sigma_S = t_{1,0}g_{0,0}t_{0,1}$. Here $t_{0,1}$ is the coupling matrix between the 0^{th} and 1^{st} layers (cross-sections) and $g_{0,0}$ is the surface Green's function calculated using decimation method [11].

The self-consistent loop starts with an initial guess of the potential profile. Anderson mixing [12] scheme is used to accelerate convergence. Once the convergence is achieved, the coherent drain current is calculated from

$$I_D = \frac{2e}{h} \int dE\ T(E)\left(f_S - f_D\right), \quad (4)$$

where transmission $T(E)$ is calculated from [10]

$$T(E) = \mathrm{tr}\left(\Gamma_{1,1}\left[A_{1,1} - G_{1,1}\Gamma_{1,1}G^\dagger_{1,1}\right]\right). \quad (5)$$

III. Simulation Results

The top gate silicon nanowire on insulator device used in our simulation is shown in Fig. 1. The silicon nanowire is placed on a thick substrate oxide layer t_{ox-sub}. The gate oxide of thickness t_{ox} is grown on the nanowire. An ultra-thin gate metal is deposited on the top of the gate oxide and the exposed regions on both sides of the gate metal are covered by oxide t_{ox-ex}. The nanowire has a square cross section of 5×5 nm^2 with a band gap E_g of 1.38 eV. The doped source-drain extension L_{ex} is 20 nm and fully ionized uniform donor concentration is assumed. The channel is undoped. The substrate oxide and the extended oxide are assumed to be SiO_2 with a dielectric constant of 3.9. The gate metal is assumed to have the same work function value as the nanowire has. The t_{ox-sub} value of 5 nm and the t_{ox-ex} value of 5 nm

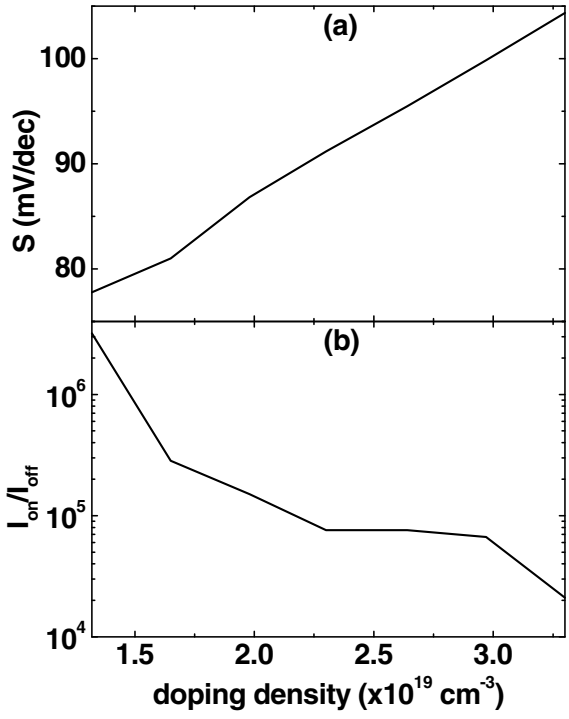


Fig. 2. Inverse subthreshold slope and on/off current ratio versus doping concentration.

are considered for Poisson solver so that the fringing electric fields emanating from the gate metals are treated correctly. The nanowire is modeled using bulk effective mass whose accuracy for wire cross section of 5×5 nm^2 and above has been varified [13], [14].

The simulated inverse subthreshold slope S and the on/off current ratio versus doping concentration are shown in Fig. 2. The simulation is performed for devices with gate length L_g of 10 nm, gate oxide thickness t_{ox} of 1 nm, and gate dielectric constant ϵ_{ox} of 3.9. The on/off current ratio reduces and the inverse subthreshold slope increases with higher doping concentration. If we double the doping concentration, the on/off current ratio is reduced by more than one order of magnitude and the inverse subthreshold slope becomes about 1.3 times. The scaling behavior of inverse subthreshold slope and on/off current ratio with gate length L_g is shown in Fig. 3. The devices used for simulation have gate oxide thickness t_{ox} of 1 nm, gate dielectric constant ϵ_{ox} of 3.9 and doping concentration of 1.65 $\times 10^{19}$ cm^{-3}. Device performance degrades with shorter gate due to short channel effects. The key quantity is the tunneling current through the conduction band that increases significantly for shorter gate devices. With longer gate length, the off current improves significantly and the on current degrades slightly that results in significant improvement in on/off current ratio. For example, the off current (mainly tunneling current) improves from .047 μA to 1.23 $\times 10^{-5}\mu$A and the on current degrades from 25.2 μA to 11.3 μA when the gate length changes from 5 nm to 10 nm. This change in gate length improves the on/off current ratio by

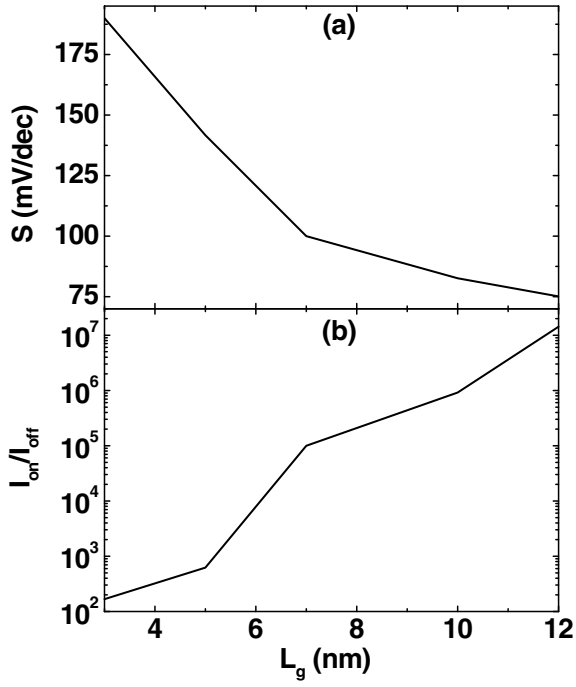


Fig. 3. Inverse subthreshold slope and on/off current ratio versus gate length.

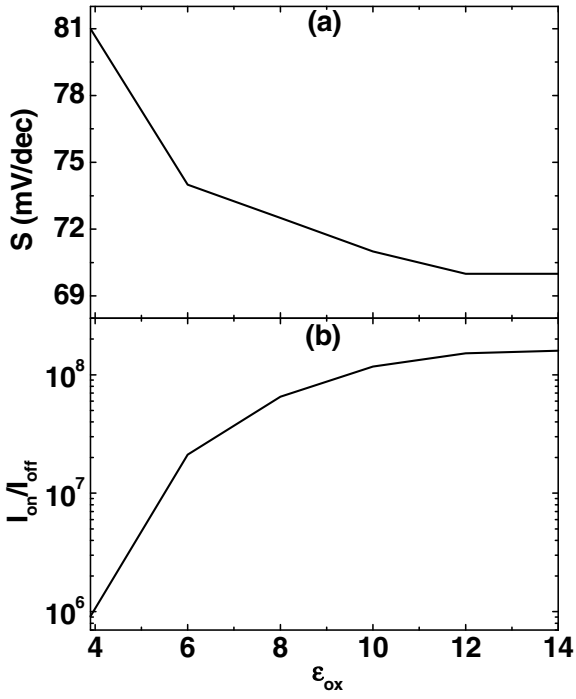


Fig. 4. Inverse subthreshold slope and on/off current ratio versus gate dielectric constant ϵ_{ox}.

three orders of magnitude and the inverse subthreshold slope from 141.68 mV/dec to 82.59 mV/dec.

Next we study the effects of gate dielectric constant and thickness on the inverse subthreshold slope and on/off current ratio. Simulation results of inverse subthreshold slope and on/off current ratio as a function of gate dielectric constant ϵ_{ox} are shown in Fig. 4. Both the inverse subthreshold slope and the on/off current ratio improve with high-K gate dielectric. The change of gate dielectric constant from 3.9 to 14 improves the on/off current ratio by two orders of magnitude and the inverse subthreshold slope from 81 mV/dec to 70 mV/dec. The physics can be understood from the band diagram shown in Fig. 5 for two different gate dielectric constant 3.9 and 14. Note that only the conduction bands are shown because the tunneling current through the valence band is zero. From the band profiles we see that the gate control is much better with high-K gate dielectric in off state. The channel potential barrier is almost equal to $E_g/2e$ that reduces the tunneling plus thermal current significantly. On the other hand, the modulation of band profile in on state with gate dielectric constant is insignificant, and therefore, the change in on current with ϵ_{ox} is very little. The off current improves from 1.23 $\times 10^{-5}\mu$A to 5.65 $\times 10^{-8}\mu$A and the on current degrades from 11.3 μA to 9.0 μA when the ϵ_{ox} changes from 3.9 to 14.

The scaling of inverse subthreshold slope and on/off current ratio with t_{ox} is shown in Fig. 6. The devices used for simulation have gate length L_g of 10 nm, gate dielectric constant ϵ_{ox} of 3.9, and doping concentration of 1.65 $\times 10^{19}$ cm^{-3}. Both the inverse subthreshold slope and the on/off

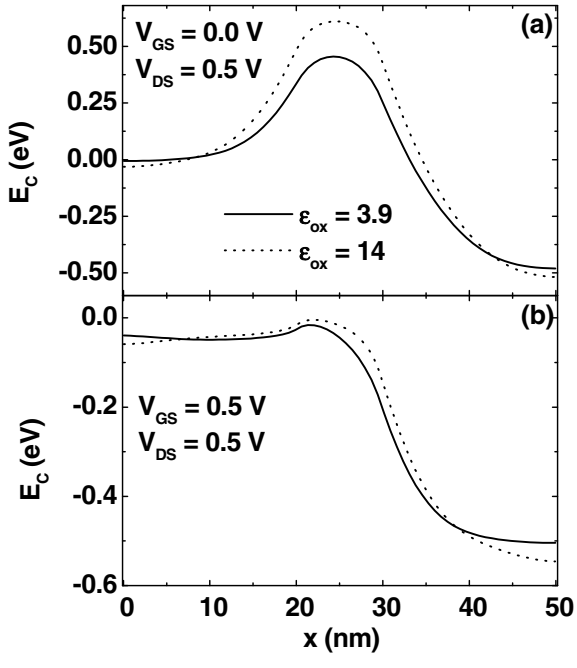


Fig. 5. Conduction band profile for two dielectric constant in (a) off state (b) on state.

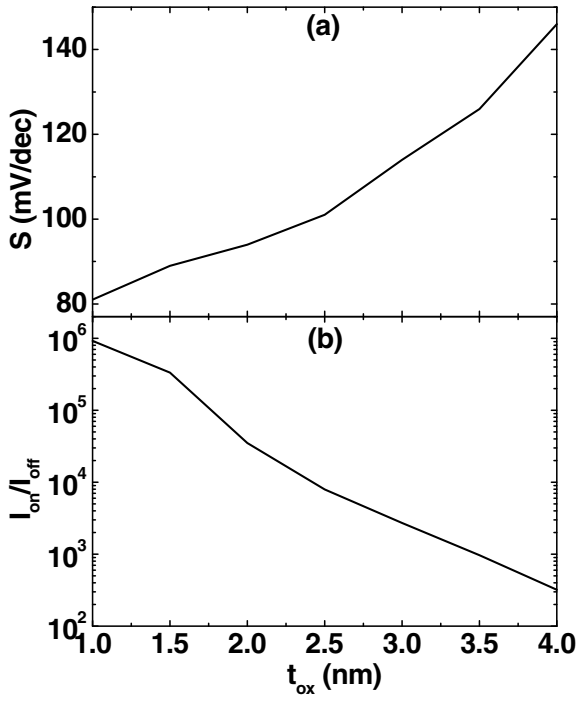


Fig. 6. Inverse subthreshold slope and on/off current ratio versus gate dielectric thickness t_{ox}.

current ratio significantly improve with thinner oxide due to better gate control with thin gate oxide. For a change of t_{ox} from 4 nm to 1 nm, the on/off current ratio improves more than three orders of magnitude and the inverse subthreshold slope improves from 146 to 81.

IV. Conclusion

A three dimensional quantum simulation is performed to study the effects of doping, gate length, gate dielectric constant, and thickness on inverse subthreshold slope and on/off current ratio of a top gate silicon nanowire on insulator transistor. The device performance in terms of inverse subthreshold slope and on/off current ratio improves with thinner high-K gate dielectric and relatively longer gate length.

References

[1] Y. Wu, Y. Cui, L. Huynh, C. J. Barrelet, D. C. Bell, and C. M. Lieber, "Controlled growth and structures of molecular-scale silicon nanowires," Nano Lett., vol. 4, no. 3, pp. 433–436, 2004.

[2] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, "High performance silicon nanowire field effect transistors," Nano Lett., vol. 3, no. 2, pp. 149–152, 2003.

[3] H. C. Lin and C. J. Su, "High performance poly-Si nanowire NMOS transistors," IEEE Trans. Nanotechnol., vol. 6, no. 2, pp. 206–212, 2007.

[4] S. M. Koo, M. D. Edelstein, Q. Li, C. A. Richter, and E. M. Vogel, "Silicon nanowires as enhancement-mode Schottky barrier field-effect transistors," Nanotechnology, vol. 16, pp. 1482–1485, 2005.

[5] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, "Performance evaluation of ballistic silicon nanowire transistors with atomic-basis dispersion relations," Appl. Phys. Lett., vol. 86, p. 093113, 2005.

[6] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim, and C. M. Lieber, "Logic gates and computation from assembled nanowire building blocks," Science, vol. 294, pp. 1313–1317, 2001.

[7] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, "Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species," Science, vol. 293, pp. 1289–1292, 2001.

[8] M. Shin, "Quantum simulation of device characteristics of silicon nanowire FETs," IEEE Transections on Nanotechnology, vol. 6, pp. 230–237, 2007.

[9] ——, "Efficient simulation of silicon nanowire field effect transistors and their scaling behavior," J. Appl. Phys., vol. 101, no. 2, p. 024510, 2007.

[10] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," J. Appl. Phys., vol. 81, no. 12, pp. 7845–7869, 1997.

[11] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, "Highly convergent schemes for the calculation of bulk and surface green functions," J. Phys. F, vol. 15, pp. 851–858, 1985.

[12] V. Eyert, "A comparative study on methods for convergence acceleration of iterative vector sequences," J. Comput. Phys., vol. 124, no. 0059, pp. 271–285, 1996.

[13] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, "On the validity of the parabolic effective-mass approximation for the i-v calculation of silicon nanowire transistors," IEEE Trans. Electron Dev., vol. 52, no. 7, pp. 1589–1595, 2005.

[14] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck, "Electronic properties of silicon nanowires," IEEE Trans. Electron Dev., vol. 52, no. 6, pp. 1097–1103, 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effects of uniaxial strain on the bandstructures of silicon nanowires

Redwan Noor Sajjad[1], Khairul Alam[2] and Quazi Deen Mohd Khosru[1]
Department of Electrical and Electronic Engineering
[1]Bangladesh University of Engineering and Technology Dhaka-1000
[2]East West University, Dhaka-1212
Email: redwan@eee.buet.ac.bd

***Abstract*— The effects of uniaxial strain on the band structures of <100> silicon nanowires of width 2.75 - 3.84 nm are studied using $sp^3d^5s^*$ orbital basis atomistic tight binding approach. The conduction band edge at Γ point has almost no variation with strain and the second valley located at 0.36$\times\pi$/a of the wire Brilluoin moves down in energy with both compressive and tensile strains. The top valence band moves up in energy with both tensile and compressive strain, and therefore, the band gap reduces with both types of strain. We notice about 7% change in band gap for an application of 2% strain. The electron effective masses at Δ_4 and Δ_2 valleys show opposite dependence on strain, and the hole effective mass of top valence band has almost similar variation with both types of strain. We notice a significant change in hole effective mass with strain.**

I. Introduction

One dimensional materials such as nanowires can be attractive building blocks for future nanotechnology because their electronic properties can be precisely controlled and they are compatible with the CMOS processes. The controlled growth of silicon nanowires, their applications as field effect transistors and logic circuits have been demonstrated experimentally [1]–[4]. In parallel with the experimental work there has been good progress in simulation [5]–[7].

Transistor scaling down to nanometer regime degrades device performance due to severe short channel effects and mobility degradation. Application of strain has been proposed to enhance mobility and has been studied extensively over the last few decades [8]–[14]. Strain was first proposed in 1950 and since then strain engineering has been one of the most crucial technological innovations to improve device performance. In 1992, the strain silicon was first used as the MOSFET channel and 70% improvement in mobility, compared to the unstrained counterpart, was observed [15], [16]. Mobility enhancement increases drain current.

Application of strain changes the atomic coordinates from their equilibrium positions. This results in changes in the crystal structures and hence in the band structures. Strain effects on bulk silicon band structures have been studied using both k.p theory [17], [18] and atomistic tight binding approach [14]. Strain splits the six fold degeneracy of conduction band minima, reduces effective mass and inter-valley scattering, and enhances mobility [10], [19]. From experimental viewpoint, strain can be applied by introducing lattice mismatch [16], [20], [21], by bending after completion of the device processes [22], [23], or by rapid thermal process [20], [24].

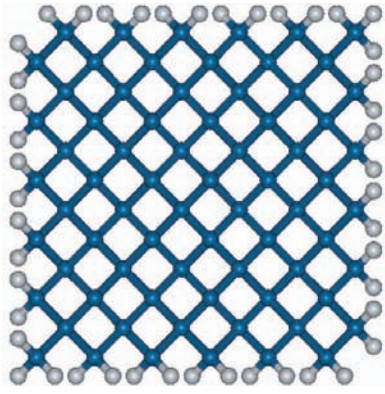


Fig. 1. Cross-section of <100> SiNW.

Strain effects on the band gaps of silicon nano clusters have been studied [25]–[27]. However, a comprehensive simulation of strained silicon nanowire is yet to explore. In this paper, we study the uniaxial strain effects on silicon nanowire band structure using $sp^3d^5s^*$ empirical tight binding approach. The nanowire studied is grown in <100> direction and the strain is applied in that direction. Nanowires of three different cross sections are used for simulation. The band gap is found to reduce with both tensile and compressive strain. This is because strain effect on conduction band minimum is negligible and the valence band moves up with strain. Variation of the electron effective masses of $\triangle_4$ and $\triangle_2$ valleys has opposite dependency on strain. The top valence band effective mass reduces with strain (note that the effective mass of valence band is negative).

II. Simulation Model

The SiNWs used in this study are grown in <100> directions using bulk bond length. During nanowire growth, the bulk bond length is assumed and the dangling bonds are passivated with hydrogen atoms. The cross section of <100> SiNW is shown in Fig. 1. The blue atoms are silicon and the gray atoms are hydrogen. The nanowire growth direction is x, which is into (or out of) the paper. The y and z directions

are <010> and <001>, respectively. The unit cell is 0.543 nm long and consists of 4 atomic layers. For band structure calculation without strain effects, the Hamiltonian is created as

$$H(k_x) = H_0(k_x) + t_{01}e^{ikx\triangle x} + t_{10}e^{-ikx\triangle x} \quad (1)$$

Here k_x is the one dimensional (1D) wave vector and $\triangle x$ is the distance between the last layer of a unit cell and the first layer of the next unit cell. The matrix elements of $H_0(k)$ are created from

$$H_0(i,j) =< \phi_{i,n}|H|\phi_{j,m} > e^{ik_x(x_m - x_n)} \quad (2)$$

and those of t_{01} from

$$t_{01}(p,q) =< \phi_{p,u}|H|\phi_{q,u} > \quad (3)$$

Here n and m label the atoms in a unit cell, and u and v label the atoms between adjacent unit cells. The basis ϕ, is the $sp^3d^5s^*$ atomic orbitals and $\phi_{j,m}$ is the j^{th} orbital of the m^{th} atom. The energy integral expressions are taken from Slater [28], and the energy integral values are taken from Boykin [29] and Zheng [30]. The band structure is obtain by calculating the eigen energies of H_{kx}.

The strain tensor is diagonal for uniaxial strain as shown in the following equation

$$\epsilon = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_2 \end{pmatrix}, \quad (4)$$

where the applied strain ϵ_1 is varied from -2% to 2%, and ϵ_2 is obtained from Poisson's ratio. The atomic coordinates of the nanowire are generated first without strain applied. Then the atomic position with strain applied is calculated from $\mathbf{r}' = (\mathbf{1} + \epsilon)\mathbf{r}$ [17], where $\mathbf{r}$ is the atomic coordinates without strain and $\mathbf{r}'$ is the atomic coordinates after strain is applied. The on-site energies are assumed have their bulk values after strain is applied, and the two-center integrals are modified according to Harrison's formula [31] $U = U_0(d_0/d)^2$, where U_0 is the bulk value of two-center integral and d_0 and d are the bulk and strain bond lengths, respectively.

III. Numerical Results and Discussions

The simulation is performed for nanowires grown in <100> direction. The band structures of 2.75nm×2.75nm wire are shown in Fig. 2. Fig. 2(a) is the band structures with no strain applied and (b) with 2% compressive strain. The bulk silicon is an indirect band gap material having conduction band minimum at 0.832×2π/a in the Δ direction. It has six equivalent Δ valleys. The nanowire is a direct band gap material. The four bulk valleys are projected at Γ point in the one dimensional Brilluoin zone of nanowire to form the conduction band minimum. The other two valleys are zone folded to 0.36×π/a in the wire Brilluoin zone. Compressive strain has almost no effects on the conduction band edge and the top valence band moves up in energy with strain. The second conduction band minimum located at 0.36×π/a of the wire Brilluoin zone moves down in energy with compressive strain.

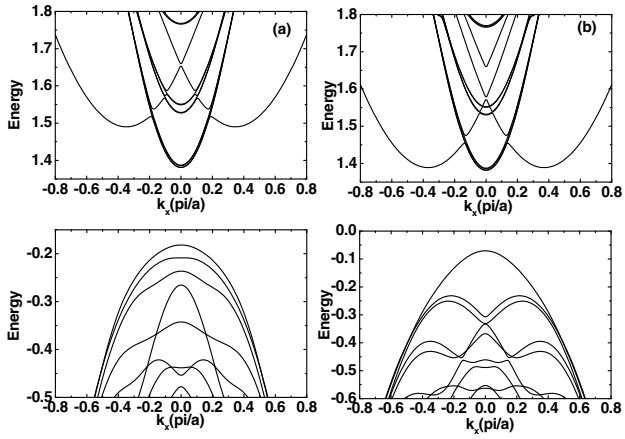


Fig. 2. The band structure a 2.75nm×2.75nm SiNW. (a) With no strain applied. (b) With 2% compressive strain.

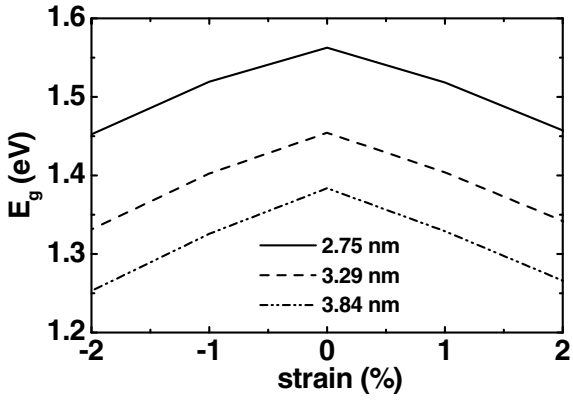


Fig. 3. Band gap variation with strain.

The results for tensile strain are not shown, but the tensile strain has the same effects on band structures. Therefore, the direct band gap will reduce with both tensile and compressive strain. This is in fact the case. The band gap variation with strain is shown in Fig. 3. The variations are shown for wires of three different cross sections. All the wires have similar band gap variation with strain, and the variation is symmetric in terms of tensile and compressive strains. The change in band gap is approximately 7% when 2% strain is applied.

The variation of conduction and valence band edges are shown in Figs. 4 and 5. The strain has almost no effects on conduction band edge and the valence band edge varies almost linearly with both tensile and compressive strain. The band gap variation and the valence band edge variation with strain are similar. The electron effective masses at Δ_4 and Δ_2 valleys and the hole effective mass of the top valence band are shown in Figs. 6, 7, and 8. The electron masses of Δ_4 and Δ_2 valleys show opposite dependence on strain. The Δ_4 mass reduces with tensile strain and increases with compressive strain, while the Δ_2 mass increases with tensile strain and reduces with compressive strain. Also the strain effect is relatively higher on wires of larger cross section for Δ_4 masses. This effect is opposite to the Δ_2 masses. The

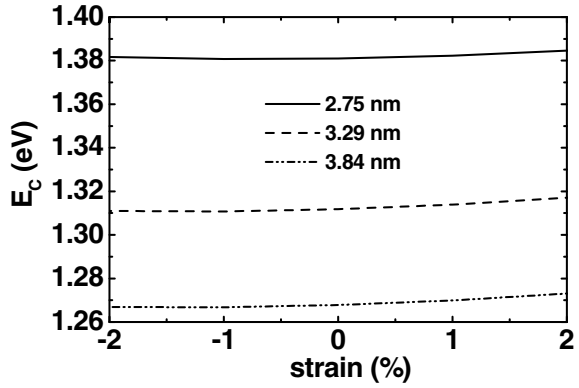


Fig. 4. Variation of conduction band edge with strain.

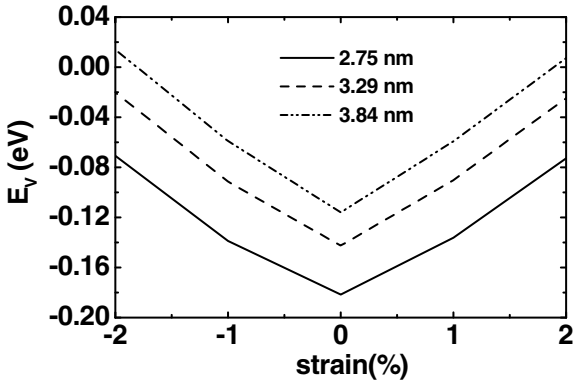


Fig. 5. Variation of valence band edge with strain.

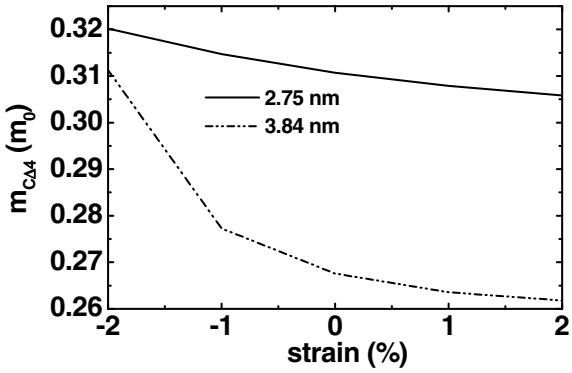


Fig. 6. Electron effective mass of Δ_4 valley versus strain.

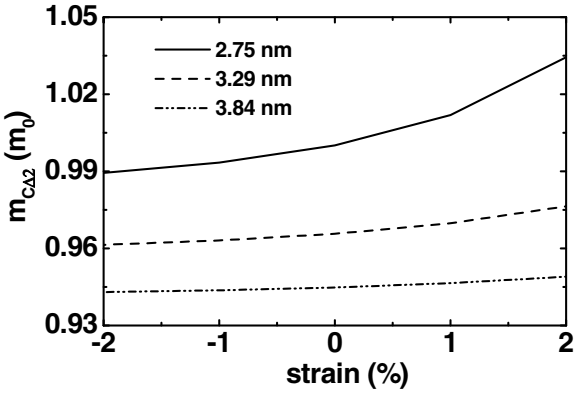


Fig. 7. Electron effective mass of Δ_2 valley versus strain.

hole effective mass of top valence band reduces (note that hole effective mass is negative) with strain and its variation with compressive and tensile strain is very similar.

IV. Conclusion

The effects of strain on band structure of <100> silicon nanowires are studied. The Δ_4 valley remains almost unaltered with strain, however, the Δ_2 valley changes significantly with strain. The direct band gap variation is a consequence of the effects of strain on the valence band. The Δ_4 and Δ_2 valley effective masses have opposite dependence on strain. The hole effective mass of the top valence band has significant change with strain.

References

[1] Y. Wu, Y. Cui, L. Huynh, C. J. Barrelet, D. C. Bell, and C. M. Lieber, "Controlled growth and structures of molecular-scale silicon nanowires," Nano Lett., vol. 4, no. 3, pp. 433–436, 2004.

[2] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, "High performance silicon nanowire field effect transistors," Nano Lett., vol. 3, no. 2, pp. 149–152, 2003.

[3] S. M. Koo, M. D. Edelstein, Q. Li, C. A. Richter, and E. M. Vogel, "Silicon nanowires as enhancement-mode Schottky barrier field-effect transistors," Nanotechnology, vol. 16, pp. 1482–1485, 2005.

[4] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim, and C. M. Lieber, "Logic gates and computation from assembled nanowire building blocks," Science, vol. 294, pp. 1313–1317, 2001.

[5] H. C. Lin and C. J. Su, "High performance poly-Si nanowire NMOS transistors," IEEE Trans. Nanotechnol., vol. 6, no. 2, pp. 206–212, 2007.

[6] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, "Performance evaluation of ballistic silicon nanowire transistors with atomic-basis dispersion relations," Appl. Phys. Lett., vol. 86, p. 093113, 2005.

[7] J. Wang, E. Polizzi, and M. Lundstrom, "A three-dimensional quantum simulation of silicon nanowire transistors with the effective-mass approximation," J. Appl. Phys., vol. 96, no. 4, pp. 2192–2203, 2004.

[8] H. H. Hall, J. Bardeen, and G. L. Pearson, "The effects of pressure and temperature on the resistance of p - n junctions in germanium," Phys. Rev., vol. 84, pp. 129–132, 1951.

[9] C. S. Smith, "Piezoresistance effect in germanium and silicon," Phys. Rev., vol. 94, pp. 42–49, 1954.

[10] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr, "The effect of general strain on the band structure and electron mobility of silicon," IEEE Trans. Electron Dev., vol. 54, no. 9, pp. 2183–2190, 2007.

[11] I. Lauer and D. A. Antoniadis, "Enhancement of Electron Mobility in Ultrathin-Body Silicon-on-Insulator MOSFETs With Uniaxial Strain," IEEE Electron Dev. Lett., vol. 26, pp. 314-316, 2005.

[12] Y. Tan, X. Li, L. Tian, and Z. Yu, "Analytical electron-mobility model for arbitrarily stressed silicon," IEEE Trans. Electron Devices, vol. 55, pp. 1386–1390, 2008.

[13] E. Fitzgerald, Y. Xie, M. Green, D. Brasen, A. Kortan, J. Michel, Y. Mii, and B.Weir, "Totally relaxed gexsi1-x layers with low threading dislocation densities grown on si substrates," J. Appl. Phys., vol. 59, no. 7, pp. 811–813, 1991.

[14] A. Khakifirooz and D. A. Antoniadis, "Scalability of hole mobility enhancement in biaxially strained ultrathin body soi," IEEE Electron Dev. Lett., vol. 27, no. 5, pp. 402-404, 2006.

[15] J. Welser, J. Hoyt, and J. Gibbons, "Electron mobility enhancement in strained-Si n-type metaloxidesemiconductor field-effect transistors," IEEE Electron Dev. Lett., vol. 15, pp. 100-102, 1994.

[16] ——, "NMOS and PMOS transistors fabricated in strained-

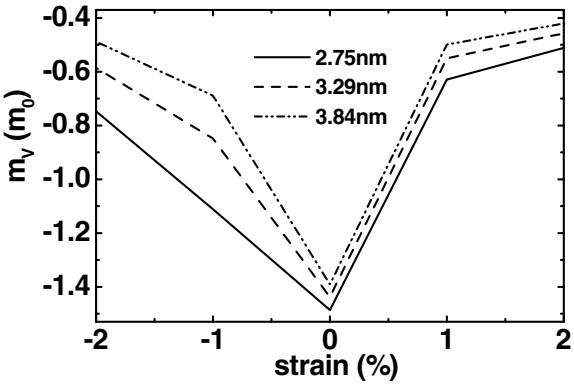


Fig. 8. Variation of hole effective mass of the top valence band with strain.

silicon/relaxed silicon-germanium structures," in Proceedings of the IEDM. New York: IEEE, 1992, pp. 1000-1002.

[17] D. Rideau, M. Feraille, L. Ciampolini, M. Minondo, C. Tavernier, H. Jaouen, and A. Ghetti, "Strained Si, Ge, and $Si_{1-x}Ge_x$ alloys modeled with a first-principles-optimized full-zone k·p method," Phys. Rev. B, vol. 74, p. 195208, 2006.

[18] L. Shifren, X. Wang, P. Matagne, B. Obradovic, C. Auth, S. Cea, T. Ghani, J. He, T. Hoffman, R. Kotlyar, Z. Ma, K. Mistry, R. Nagisetty, R. Shaheed, M. Stettler, C. Weber, and M. D. Giles, "Driving current enhancement in p-type metal-oxide-semiconductor field-effect transistors under shear uniaxial stress," Appl. Phys. Lett., vol. 85, p. 61886190, 2004.

[19] S. Takagi, J. L. Hoyt, J. J.Welser, and J. F. Gibbons, "Comparative study of phononlimited mobility of two-dimensional electrons in strained and unstrained-si metal-oxide-semiconductor field-effect transistors," J. Appl. Phys., vol. 80. p. 1567, 1996.

[20] K. Rim, J.-L. Hoyt, and J.-F. Gibbons, "Transconductance enhancement in deep submicron strained si n-mosfets," in Proceedings of the IEDM. New York: IEEE, 1998, pp. 707-710.

[21] J.-L. H. et al., "Strained-silicon mosfet technology," in Proceedings of the IEDM. New York: IEEE, 2002, pp. 23-26.

[22] S. Maikap, C. Yu, M. L. S. Jan, and C. Liu, "Mechanically strained Si NMOSFETs," IEEE Electron Dev. Lett., vol. 25, no. 1, pp. 40-42, 2004.

[23] A. Lochtefeld and D. Antoniadis, "Investigating the relationship between electron mobility and velocity in deeply scaled nmos via mechanical stress," IEEE Electron Dev. Lett., vol. 22, no. 12, p. 591593, 2001.

[24] A. S. et al., "Local mechanical-stress control (lmc) : A new technique for CMOS performance enhancement," in Proceedings of the IEDM. New York: IEEE, 2001, p. 433436.

[25] X.-H. Peng, A. Alizadeh, N. Bhate, K. K. Varanasi, S. K. Kumar, and S. K. Nayak, "First-principles investigation of strain effects on the energy gaps in silicon nanoclusters," Journal of Physics: Condensed Matter, vol. 19, p. 266212, 2007.

[26] G. Klimeck, F. Oyafuso, T. Boykin, R. Bowen, and P. von Allmen, "Development of a nanoelectronic 3-D (nemo 3-D) simulator for multi-million atom simulations and its application to alloyed quantum dots," Computer Modeling in Engineering and Science, vol. 3, p. 601, 2002.

[27] T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, "Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory," Phys. Rev. B, vol. 66, p. 125207, 2002.

[28] J. C. Slater and G. F. Koster, "Simplified LCAO method for the periodic potential problem," Phys. Rev., vol. 94, no. 6, pp. 1498–1524, 1954.

[29] T. B. Boykin, G. Klimeck, and F. Oyafuso, "Valence band effective-mass expressions in the $sp^3d^5s^*$ empirical tight-binding model applied to a Si and Ge parametrization," Phys. Rev. B, vol. 69, pp. 115 201/1–10, 2004.

[30] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck, "Electronic properties of silicon nanowires," IEEE Trans. Electron Dev., vol. 52, no. 6, pp. 1097–1103, 2005.

[31] W. A. Harrison, Electronic structure and properties of solids. San Francisco: Freeman, 1980.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Highly Oriented Carbon Nanotubes by Chemical Vapor Deposition

Sharif M. Mominuzzaman[1], Ishwor. Khatri[2], Zhang Jianhui[2], Tetsuo Soga[2] and Takashi Jimbo[2]

[1]Department of Electrical an Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh.
[2]Department of Environmental Technology and Urban Planning, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
E-mail: momin@eee.buet.ac.bd

Abstract - In the present work we have synthesized low density aligned carbon nanotubes using ethanol as a precursor and ferrocene as a catalyst by simple thermal chemical vapor deposition (CVD) method on silicon substrates. The synthesis involves the pyrolysis of ethanol/ ferrocene solutions. Mist of the hydrocarbon/catalyst is produced by medical nebulizer and nitrogen gas is used as the carrier to the large size (75mm diameter and 1metre long) tube which is designed to produce nanotubes in large quantity once growth process is optimized for particular application. The growth temperature of the chamber is varied in wide range. In the synthesis process, temperature is found to be crucial for the formation of carbon nanotubes. Different types (in terms of size and shape) of nanotubes are grown. at 800°C Short and closely uniform sized nanotubes are formed at 850°C. Furthermore, very interestingly we obtained low density vertical aligned nanotubes which has potential application in various electronic devices such as, field emitter and conducting electrodes, etc. Our effort to reproduce the straight uniformed nanotube along with the low density vertically aligned nanotubes was successful in spray pyrolysis method. Long nanotubes of different length and diameters are observed at 900°C. The tubes are observed to be temperature dependent. TEM investigations of these nanotubes also will be reported.

I. Introduction

The Buckminsterfullerene, i.e., C_{60}, a novel prize winning molecule was discovered in 1985 during experiments carried out by laser vaporization of carbon species from the solid graphite in order to understand the mechanisms by which long-chain carbon molecules are formed in interstellar space and circumsteller shells. As this discovery by Kroto *et al.* [1] created intense interest among scientists community helical microtubules of graphitic carbon, i.e., carbon nanotubes are observed by Sumio Iijima [2] during the experiments using apparatus similar to that used by [3](for mass production of C60. The form of nanotube synthesized this time was multiwalled. Two years later in 1993, with the discovery of single walled nanotubes[4], a new field related to nanotechnology has emerged due to the special properties of this nano dimensional material. In fact, a carbon nanotube has a cylindrical structure having a diameter of few nano-meters and a very large aspect ratio, whereas, carbon atoms are generally arranged in a hexagonal honeycomb network. Depending on the structures, the carbon nanotube may show metallic or semiconducting. For metallic nanotube, the conductivity can be very high compared to typical conductors. Further, carbon nanotube has very special mechanical properties. It has extra high thermal conductivity with superior mechanical strength and tera level of Young's modulus. As exceptional electrical, electronic, optical and mechanical properties are observed carbon nanotubes already revealed promising applications in various diversified field of science and technology such as nanoelectronics, advanced mechanical and biomedical applications. Carbon nanotubes (NT) are among the most promising materials for various areas of science and technology such as nanoelectronics, advanced composites and others [5].

Carbon nanotube can be synthesized by number of techniques such as, laser deposition, arc-discharge, chemical vapor deposition (CVD). The CVD growth is suitable for preparing the carbon nanotube both bulk and on substrate. The arc discharge and the laser deposition have relatively low yields of carbon nanotubes. When using the arc discharge and the laser deposition, it is difficult to control the diameter and the length of the carbon nanotube. Further, in the arc discharge and the laser deposition, clusters of amorphous carbon besides the carbon nanotubes are produced in a large amount, and thus a complicated purifying process must be performed.

From the technological aspect, it is really a difficult and challenging task to grow CNT on a substrate. Chemical vapor deposition (CVD) methods, such as thermal chemical vapor deposition, low pressure chemical vapor deposition, plasma enhanced chemical vapor deposition (PECVD) and catalytic chemical vapor deposition (CCVD) are generally used to form carbon nanotubes on a substrate and can be grown much lower temperature than laser ablation and arc-discharge processes. Furthermore, ability of large-scale production at lower cost makes the CCVD method attractive to the scientific community; therefore, many research groups in the whole world are using CCVD method to grow CNTs.

In the case of catalytic chemical vapor deposition (CCVD) methods, metal particles are observed to be essential for the formation of nanotubes by the decomposition of

carbon feedstock, such as hydrocarbons. Transition metals such as iron, cobalt, nickel are found to be suitable catalyst for the CNT growth. In some cases, alloying them with each other or with other non-transition metals dramatically changes their catalytic activities. Possibly the role of the non-transition metal is to disperse and to stabilize the catalyst metal. The size of the metal particles plays a crucial role in determining the nature of the nanotubes grown. One way of producing catalyst nanoparticles in the pyrolysis zone is by spraying a colloidal or true solution of a hydrocarbon and metal salt (precatalyst) that decompose at the synthesis temperature [6]. In other words, active catalyst nanoparticles are formed directly in the pyrolysis zone and participate in NT nucleation and growth, chemisorbing the carbon that appears from hydrocarbon decomposition. Ultrasonic spraying is easy to control and it offers a high efficiency. We are working on various nanostuctured carbon [7-10] for its electronic application. In the present work we have synthesized low density aligned carbon nanotubes using ethanol as a precursor and ferrocene as a catalyst by simple thermal catalytic chemical vapor deposition (CCVD) method on silicon substrates and similar results are not reported so far by earlier observations. The CNTs are characterized by scanning electron microscopy (SEM) and Transmission electron microscopy (TEM) analyses.

II. Experimental Details

The synthesis involves the pyrolysis of ethanol/ ferrocene solutions. Mist of the hydrocarbon/catalyst is produced by medical nebulizer which is connected to big size quartz tube (75mm of internal diameter and 1000mm in length), designed to produce nanotubes in large quantity. Nitrogen (N) is used as carrier gas to generate the ethanol/ferrocene mist in the nebulizer. The schematic of the spray pyrolysis setup is illustrated in Fig. 1. The quartz tube was fitted into a 700mm long cylindrical furnace. One end of the quartz tube was attached to the nebulizer and other to the gas bubbler. The quartz tubing was heated by a furnace (Toyo), equipped with a very precise temperature controller (±1°C).

Single-crystal silicon (100) substrates were cleaned twice in acetone for 5 minuets followed by cleaning in methanol and de-ionized water by ultrasonicator. The substrate was dried by nitrogen blower before putting in the quartz boat and placed in the centre of the 1000mm long quartz tube. The nebulizer container was filled with 50mL of the mixture of 2 wt% ferrocene in ethanol. To make uniform solution, ferrocene and ethanol was mixed in ultrasonic vibrator for 5 minuets, earlier. Nitrogen was used as carrier gas and flow was regulated my mass flow controller. To flush air/oxygen, nitrogen flow (1L/min) is maintained as the temperature of the furnace is increased. 30 minuets was allowed to reach the set temperature (600-950°C) and kept 10 minuets at the stable temperature. The nebulizer was switched on and nitrogen flow is increased to 1.5L/min and ferrocene/ethanol mist is allowed to enter in the quartz tube. The spraying time lasted about 15 minuets. Once the solution is exhausted, the N_2 flow rate is reduced to 0.5L/min and maintained until the furnace temperature is cooled down to below 100°C.

The analyses of the obtained CNTS were made by SEM (Hitachi S-3000H, scanning electron microscopy) and Transmission electron microscopy (TEM) observations.

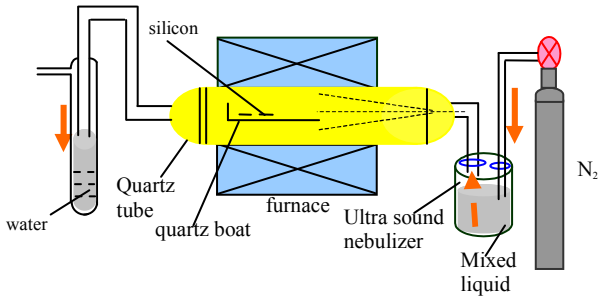


Fig. 1. The experimental set-up for CNT synthesis.

III. Results and Discussions

The spray pyrolysis experiments carried out in the temperature range of 600 to 925°C. Below 800°C, only small amount of carbonaceous material, such as, amorphous carbon was detectable. At 800°C, the initiation of nanotube growth is observed to appear. The lengths of the nanotubes are very short — limited to only few microns. Diameter and length of the nanotubes are very random. Possibly at this temperature the catalyst particles are not uniform. Therefore, though the nanotubes grown has initiated to some extend but with widely varying dimensions. Furthermore, there might be lot of defects/dangling bonds present in the graphene sheet which restricts its natural growth mechanism and limits the nanotube length. Also due to the presence of defects/disorders the diameter of the nanotubes are not uniform even along the same nanotube. SEM image of sample grown at 800°C is shown in figure 2. The CNTs synthesized at 850 °C are completely different to the CNTs grown at 800 °C. The nanotubes are vertically grown. They are well aligned. The CNTs are also uniform in terms of diameter and length. Possibly defects such as, dangling bonds etc. are reduced at this temperature, therefore, the tubes are straight and the diameter remain almost same along the tube. This kind of short and straight nanotubes is promising for various applications. The nanotubes grown at 850°C are shown in Figs. 3(a)-(c).

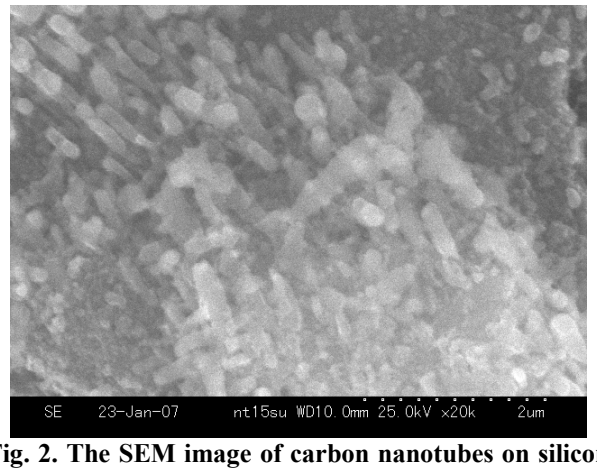
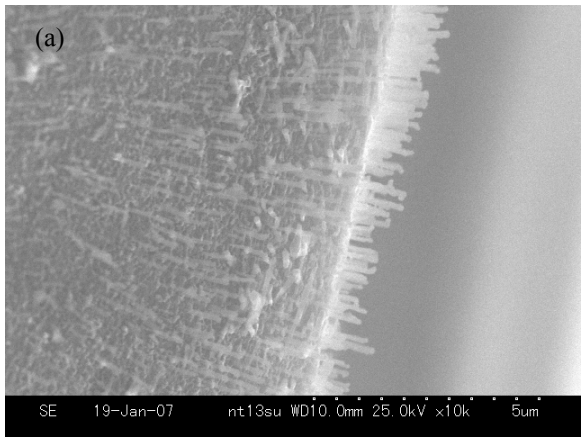
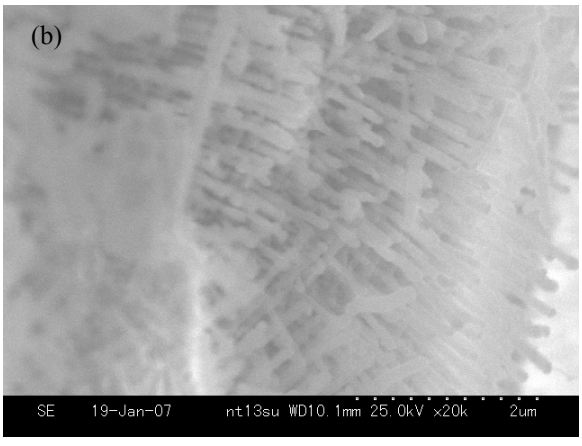


Fig. 2. The SEM image of carbon nanotubes on silicon substrate synthesized at 800 °C.





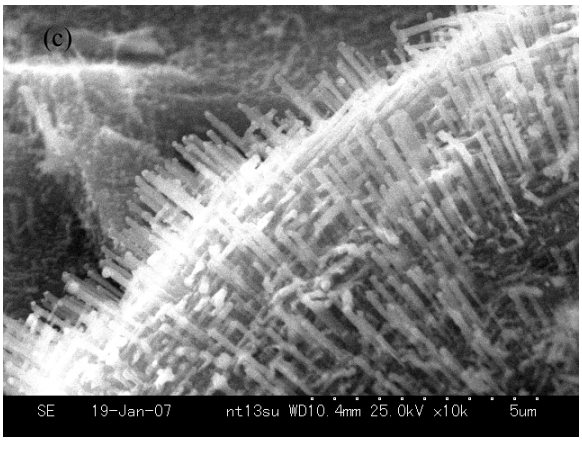


Fig. 3. The SEM image of carbon nanotubes on silicon substrate synthesized at 850 °C.

The diameter of the grown tubes depends strongly on the growth temperature; a decrease of average diameter with increasing temperature was found at 900 °C. The nanotubes become very long and curly in morphology (Fig. 4). With further increase of temperature, the nanotubes diameter has increased and length has decreased. Fig. 5 shows the SEM image of the nanotube grown at 925 °C.

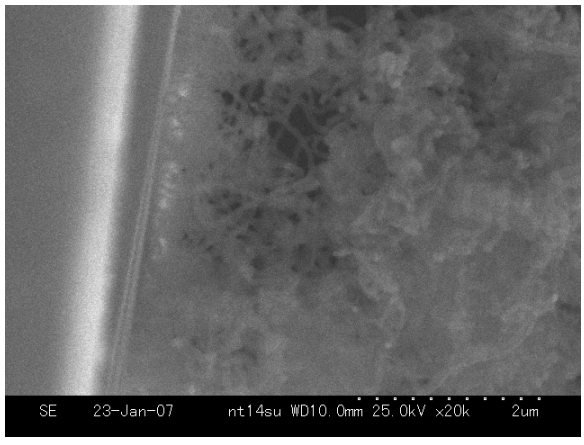


Fig. 4. The SEM image of carbon nanotubes on silicon substrate synthesized at 900 °C.

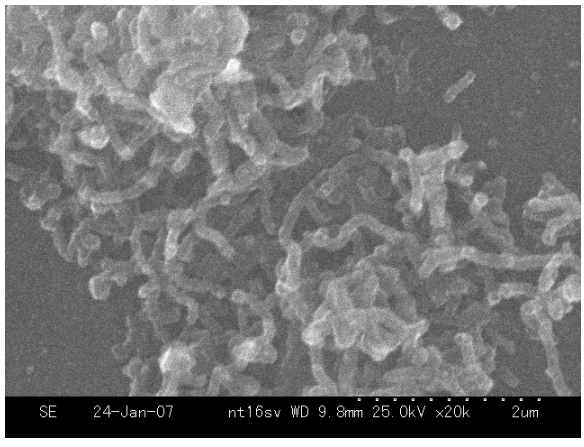


Fig. 5. The SEM image of carbon nanotubes on silicon substrate synthesized at 925 °C.

In the synthesis process, temperature is found to be crucial for the formation of carbon nanotubes. The nanotube grown at 850 is very different and interesting one. The nanotubes are also different from reported earlier [11]. The nanotubes are highly oriented in vertical direction, therefore, can be used in field emission devices. Recent work has demonstrated advances in using CNT films in wide variety of applications as transparent conductor [12]. Compatibility of CNT films as the transparent electrode using PEDOT:PSS/P3HT:PCBM structure is reported[13]. The vertical nanotube synthesized in the present work is expected to be a good electrical conductor that can be used on highly doped Si and can be used as base to fabricate efficient polymer-fullerene bulk heterojunction organic solar cell with improved performance due to its better collection efficiency. Furthermore, these vertical CNTs might serve as super cooling purpose in nanotechnology due to its extraordinary high thermal conductivity. electrode on highly doped Si substrate. As the CNTs grown at 850°C have scope in various fields, TEM observations are done in order to investigate an insight into the nanotube structures. TEM analysis reveals that these nanotubes are straight and surface is smooth. The TEM image is shown in Fig. 6.

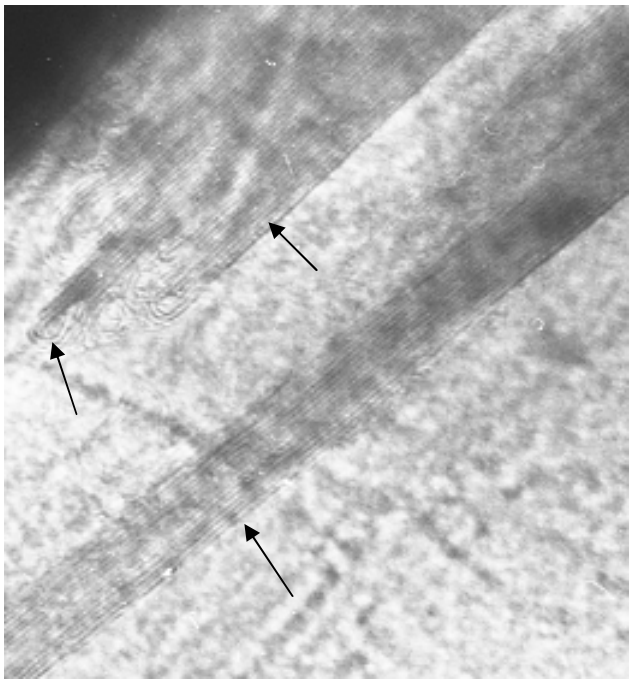


Fig. 6 The TEM image of carbon nanotubes on silicon substrate synthesized at 850 °C. The arrows shown indicate the straight tube and smooth surface.

IV. Conclusions

In the present work we have synthesized low density aligned carbon nanotubes using ethanol as a precursor and ferrocene as a catalyst by simple thermal chemical vapor deposition (CVD) method on silicon substrates. The synthesis involves the pyrolysis of ethanol/ ferrocene solutions. Mist of the hydrocarbon/catalyst is produced by medical nebulizer and nitrogen gas is used as the carrier to the large size (75mm diameter and 1metre long) tube which is designed to produce nanotubes in large quantity once growth process is optimized for particular application. The growth temperature of the chamber is varied in wide range. The tubes are observed to be temperature dependent. Different types (in terms of size and shape) of nanotubes are grown at 800°C Short and closely uniform sized nanotubes are formed at 850°C[Fig. 2]. Furthermore, very interestingly we obtained low density vertical aligned nanotubes in the edge of the substrates [Figs. 3(a) and (b)] which has potential application in various electronic devices such as, field emitter. Our effort to reproduce the straight uniformed nanotube along with the low density vertically aligned nanotubes in the edge was successful and similar results are not reported so far by earlier observations in spray pyrolysis method [3-5]. Long nanotubes of different length and diameters are observed at 900°C as shown in Fig. 4. Raman spectral analyses and TEM investigations of these nanotubes also will be reported for the first time.

V. Acknowledgements

One of the authors (SMM) is grateful to Bangladesh University of Engineering and Technology for granting leave to carry out the research. This work was partly supported by JSPS Research for Future Program.

References

[1] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley, "C60: Buckminsterfullerene", Nature, vol. 318, pp. 162-163, 1985.

[2] Sumio Iijima,"Helical microtubules of graphitic carbon", Nature, vol. 354, pp. 56-58, 1991.

[3] W. Krätschmer, Lowell D. Lamb, K. Fostiropoulos and Donald R. Huffman, "Solid C_{60}: a new form of carbon," Nature, vol. 347, pp. 354–358, 1990.

[4] Sumio Iijima and Toshinari Ichihashi, "Single-shell carbon nanotubes of 1-nm diameter", Nature, vol. 363, pp. 603-605, 1993. Nature 347, 1990, 354-358)

[5] John Robertson, "Growth of nanotubes for electronics," Materials Today, vol. 10, pp. 36-43, 2007.

[6] P Nemes-Incze, N Daroczi, Z Sarkozi, A A. Koos, K. Kertesz, O. Tiprigan, Z. E. Horvath, Al. Darabont, L. P. Biro, "Synthesis of bamboo- structured multiwalled carbon nanotubes by spray pyrolysis method, using a mixture of benzene and pyridine" Journal of Optoelectronics and Advanced Materials, vol. 9, pp. 1525-1529, 2007.

[7] S. M. Mominuzzaman, Mahbub Alam, T. Soga and T. Jimbo, "Rearrangements of sp^2/sp^3 hybridized bonding with phosphorus incorporation in pulsed laser deposited semiconducting carbon films by X-ray photoelectron spectroscopic analysis", Diamond and Related Materials, vol. 15, pp. 1795-1798, 2006.

[8] S. M. Mominuzzaman, M.Rusop, T. Soga, T. Jimbo and M. Umeno, "Nitrogen Doping in Camphoric Carbon Films and its Application to Photovoltaic Cell", Solar Energy Materials and Solar Cells, Volume 90, Issues 18-19, pp.3238-3243, 2006.

[9] M.Rusop, S. M. Mominuzzaman, T. Soga, T. Jimbo and M. Umeno, "Photovoltaic Properties of n-C:P/p-Si Cells Deposited by XeCI Eximer Laser Using Graphite Target", Solar Energy Materials and Solar Cells, volume 90, pp. 3205-3213, 2006.

[10] M.Z. Islam, M. Alam, S. M. Mominuzzaman, M.Rusop, T. Soga, T. Jimbo and M. Umeno, "Study of pulsed laser-deposited phosphorus-doped carbon/p-silicon photovoltaic cell", Journal of Crystal Growth, vol. 288, pp. 195-199, 2006.

[11] R. Kamalakaran, M. Terrones, T. Seeger, P. Kohler-Redish, M. Ruhle and YA Kim, T Hayashi and M. Endo, "Synthesis of thick and crystalline nanotube arrays by spray pyrolysis", Appl Phys Lett , vol. 77, pp. 3385–3387, 2000.

[12] Daihua Zhang, Koungmin Ryu, Xiaolei Liu, Evgueni Polikarpov, James Ly, Mark E. Tompson and Chongwu Zhou, "Transparent, Conductive, and Flexible Carbon Nanotube Films and Their Application in Organic Light-Emitting Diodes", NANO LETTERS, vol. 6, pp. 1880-1886, 2006.

[13] Michael W. Rowell, Mark A. Topinka, Michael D. McGehee, Hans-Jürgen Prall, Gilles Dennler, Niyazi Serdar Sariciftci, Liangbing Hu and George Gruner,"Organic solar cells with carbon nanotube network electrodes", Appl. Phys. Lett., vol. 88, pp. 2335061-2335063, 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Fully Parallel Single and Two-Stage Associative Memories for High Speed Pattern Matching

Md. Anwarul Abedin[1], *Tetsushi Koide*[2] *and Hans Juergen Mattausch*[2]
[1]Dhaka University of Engineering and Technology, Gazipur, Bangladesh
[2]Research Institute for Nanodevice and Bio Systems, Hiroshima University, Japan
E-mail: abedin14@yahoo.com

Abstract – A hardware realization of single and two-stage fully parallel associative memories for high speed reliable pattern matching is proposed. We have designed, fabricated and tested the single stage associative memory test chip designed in 0.35 μm two-poly, three-metal CMOS technology which gives very high speed pattern matching performance. But in some applications single stage search or single winner search makes the system less reliable. To increase the reliability of pattern matching system, we have also introduced a cascaded fully parallel associative memory with two-stage winner search. In two-stage pattern matching architecture we have used two different types of associative memories. One is based on the *k*-nearest-matches search and other one is a special type of associative memory in which winner search is done only among the activated reference patterns. The activation in the second associative memory is done by first associative memory after searching the *k*-nearest-matches. We have already designed, fabricated and tested the associative memories separately. The complete two-stage pattern matching system is tested here with MATLAB software and hardware realization is currently under the design process.

I. Introduction

Associative memories have been studied and used as a possible solution for speeding up time-consuming content related searches and for allowing access to data by name or partial content rather than by location or address. In its simplest form, an associative memory can be viewed as a hardware device consisting of N fixed-size cells, each being marked as empty or storing a data word or record. When presented with a search key and a mask specifying the relevant fields of the stored words, the associative memory responds by marking all the words that match the specified key or, more generally, satisfy the search requirements. An associative memory based system can perform recognition by calculating the distances between input patterns and stored reference patterns (Fig. 1). As a measure to express the differences between input data and reference data, the term "distance" is used. The reference pattern with minimum distance is referred to as the "winner" and the reference pattern with the next smallest distance is referred to as the "nearest-loser". Architectures for fully-parallel winner-search according to the Hamming distance [1] and the Manhattan distance [2] have been proposed. Both Hamming and Manhattan distances can be represented by,

$$D = \sum_{i=1}^{W} |S_i - R_i| \qquad (1)$$

Where, $S = \{S_1, S_2, \ldots, S_W\}$ and $R = \{R_1, R_2, \ldots, R_W\}$ are input and reference data, respectively. D is called the Hamming distance, when S_i and R_i are 1-bit binaries. D is called the Manhattan distance, when S_i and R_i are n-bit binaries ($n > 1$). Digital word-parallel and bit serial associative memories for Hamming distance [3] and Manhattan distance [4] search have also been proposed. Since the complete system is realized as a digital circuit, the size is relatively large compared with [1, 2].

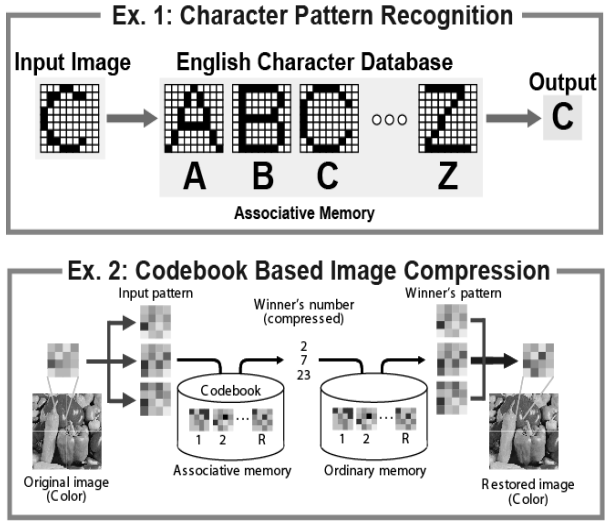


Fig. 1. Application examples of associative memory based system.

However, in many real applications, particularly in the fields of image recognition and authentication, the Euclidean distance defined by,

$$D_e = \sqrt{\sum_{i=1}^{W} (S_i - R_i)^2} \qquad (2)$$

gives the correct distance between two points in a W-dimensional vector space and is known to give better results than Hamming or Manhattan distance.

The most important points for a Euclidean distance hardware implementation are the circuit designs for the square and the square root operations. In particular, the parallel processing for Euclidean distance measurement between the input pattern and many reference patterns, required for a wide range of applications, is difficult. Here, we present a solution for fully-parallel winner search capability using Euclidean distance, which consists of mixed digital-analog distance calculation circuitry and an analog winner search circuit and can achieve compact sin-

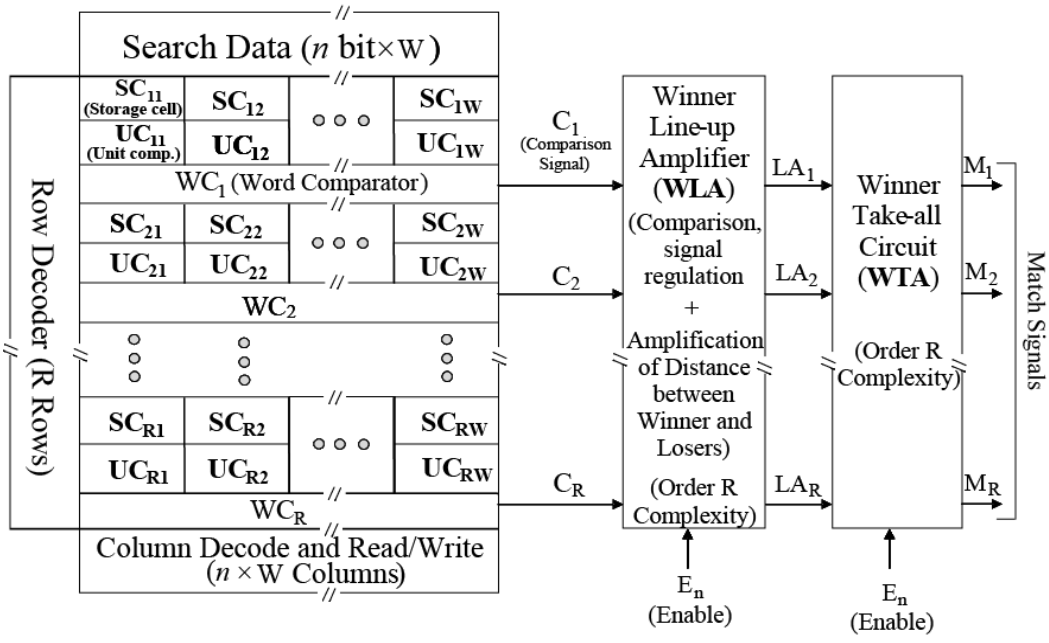


Fig. 2. Block diagram of the mixed digital-analog fully-parallel associative memory for Euclidean distance search.

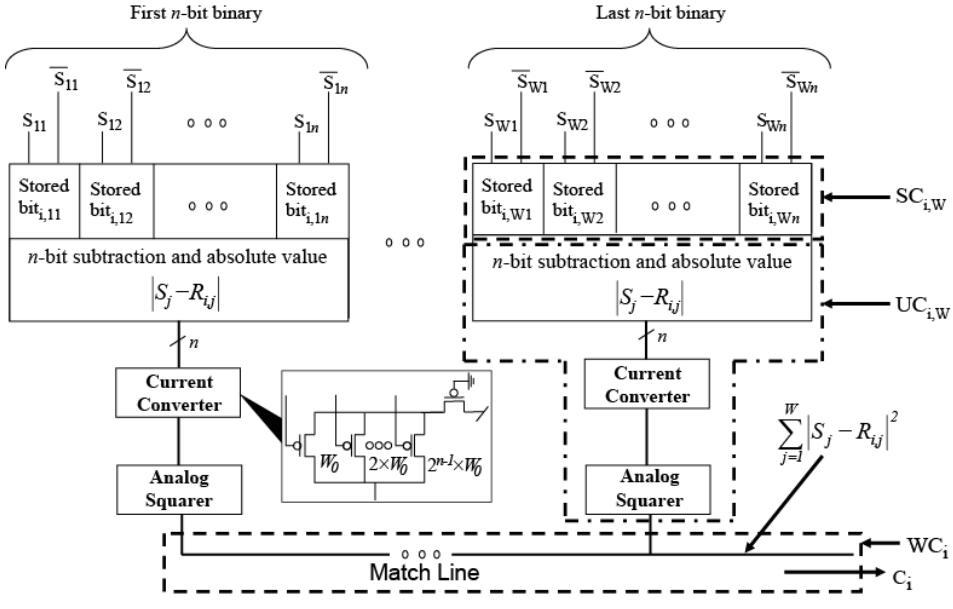


Fig. 3. Key-circuit blocks in the proposed associative memory with fully-parallel nearest Euclidean distance search.

gle-chip implementation in conventional CMOS technology as well as short nearest match times up to large distances The prototype of associative memory we used here has been already designed [5] and has mixed analog-digital fully-parallel architecture for nearest Euclidean distance search..

II. Single Stage Fully Parallel Associative Memory

A. Architecture

The block diagram of the proposed associative memory with fully-parallel single-stage winner search capability realizing the Euclidean distance measure is shown in Fig. 2. The memory part consists of conventional read/write periphery for storing the reference-data words and for reading out the nearest-match data. A row of the memory field contains W storage cells (SC), each with n bits plus the circuitry for unit comparison (UC) and word comparison (WC). The word-comparison results C_i are transferred to the winner-line-up amplifier (WLA). The outputs signals LA_i of the WLA are finally evaluated by the winner-take-all circuit (WTA) to decide on the row, which contains the winner data.

The memory-field structure of the proposed Euclidean distance search hardware is shown in Fig. 3 in more detail. Here, digital n-bit subtractor and absolute value calculation units compare the W binaries, each with n-bit, in all rows of the memory field storing the reference data in parallel with the input data. The digital absolute

unit-distance values are then converted into analog currents using current converters (CC). To realize the CC function the gates of the CC-transistors are connected to the corresponding *n*-bit output-signal lines of the unit comparator and their drains are connected together to add the analog currents of all CC-transistors. The width of each CC-transistor, $2^{n-1} \times W_0$, varies depending on its bit position in the binary so as to correctly distinguish the weight of each bit. The analog currents from each *CC* are then squared using analog current squarer circuits [6].

Finally, the output currents from all squarer circuits are added to get a Euclidean-distance-equivalent current which constitutes the match line current and is fed to the *WLA* for further processing. The *WLA* amplifies the differences of current signals between winner and losers and regulates the winner signal to a suitable level for further distance amplification by the *WTA*. The initial job of the *WTA* circuit is to amplify winner-loser distances by voltage-current-voltage transformations. The final decision circuit in the *WTA* consists of inverters with an adjusted switching threshold. It generates a "1" for the winner row and a "0" for each loser row.

B. Layout Design and Measurement Results

The photomicrograph of the fully parallel single-stage winner search associative memory is shown in Fig. 4. It contains 64 reference words with 16 binaries each 5-bit long and the core area of the fabricated test chip is 5.12 mm^2.

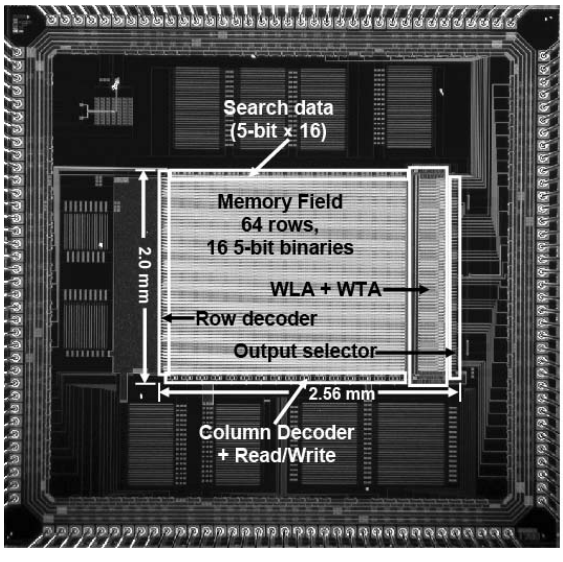


Fig. 4. Photomicrograph of the single-stage winner search associative memory test chip, designed in 0.35 μm CMOS technology.

Fig. 5 shows the measured nearest-match times, which are the time period from rising edge of the enable signal (*En*) to the winner decision at the output of the winner-take-all circuit, as a function of the distance between the winner and the input-data word. The graph shows that the system can recognize the winner perfectly for a wide range of input-winner distances from 1 to 32 (in practical applications winner patterns with large input-winner distance are seldom). Here, the results for the most difficult search condition of (i) winner-nearest loser distance = 1 and a relatively easy search condition of (ii) winner-nearest loser distance = 4 are plotted, from which, we obtained very fast winner search times around 100 ns at a relatively small average power dissipation of less than 183 mW.

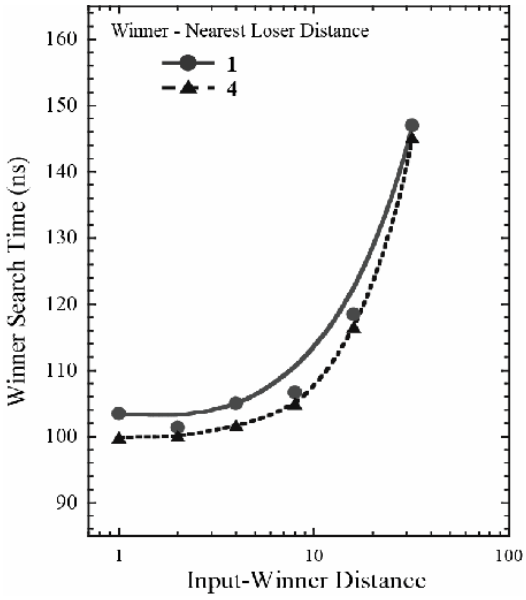


Fig. 5. Measured winner search time of the single winner search associative memory.

III. Two-stage Search Associative Memory

As we mentioned before, in some applications single-stage winner search or single winner search makes the system less reliable. To increase the reliability of pattern matching system, here we will introduce a cascaded fully parallel associative memory with two-stage winner search. In this architecture a *k*-nearest-matches search associative memory will be used in the first stage. In next few sections we will discuss about the architecture and test chip information.

A. *K*-nearest-match Associative Memory Architecture

The structural blocks of the proposed compact associative memory with fully-parallel search capability for *k*-nearest-matches search is shown in Fig. 6. The memory part and the winner search circuit are almost similar to the previously mentioned single stage winner search circuit. The change here is that here, finally from the output of each WTA one feedback line is added to the corresponding match line through a feedback element. It forces the previous winner line to become a loser when the enable signal is raised again.

The winner search circuit which can search *k*-nearest-matches is shown in more detail in Fig. 7. To provide *k*-nearest-matches search a feedback loop is added from the output decision circuit to the match line. The principle operation of this feedback circuit is to provide a "low" signal to the output when the corresponding line's output selector generates "1". Since the output of the feedback circuit is connected to the match line via a pMOS transistor whose source is connected to V_{DD}, it charges

the corresponding match line to V_{DD} and thus makes the corresponding row to act as a loser row during subsequent searches. So, in the next clock cycle the system can search the next winner which was a nearest loser in the previous clock cycle. The circuit continues to search the 2^{nd}, 3^{rd}, 4^{th}, ….. n^{th} winners one by one in every clock cycle until the user stops the winner search enable signal. As the feedback element we have used a NOR type SR flip-flop shown in Fig. 8. The SR flip-flop is suitable for our required feedback element because it maintains a stable output even after the inputs are turned off. This feedback element has a selector signal (reset line of the SR flip-flop), and by controlling it, the complete circuit can be used either as a single winner search circuit or as a *k*-nearest-matches search circuit.

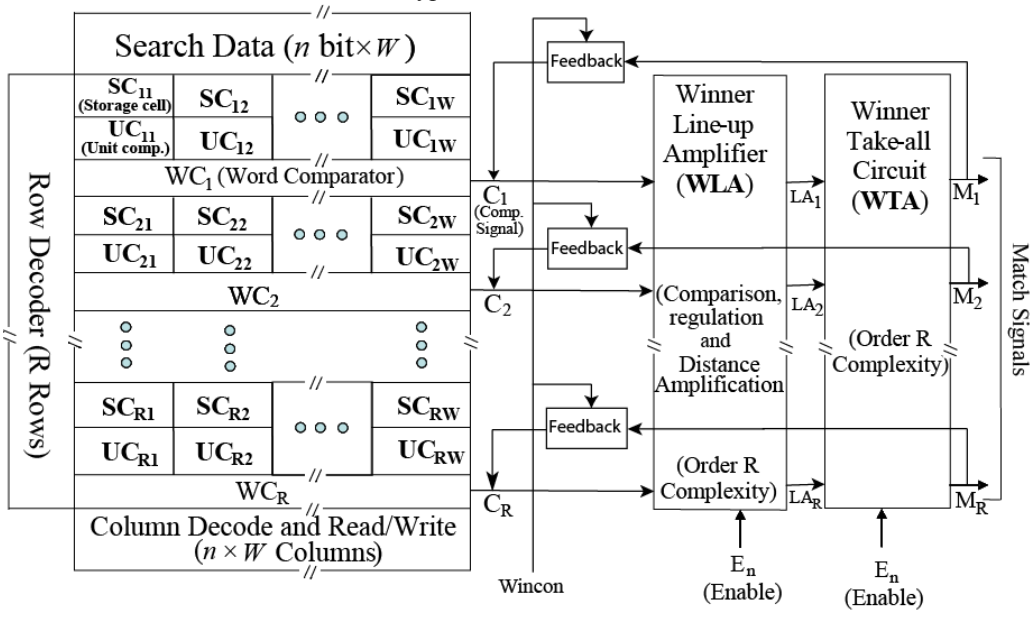


Fig. 6. Block diagram of the mixed digital-analog fully-parallel associative memory for *k*-nearest-matches search.

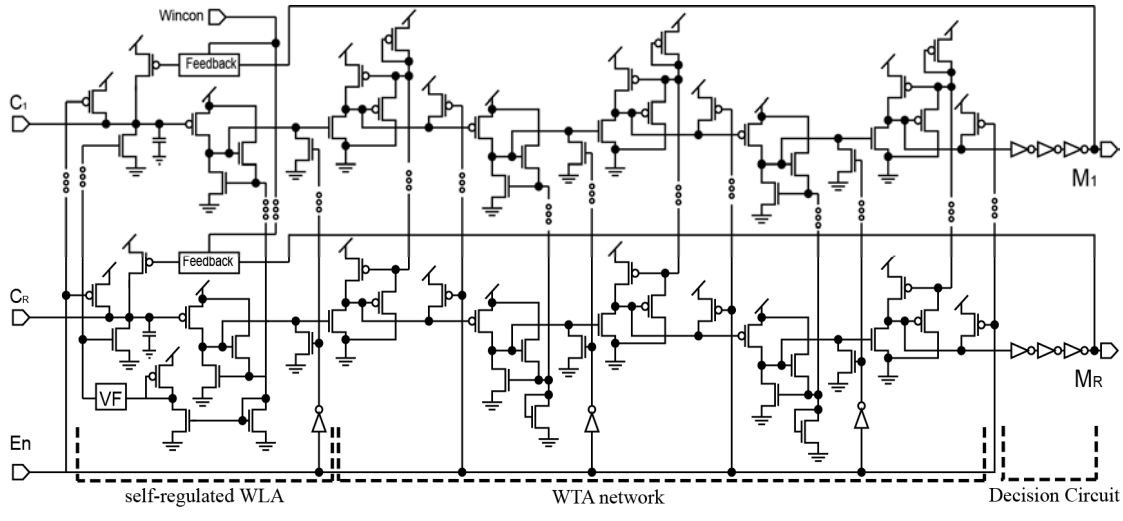


Fig. 7. *K*-nearest-matches search circuit.

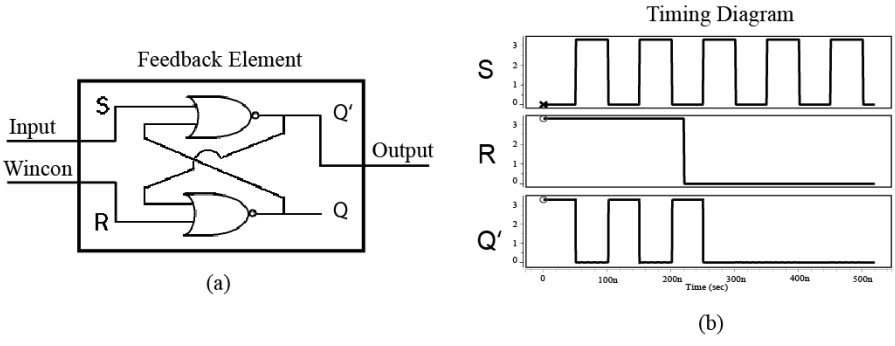


Fig. 8. Detailed structure of the NOR-type SR Flip-Flop used as feedback element with its timing diagram.

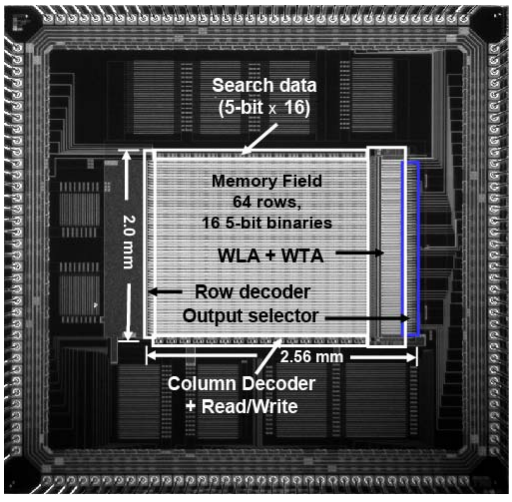


Fig. 9. Photograph of the test chip for *k*-nearest-matches search type associative memory.

B. Layout Design and Measurement Results

Fig. 9 shows the fabricated associative memory for *k*-nearest-matches search used as the first associative memory, which measures 5.12 mm^2 in 0.35 μm, 2-poly, 3-metal CMOS technology. It contains 64 reference words with 16 binaries each 5-bit long. Fig. 10 shows the *k*-nearest-matches search results from the fabricated test chip. Here the search result up to fourth-nearest match or fourth winner is shown. The input-winner distances (Euclidean) of the reference patterns are 1, 2, 3, and 4, respectively. The average power dissipation of the associative memory is around 184 mW. As the second associative memory we have used an associative memory which can search only one row as winner from the previously activated reference pattern rows.

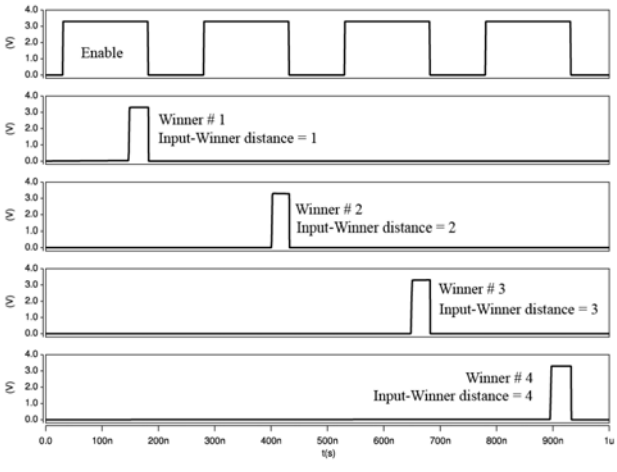


Fig. 10. *K*-nearest-matches search results from the fabricated test chip.

C. Complete Model of the Two-stage Winner Search Associative Memory

Block diagram of the proposed cascaded associative memory architecture is shown in Fig. 11. It contains two associative memory blocks, one for the main reference data and another for the additional feature reference data.

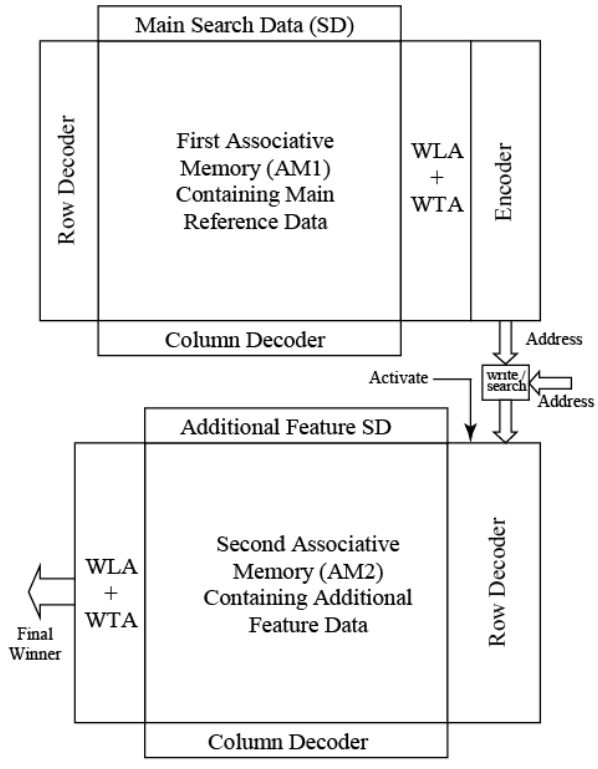


Fig. 11. Block diagram of the two-stage winner search system.

As the first associative memory we have used the *k*-nearest-matches search fully parallel architecture which contains the main reference data. The second associative memory contains some characteristic feature data of the input pattern stored in the first associative memory. The features are selected in such a way that they have the minimum dependency on size and variation of data. Given a segmented character we have extracted its moment based features as follows [7]: Total mass (number of pixels in a binarized character), Centroid, Elliptical parameters (i.e. Eccentricity (ratio of major to minor axis) and Orientation (angle of major axis)), and Skewness. The winner search in the second associative memory is done based on the information of winners searched in the first associative memory. The information of winners searched by the first associative memory is transferred to second associative memory and the corresponding rows are activated. The winner search in the second associative memory is done only among these activated rows. As shown in Fig. 12, *k* winners are searched by the first associative memory. Then the corresponding rows in the second associative memory are activated for the final winner search. At last the single winner searched by the second associative memory is taken as the final winner.

To connect the second associative memory with the first one we have added a multiplexer before the row decoder. It initially provides addresses to store the feature reference data in the memory and during the search operation it provides the address of the *k*-nearest-matched lines selected by the first associative memory to activate these lines for final winner search in second stage.

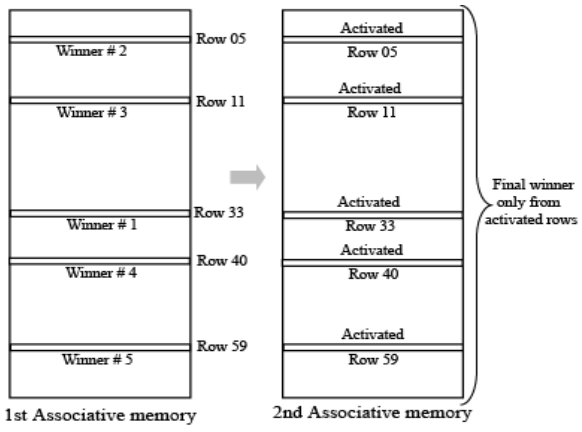


Fig. 12. Operation example of two-stage winner search system.

D. Simulated Performance

We have examined the system performance in a real application for hand written English character recognition using MATLAB software. A number of 35 datasets of English characters written by four different writers were used for the experiment. Fig. 13 shows some gathered data samples. We have tested our proposed system and compared the results with a system which uses only one associative memory. The results are shown in tabular form in Table I. From the table we can see that the proposed system has improved the reliability of pattern matching and the misclassification rate is reduced from 13.0% to 5.3%. Due to the large variation of the hand written characters the misclassification rate of the writer "A" is larger than the other writers.

Fig. 13. Handwritten samples from different writers used as input data

Table I
Classification results for two data sets from four different writers

Writer	Test Set 1 (26 samples)		Test Set 2 (26 samples)	
	Miss-classify		Miss-classify	
	Single-stage	Two-stage	Single-stage	Two-stage
A	4	2	4	2
B	4	1	3	1
C	3	1	3	1
D	2	1	4	2
Total	12.5%	4.8%	13.5%	5.8%

V. Conclusion

In this paper, a hardware realization of single-stage and two-stage pattern matching system using mixed digital-analog fully parallel associative memory is proposed. In the single-stage associative memory only one winner is searched in single stage. From the measurement results it is clear that that architecture can be used for high speed pattern matching system as it can search the winner from the 64 reference patterns only around 100 nsec. For the two-stage winner architecture in the first stage, winner search is done in sequential order to search not only the best nearest-match data but also a freely chosable number of *k* nearest-matches from the main reference data. Then, in the second stage, the winner is searched in the additional feature memory among the *k*-nearest-matched rows, determined by the first associative memory, to get the final winner. Since the proposed architecture is applicable for different distance search associative memories, we can use hybrid distance measures like Hamming / Manhattan distance in the first associative memory and Euclidean distance in the second. MATLAB simulation results shows that the proposed two-stage system improved the reliability of the pattern matching and the misclassification rate is reduced from 13.0% to 5.3% compared with the single stage pattern matching.

Acknowledgements

Part of this work is supported by Monbukagakusho Scholarship and "Interdisciplinary Research on Integration of Semiconductor and Biotechnology" program of the Ministry of Education, Culture, Sports, Science and Technology of Japan. The test-chip in this study has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo with the collaboration of Rohm Corporation and Toppan Printing Corporation.

References

[1] H. J. Mattausch, *et. al., IEEE Journal of Solid-State Circuits*, vol. 37, pp. 218-227, 2002.

[2] Y. Yano, *et. al., Int. Conf. on Solid State Devices and Materials (SSDM'2002)*, pp. 254-255, 2002.

[3] Y. Oike, *et. al., Proceedings of IP Based System-on-Chip Design Forum & Exhibition (IP-SOC)*, pp. 127-130, 2004.

[4] Y. Oike, *et. al., Proceedings of IEEE Custom Integrated Circuits Conference (CICC)*, pp. 295-298, 2004.

[5] M. A. Abedin, *et. al. Japanese Journal of Applied Physics (JJAP)*, Vol. 46, No. 4B, pp. 2231-2237, 2007.

[6] K. Bult, *et. al. IEEE Journal of Solid-State Circuits*, vol-22, No. 3, pp. 357-365, 1987.

[7] Y. Shirakawa, *et. al., Int. Conf. on Solid State Devices and Materials (SSDM'2004)*, pp. 362-363, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Modified CAEN-BIST Algorithm for Better Utilization of Nanofabrics

Babak Zamanlooy and Ahmad Ayatollahi

Department of Electrical Engineering, Iran University of Science and Technology
Narmak, Tehran, Iran
Babak_Zamanlooe@ee.iust.ac.ir, Ayatollahi@iust.ac.ir

Abstract - In the last 40 years there has been an exponential increase in the number of transistors per processor. This increase has been according to Moore's law that predicted the number of transistors that could be placed on the chip would double every two years. However, there are some challenges like leakage currents, process variation, costs and reliability issues like NBTI and HCI that may result to the end of scaling. A solution to CMOS scaling problems is based on architectures that use nanoelectronic devices. The main problem with these devices is their high defect rates due to their bottom up assembly. The reconfigurability of these architectures enables them to manage the high amount of defects. So, to use these architectures the defects should be found and after that in the mapping phase the defective elements are configured around. One of the architectures based on nanoelectronic devices is the nanofabric architecture proposed by Goldstein and Budiu [3]. Different algorithms have been proposed for finding defect map of nanofabric. One of them is proposed by Brown and Blanton [2]. In this paper a modification in CAEN-BIST algorithm has been proposed which improves the nanoblock defect density, test time and utilization of nanoblocks.

I. Introduction

In the last 40 years there has been an exponential increase in the number of transistors per processor. This increase has been according to Moore's law that predicted the number of transistors that could be placed on the chip would double every two years. However, there are some challenges like leakage currents, process variation, costs and reliability issues like NBTI and HCI that may result to the end of scaling [1], [2].

A solution to CMOS scaling problems is based on architectures that use nanoelectronic devices. Because of the use of bottom-up approaches for building nanoelectronic architectures these circuits will have a regular structure. One possible organization that has gained popularity is that of the crossbar array, shown in Fig. 1. The crossbar array is used in many architectures like nanofabrics, FET based array, CMOL and FPNI [3]-[6]. This architecture consists of two sets of orthogonal wires. At the cross point of any two wires is a bistable, reconfigurable switch. When the latch is in the "off" state, the wires are insulated from one another and have no contact. In the "on" state, the wires are connected

Because of using bottom up approaches for fabricating these architectures, they will have a high amount of defects. To overcome this problem, the reconfigurable nature of nano switches is used. The flow for using these reconfigurable fabrics is shown in Fig. 2. In this flow an initial defect independent step exists which includes technology mapping and global place and routing and generates a soft configuration. In the next step that is defect aware, the soft configuration is converted to hard configuration using defect map. So, one of the important parts of this flow is finding defect map [7].

The architecture which is investigated here is the nanofabric proposed by Goldstein and Budiu [3]. Different algorithms have been proposed for finding defect map in nanofabrics [8]-[10]. One of them is the algorithm proposed by Brown and Blanton [8]. A modification in this algorithm is proposed here that improves the utilization of nanoblocks.

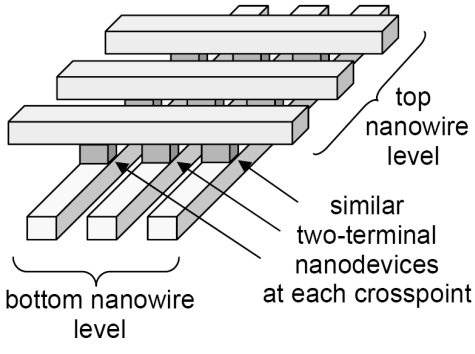


Fig. 1 Nanowire Crossbar

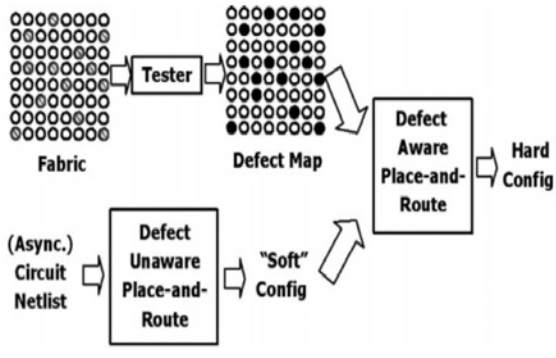


Fig. 2 The tool-flow for using molecular reconfigurable fabrics for computation.

The architecture of nanofabric is described in section II, while the CAEN-BIST algorithm and our modified version is described in section III, finally summary and conclusions are drawn in section IV.

II. Nanofabric

Nanofabric architecture proposed by Goldtein and Budiu [3] is regular and consists of interconnected nanoblocks, which can be fabricated on top of conventional CMOS circuits that are used to apply configuration bits and power supply. The architecture of the nanofabric is shown in Fig. 3.

The nanoBlock is the fundamental unit of the nanoFabric. It is composed of molecular logic arrays for implementing desired logic which is similar to programmable logic arrays, latches used for signal restoration and latching and I/O area to connect the nano block to its neighbours through the switch block. Schematic of nanoblock is shown in Fig. 4.

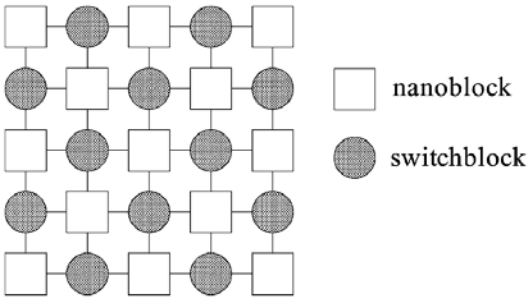


Fig. 3 Nanofabric architecture

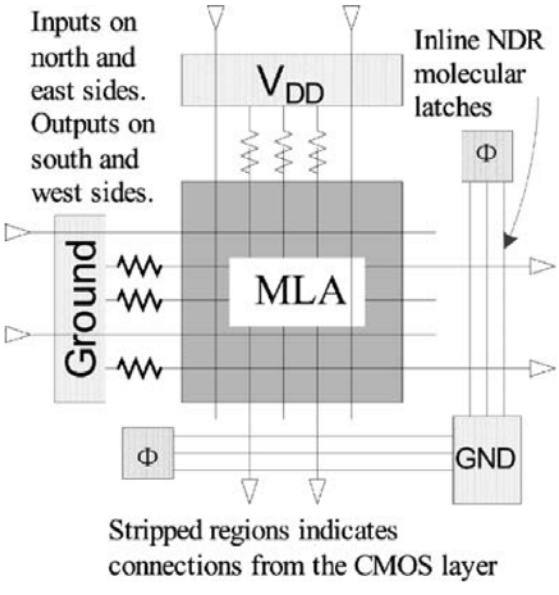


Fig. 4 Schematic of nanoblock

The region between a set of nanoBlocks is known as a switchblock, as depicted in Fig. 5. A switchblock is also reconfigurable and serves to connect wire segments of adjacent nanoBlocks. The schematic of switchblock is shown in Fig. 5.

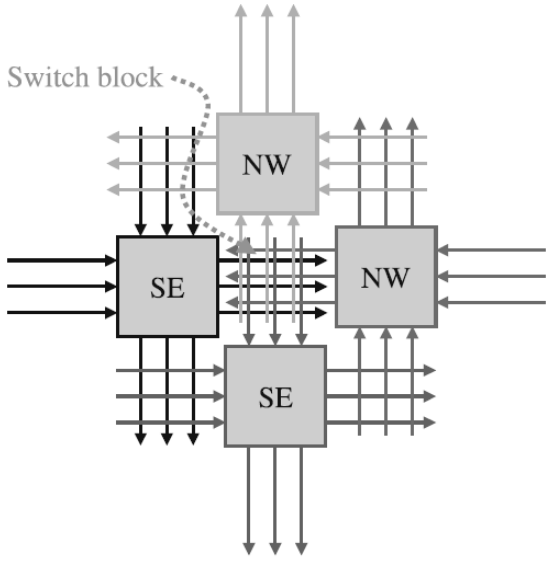


Fig. 5 Schematic of switchblock

III. CAEN-BIST Algorithm

A. CAEN-BIST Algorithm

The testing strategy presented by Brown and Blanton [8] is a built-in self-test algorithm for the nanoFabric called CAEN–BIST. This algorithm enables the nanoFabric to test itself and store the results of the tests internally.

The process of testing a nanoBlock involves a set of test patterns and a set of nanoBlock configurations. The nanoBlock is configured for testing as shown in Fig. 6. A walking binary sequence is applied to the inputs of the block. For a nanoBlock with three inputs, the following test patterns are applied: 100, 010, 001 as shown in Fig. 6. This configuration causes the walking sequence to appear as outputs on the vertical wires. If a walking sequence does not appear correctly, a fault is detected in the nanoBlock. After the test sequence is applied, the configuration is shifted, as shown in Fig. 7. The walking sequence is repeated until all connections in the nanoBlock have been tested. For a nanoBlock with k connections, k test patterns are applied to k different configurations. This test set ensures that every connection in the nanoBlock is tested in both the on and off configurations. This test set provides 100% fault coverage for single stuck-line, bridging, and connection faults.

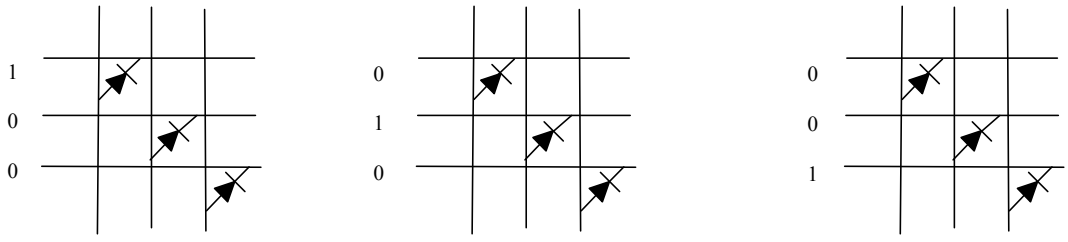


Fig. 6 A walking sequence of ones is applied to the horizontal wires and the output response is transmitted along the vertical wires.

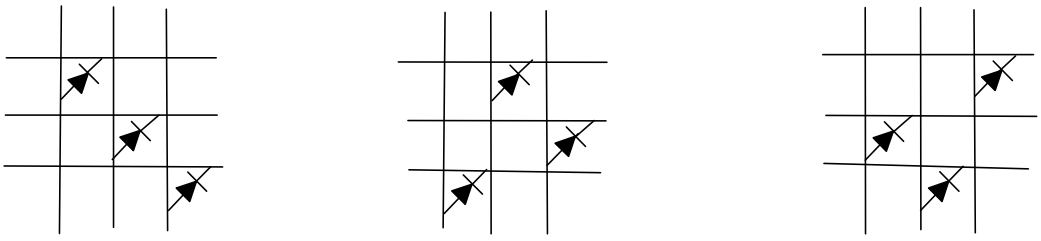


Fig. 7 The configuration of the nanoBlock is shifted after all test patterns have been applied and the process is repeated until all connections have been tested.

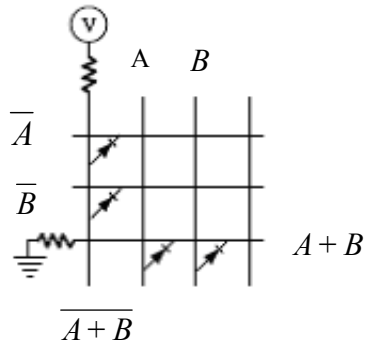


Fig. 8 OR gate implementation with nanoblock

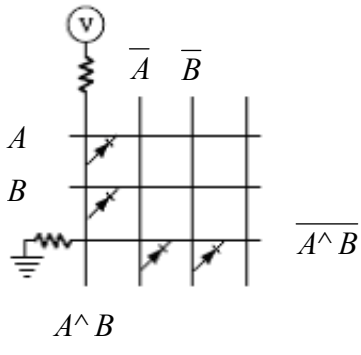


Fig. 9 AND gate implementation with nanoblock

B. Modifed CAEN-BIST Algorithm

Before investigating the modified BIST algorithm, it should be noted that the modified algorithm is based on an assumption. The assumption is that the nanoblocks are used for implementing AND and OR gates. This assumption is based on the fact that most apparent advantage of nanowire crossbars over conventional MOS is their ultra high density of crosspoint devices, and the utilization of crosspoint devices decreases when the crossbar becomes larger, leading to lower logic density. Indeed, the achievable complementary logic density in a crossbar is $O(n^{-1})$, where n is the number of nanowires in each of the orthogonal arrays and is referred to as *the* dimension of the nanowire crossbar. Moreover, the delay of a crossbar logic circuit is $O(n)$ [11]. So, it's better to have a small nanowire crossbar. It seems that the smallest crossbars are the one that implement AND and OR functions. It's obvious that combination of these gates can implement any logic function. The implementation of AND and OR gates using nanoblock is shown in figures Fig. 8 and Fig. 9. To illustrate how a nanofabric can be used to implement a real circuit, the implementation of C17 circuit, shown in Fig. 10 , which is an ISCAS85 benchmark circuit is shown in Fig. 11 [9] .

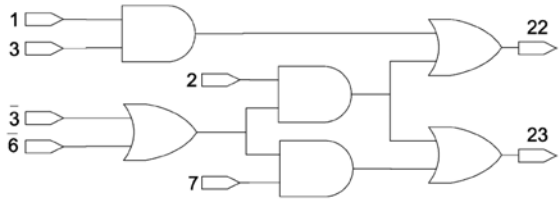


Fig. 10 C17 Schematic

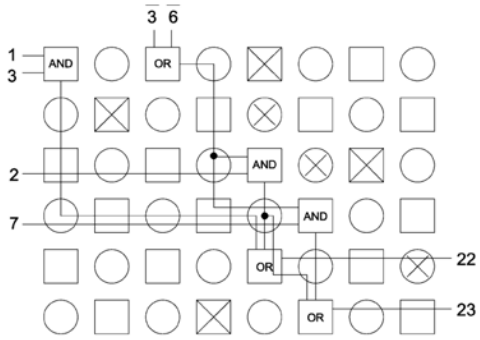


Fig. 11 The mapping of c17 to a defective nanofabric

The CAEN-BIST algorithm investigated in the last section finds defective nanoblocks and switchblocks, which are configured around in the reconfiguration stage. The idea proposed here is that some of defective blocks can be used. The nanoblocks are used for implementing AND and OR gates. The crossbar array has 9 crosspoints. But if we note, in the implementation of AND and OR gates only four similar crosspoints are important. The other point is that these crosspoints are in test groups 1 and 3, shown in Fig. 7 . So, the defective nanoblocks that only their group 2 test has failed can be used. In other words, we only need to apply test patterns to test groups 1 and 3. This will result to a reduction in nanoblock defect dentsity versus component defect density and test time. Also, the utilization of nanofabrics is improved. Although, we investigated the idea for 2 input AND and OR gates, it will work for gates with more inputs. In the following paragraphs the effect of modified CAEN-BIST algorithm is investigated for K*K crossbars with K^2 crosspoints.

If the component defect density is d and defective components are normally distributed, in a nanoblock with K^2 crosspoints the nanoblock defect density (D) is calculated using Eq. 1:

$$D = 1 - (1 - d)^{K^2} \quad (1)$$

In the modified CAEN-BIST algorithm, the number of effective crosspoints is reduced to K^2-K. Simulations for different values of K (K=3,4,5) , shown in Fig. 12 confirm that the nanoblock defect density is reduced.

To test a nanoblock in CAEN-BIST algorithm, k configurations and k^2 test patterns are needed. In the modified algorithm, K-1 configurations and K*(K-1) test patterns are needed so, test time that was $Kt_{config}+K^2t_{tes}t$ is reduced to $(K-1)t_{config}+K(K-1)\ t_{test}$.

Also, it can be said that if the probability of defective nanodevices is uniform this will result to a 1/k improvement in the utilization of nanoblocks, because the effective crosspoints are reduced form K^2 to K(K-1). The improvement results are shown briefly in Table I. Although the idea is investigated for two input AND and OR gates it is useful for gates with more inputs.

Table 1 Improvement Results of Modified CAEN-BIST Algorithm

	Modified CAEN-BIST Algorithm Improvement
Nanoblock Defect Density	$\frac{1-(1-d)^{K(K-1)}}{1-(1-d)^{K^2}}$
Test Time	$\frac{(K-1)t_{config}+K(K-1)\ t_{test}}{Kt_{config}+K^2t_{test}}$
Utilization	$\frac{1}{k}$

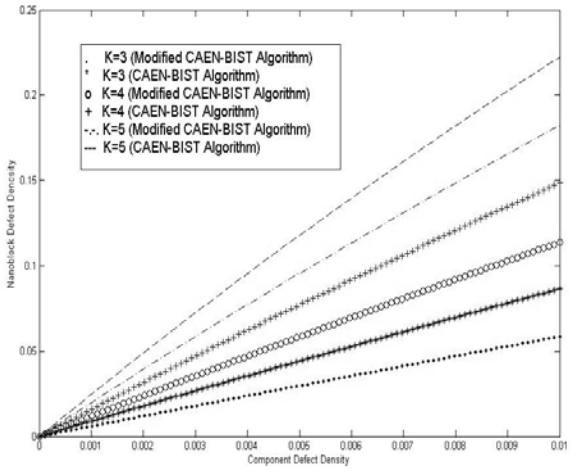


Fig. 12 Comparison of nanoblock defect density in CAEN-BIST algorithm and its modified version

IV. Summary and Conclusion

The architectures based on nanoelectronic components are proposed as a solution for CMOS scaling problems. The main problem of these architectures is their high defect rates, which is managed using reconfigurable nature of these devices. To use the reconfigurability of these devices a defect map is needed. One of the algorithms proposed for finding defect maps is ĆAEN-BIST algorithm. A modified version of CAEN-BIST algorithm has been presented in this paper. This modification is based on the assumption that the nanoblock is used for implementing AND and OR gates and results to reduction of K^2 crosspoints to K(K-1) effective crosspoints, which reduces nanoblock defect density, test time and improves its utilization.

References

[1] M. Haselman and S. Hauck, "The Future of Integrated Circuits: A Survey of Nano-electronics," Submitted to Proceedings of IEEE.

[2] S.-L. Jeng, J.-C. Lu, and K. Wang, "A Review of Reliability Research on Nanotechnology," IEEE Transactions on Reliability, vol. 56, pp. 401-410, 2007.

[3] S. C. Goldstein and M. Budiu, "NanoFabrics: Spatial Computing Using Molecular Electronics," in Proc. International Symposium on Computer Architecture, Sweden, pp. 178-189, 2001.

[4] A. Dehon, "Array-Based Architecture for FET-Based, Nanoscale Electronics," IEEE Transactions on Nanotechnology, vol. 2, pp. 23-32, 2003.

[5] D. B. Strukov and K. K. Likharev, "CMOL FPGA: a reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices," Nanotechnology, vol. 16, pp. 888-900, 2005.

[6] G. S. Snider and R. Stanley Williams, "Nano/CMOS architectures using a field-programmable nanowire interconnect," Nanotechnology, vol. 18, pp. 1-11, 2007.

[7] M. Mishra and S. C. Goldstein, "Defect tolerance at the end of the roadmap," in Proc. International Test Conference, USA, pp. 1201-1210, 2003.

[8] J. G. Brown and R. D. Blanton, "A Built-in Self-test and Diagnosis Strategy for Chemically Assembled Electronic Nanotechnology," Journal of Electronic Testing, vol. 23, pp. 131-144, 2007.

[9] Z. Wang and K. Chakrabarty, "Built-in Self-test and Defect Tolerance in Molecular Electronics-based Nanofabrics," Journal of Electronic Testing, vol. 23, pp. 145-161, 2007.

[10] M. Tehranipoor and R. M. P. Rad, "Built-In Self-Test and Recovery Procedures for Molecular Electronics-Based Nanofabrics," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 26, pp. 943-958, 2007.

[11] M. Dong and L. Zhong, "Logic Synthesis with Nanowire Crossbar: Reality Check and Standard Cell-based Integration," in Proc. Design, Automation and Test in Europe, Germany, pp. 268-271, 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Phase Noise Reduction in CMOS LC Oscillators Using Tail Noise Shaping and G_{m3} Boosting

Kapil Jainwal
Electrical Engineering
Indian Institute of Technology, Bombay
Powai, Mumbai 400076, India
kapiljainwal@ee.iitb.ac.in

Jayanta Mukherjee
Assistant Professor, Electrical Engineering
Indian Institute of Technology
Powai, Mumbai 400076, India
jayanta@ee.iitb.ac.in

Abstract — **This paper presents a new technique for Low-Phase Noise CMOS Quadrature differential oscillator with better phase noise. The design uses tail noise shaping along with G_{m3} boosting circuit and third harmonic tuned LC tank for improved phase noise performance. The proposed quadrature oscillator circuit shows a 4 dB phase noise improvement over standard fixed bias quadrature oscillator. Post Layout Simulation results in a 0.18-um CMOS fabrication process show that the phase noise of the proposed quadrature oscillator to be -114 dBc/Hz at 1-MHz frequency offsets respectively at a center frequency of 5-GHz. The improvement is higher than the sum of reductions provided by the individual techniques. The power consumed increases by a very small amount and thus the Figure of Merit (FOM) shows a 4 dB improvement.**

I. Introduction

One of the major challenges in the design of complementary metal oxide semiconductor (CMOS) single-chip transceiver system is the integration of a low power low noise oscillator with quadrature output. The availability of accurate quadrature signals is a prerequisite for the implementation of image-rejection transceivers. The parallel coupled differential topology is the most popular circuit for quadrature signal generation.

Phase noise of the oscillator is one of the most critical parameter for the quality of signal in transceiver frontends. The conventional Differential Quadrature oscillator shows a poor phase noise performance as compared to a single phase VCO [1], [2], [3], [11]. One of the major sources of phase noise of oscillator circuit is the flicker noise of the biasing transistor with a power spectral density inversely proportional to the frequency. In [4], it is shown that the contribution of the tail noise to the overall Phase Noise is mainly in the flicker noise region. In [1], a technique of tail noise shaping has been shown to reduce the overall Phase Noise significantly.

Further in [5], [6] and [13], it has been shown that by increasing the third harmonic component of the output voltage waveform, the average output voltage can be increased and thus the overall Phase Noise is decreased. This paper presents a technique in the design of a CMOS oscillator, by which both the flicker noise dominated $1/f^3$ shaped part of the Phase Noise spectrum and thermal noise dominated $1/f^2$ shaped part of the Phase Noise spectrum are reduced through a combination of the tail noise shaping scheme along with a G_{m3} boosting circuit which increases the third harmonic component of the output voltage waveform.

II. Design of Low Phase Noise CMOS LC Tank Oscillator

A. Tail Noise Shaping

The flicker noise in MOSFET is either theorized as a random variation in the carrier-density in the MOSFET channel or as a variation in the mobility of the carriers in the channel or alternatively a unified model that approximates one of these two processes depending on the bias voltage [7]. As shown in [6] switching of tail transistor with the oscillator output produces a lower phase noise in the $1/f^3$-region as compared to the fixed bias topology by the periodic switching on and off of the tail transistor thereby ceasing trap formation.

Fig 1(a) shows a fixed bias quadrature oscillator. In [6], the corresponding tail noise shaping circuit is described and is the one given in Fig 1(b). The scheme employs separate biasing for the coupling NMOSFETS, thereby achieving significant phase noise reduction. Increasing higher

harmonic content in the switching signal leads to better noise reduction [10].

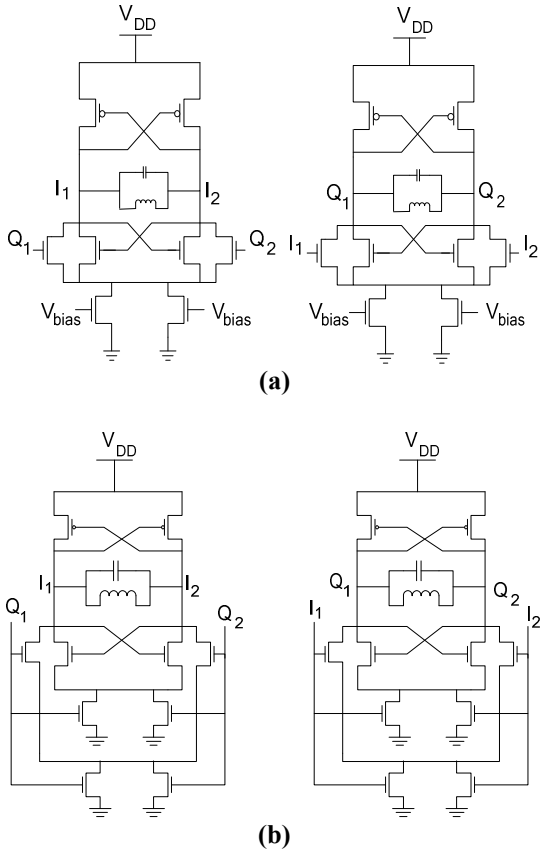


Figure 1. (a) Standard fixed bias parallely coupled quadrature oscillator circuit (b) Modified quadrature oscillator circuit using tail noise shaping (c) G_{m3} boosting circuit

B. G_{m3} Boosting circuit.

The dependence of phase noise on the average voltage across the tank can be given by the phase noise model for an oscillator proposed by Leeson [8], [9]:

$$L\{\Delta\omega\} \propto \frac{1}{V_0^2} \cdot \frac{kT}{C} \cdot \left(\frac{\omega_0}{Q}\right)^2 \cdot \frac{1}{\omega_m^2}$$

where, V_0, is the output voltage across the tank, C is the tank capacitance, Q is the Q-factor of the tank .From Equation (1), it can be seen that the phase noise can be reduced by increasing V_o, which is the voltage across the resonator and is proportional to the slope at the zero crossing voltage across the tank. . Such a result can be achieved by boosting the third harmonic component of the output voltage waveform [5], [10], [12]. The output waveform then has a steeper slope at zero crossing points.

Figure 2 shows a third harmonic boosting circuit [5] based on two PMOSFETS. The cross-coupled transistors amplify the output voltage with the resistance R_p and capacitance C_p tuned such that only the third harmonic component is filtered out. Thus the third order transconductance (G_{m3}) is boosted only.

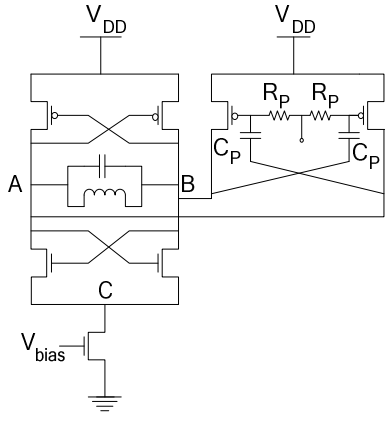


Figure 2. G_{m3} boosting circuit

In the present design both of these has been employed to reduce phase noise. The final circuit using both techniques namely tail noise shaping and G_{m3} boosting is shown in Figure 3.

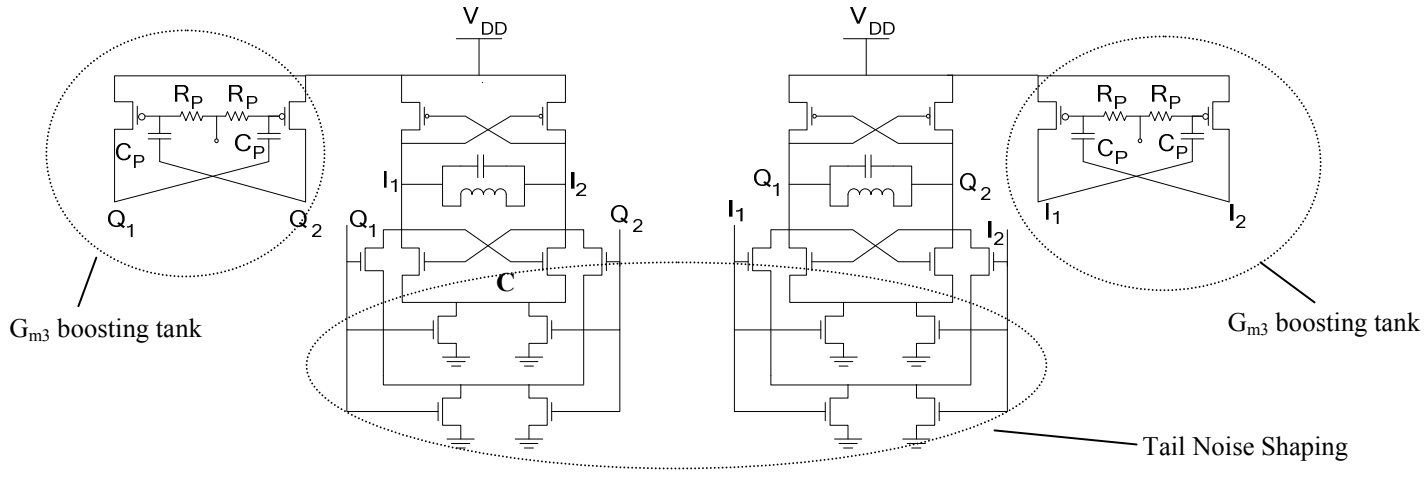


Figure 3. Quadrature Oscillator circuit with tail noise shaping and G_{m3} boosting circuit

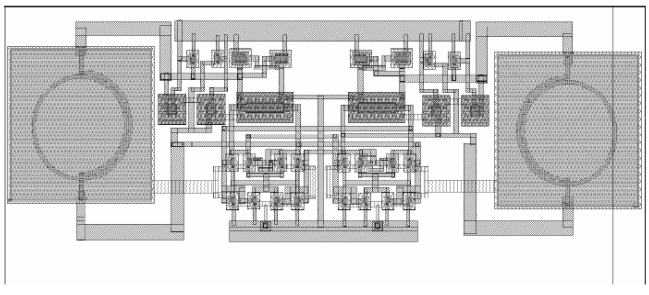


Figure 4: Layout for quadrature oscillator circuit with switching and biasing scheme along with Gm3 boosting Circuit

IV. Simulation Results

An important issue that crops up is whether the simulator is capable of simulating the phase noise reduction due to tail noise switching. Figure 5 compares the Phase noise results in a differential oscillator whose tail transistor is switched by a pulse train with one having its tail transistor biased by a fixed voltage. The Phase Noise reduction due to switching of tail transistor in differential oscillators is captured by the simulator. This has been also demonstrated in [13].

The proposed quadrature oscillator circuit shown in Figure 3 was simulated in the UMC 0.18-um CMOS technology using Spectre RF simulator in cadence. Figure 4 shows the corresponding layout. The models used for simulation were BSIM3. The simulation was carried out over all four process corners. Figure 6 shows the comparison of the Post Layout Phase Noise simulation results obtained for the various configurations in the typical model. The variation in simulation results over the other process corners was minor. The capacitors used in the G_{m3} boosting circuit were MIM (Metal-Insulator-Metal) type, which from simulation results is found to give better phase noise performance. The inductor used in the tank was of single metal spiral type. The capacitors and inductors were adjusted so that the frequency of oscillation remained around 5 GHz for all cases so that comparison between circuits is fair.

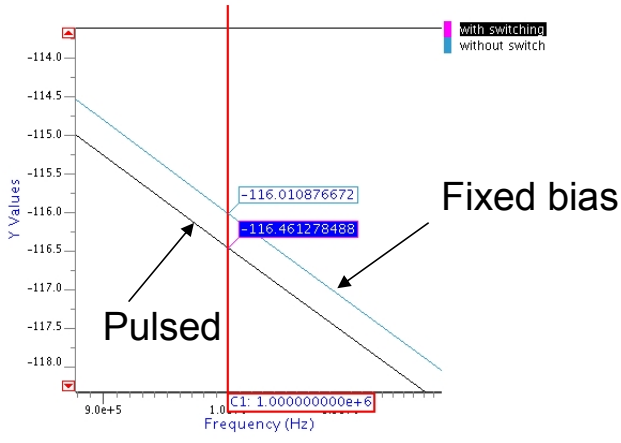
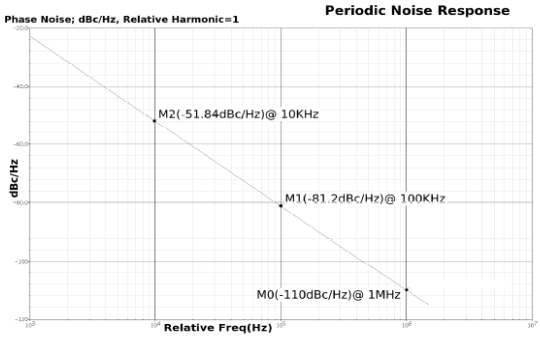


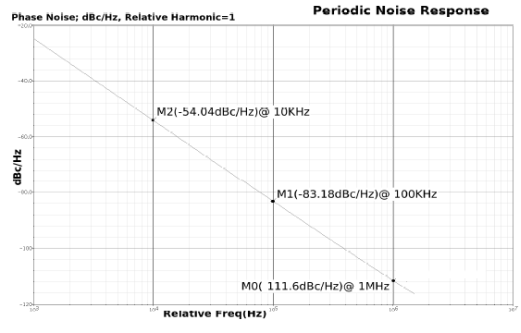
Figure 5: Comparison of the Phase Noise results obtained using Eldo RF for a differential oscillator biased with a fixed voltage to that biased by a pulsed voltage

It can be seen that tail noise shaping reduces the Phase Noise to -111.6 dBc/Hz at 1 MHz offset from -110dBc/Hz for a simple parallely coupled quadrature oscillator with fixed bias. Simulations results show that the G_{m3} boosting circuit provides on average an improvement of about 1dB. However with both the techniques applied, the Phase Noise reduces to -114dBc/Hz. Thus the proposed design shows an improvement of 4dB over the simple fixed bias quadrature oscillator. This improvement is more than the sum of the individual

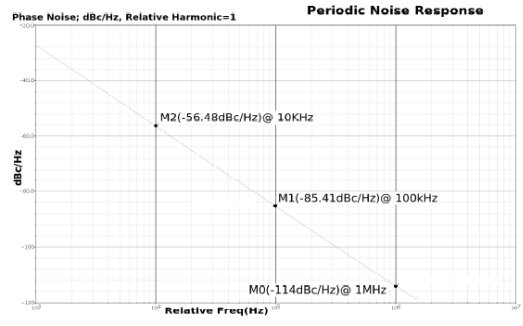
techniques. The explanation for this improved result is that the two techniques are mutually helpful. For example, the *Gm3* boosting circuit increases higher harmonic content in the tail transistor switching which in turn causes the switching of tail transistor noise to be more efficient. Similarly switching on and off of the tail transistor causes the higher harmonic content of the output to increase which in turn reinforces the Gm3 boosting circuit. The power consumption of the various schemes shows little difference and hence the Figure of merit (FOM) also increases by 4dB.



a. Simple fixed bias quadrature oscillator



b. Quadrature oscillator with tail noise shaping



c. Quadrature oscillator with tail noise shaping and *Gm3* boosting

Figure 6: Comparison of Phase Noise obtained for various configurations

V. Conclusion

In present paper, we have presented a new design approach for Low Phase Noise CMOS quadrature oscillator for improved phase noise performance over conventional quadrature oscillator. Phase noise is minimized by switching and separation of biasing transistors along with a G_{m3} boosting technique. The switching technique improves phase noise performance by reducing the trap formation and thereby improves the phase noise performance. The G_{m3} boosting circuit further improves the phase noise by enhancing 3rd harmonics. The two techniques are mutually reinforcing and the overall Phase Noise reduction is more than the contributions of the individual techniques. The post layout simulation results in 0.18um CMOS fabrication process shows that the phase noise of the present design, is -114 dBc/Hz at an offset of 1MHz. The power consumed increases only negligibly and the Figure of Merit of the oscillator increases by 4 dB.

Acknowledgement

We wish to thank Europractice for providing the UMC 0.18 um design kit used in this work. Also we would like to thank the VLSI Lab of IIT Bombay where all the simulation was done.

References

1. Chan-Young Jeong, and Changsik Yoo," 5-GHz Low-Phase Noise CMOS Quadrature VCO", IEEE Microwave & Wireless Compon. Lett. vol. 16, no. 11, Nov. 2006
2. Sheng-Lyang Jang, Yun-Hsueh Chuang, Chien-Feng Lee and Shao-Hua Lee," A 4.8-GHz Low-Phase Noise CMOS Quadrature VCO", IEEE J, 2006.
3. Hayman N.Shanan, Michael P. Kennedy, "A Technique to Reduce Flicker Noise Up-Conversion in CMOS LC Voltage-controlled Oscillator", IEEE J, 2004.
4. Mukherjee J., Roblin P. and Akhtar S. "An Analytic Circuit-Based Model for White and Flicker Phase Noise in LC Oscillators," IEEE Trans. on Ckts. and Sys I: Regular Papers, vol 54, July 2007, pp 1584-1598
5. Huijung Kim, Woonyn Kim, S.Kang ," A Low Phase Noise LC VCO in 65 nm CMOS Process Using Rectangular Switching Technique", IEEE J. Solid-State Circuit, vol. 17, no. 8, pp. 610-612, Aug 2007.
6. E. A. M. Klumperink, S. L. J. Gierkink, A. P. van der Wel, and B. Nauta, "Reducing MOSEF 1/f noise and power consumption by switched biasing,"IEEE J. Solid-State Circuits, vol. 35, no. 7, pp. 994–1001, Jul.2000.
7. Hung K.K., Ko P.K., Hu C. and Cheng Y.C. "A unified model for flicker noise in metal-oxide-semiconductor field-effect transistors," IEEE Trans. Electron Devices, vol 37, pp 654-665, Mar. 1990
8. B. Razavi, "A study of phase noise in CMOS oscillators," IEEE J. Solid-State Circuits, vol. 31, no. 3, pp. 331–343, Mar. 1996.
9. Thomas H. Lee, Ali Hajimiri, "Oscillator Phase Noise: A Tutorial", IEEE J. Solid-State Circuits, vol. 35, no. 3, pp. 326–335, March. 2000.
10. C.C. Boon, M.A. Do, K.S. Yeo, J.G. Ma and X.L. Zhang "RF CMOS Low-Phase-Noise LC Oscillator through Memory Reduction Tail Transistor," in IEEE Trans. on Circuits and Systems-II: Express Briefs, pp. 85–90, vol.51, no.2, Feb 2004.
11. Jeong Chan-Young and Yoo Changsik "5-GHz Low Phase Noise CMOS Quadrature VCO," IEEE Microwave and Wireless Component Letters, pp. 609–611, vol.16, no.11, Nov 2006.
12. S. L. J. Gierkink, "A low-phase-noise 5-GHz CMOS quadrature VCO using super-harmonic coupling," IEEE J. Solid-State Circuits, vol. 38, no. 7, pp. 1148–1154, Jul. 2003.
13. Ahmed K. Kassim I , Khaled Sharaf ,and Hani Ragaie"Tail Current Flicker Noise Reduction in LC VCOs by Complementary Switched Biasing", IEEE ICM Dec 2003, pp 102-105

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Reversible Multipliers: Decreasing the Depth of the Circuit

Fateme Naderpour, Abbas Vafaei

Department of Computer Engineering
The University of Isfahan, Isfahan, Iran

naderpour@comp.ui.ac.ir , abbas_vafaei@eng.ui.ac.ir

Abstract - There are many arithmetic operations which are performed, on a computer arithmetic unit, through the use of multipliers (e.g., exponential and trigonometric functions). Consequently, optimized multipliers are on demand while designing an arithmetic unit. On the other hand, given the advent of quantum computer and reversible logic, design and implementation of digital circuits in this logic has gained popularity. In reversible circuit design, decreasing three parameters is of interest: quantum cost, depth of the circuit and the number of garbage outputs. In this paper, we propose a novel reversible multiplier with the aim of decreasing the depth of the circuit while neither scarifying any extra quantum cost nor garbage outputs. The partial products, as is the case for prior works, are generated in parallel using Peres gates and thereafter a reversible multi-operand adder consisting of reversible full-adders and half-adders produces the final product.

Keywords- Reversible circuits, Multiplier, Depth of a reversible circuit, Quantum cost, Garbage output.

I. Introduction

Power dissipation is one of the important considerations in digital circuit design. A part of this energy loss is due to switches and materials. The other part arises from Landauer's principle [1] where he proved that in irreversible circuits losing one bit of information dissipates *KT*.Ln2 joules of heat energy, where *K* is the Boltzmann's constant and *T* is the absolute temperature. Although the amount of dissipating heat for room temperature is small, it can not be neglected in some applications (e.g., quantum circuit design). Furthermore, Bennett [2] showed that reversible circuits do not lose information due to the one-to-one mapping between inputs and outputs; hence no extra energy loss.

In the design of reversible circuits two restrictions should be considered:

- Fan-out is not permitted
- Loops are not permitted

Due to these restrictions, synthesis of reversible circuits can be carried out from the inputs towards the outputs and vice versa [3].

From the point of view of reversible circuit design, there are three parameters for determining the complexity and performance of circuits [4]:

- **Quantum cost (QC):** the number of 1×1 or 2×2 reversible gates which are used in circuit.
- **The number of garbage outputs (GO):** The number of dummy (unused) outputs which are added in order to make the circuit reversible.
- **Depth:** The number of 1×1 or 2×2 reversible gates which are in the longest path from input to output.

Reduction of these three parameters is the bulk of the work in reversible circuit design. However, minimizing all these parameters together is not easy. In this paper, we aim to design a reversible multiplier with the minimized depth while not increasing the QC or GO of the circuit with respect to its previous counterparts.

The rest of paper is organized as follows: a background of reversible gates and circuits are explained in Section 2. In Section 3, a brief explanation of multipliers is presented while in Section 4 the new reversible design is proposed. Section 5 compares the new design with previous works and finally Section 6 concludes the paper.

II. Background of Reversible Circuits

An *n*×*n* reversible circuit consists of *n* inputs and *n* outputs with mapping of each input assignment to a unique output assignment and vice versa.

A. Reversible Gates and Circuits

There are two main types of reversible gates: Toffoli [5] and Fredkin [6]. An *n*×*n* Toffoli gate passes the first (*n*-1) inputs to outputs unaltered (as control signals) and for the last output the n^{th} input inverts (as target signal) if all the previous (*n*-1) signals are '1'. Assuming x_i as input and y_i as output, then [3]:

$$y_i = x_i \qquad 1 < i < n-1$$
$$y_n = x_n \oplus (x_1.x_2.....x_n)$$

Fig. 1 depicts a 3×3 Toffoli gate.

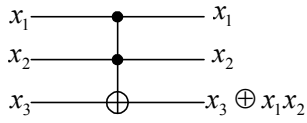


Fig. 1: 3×3 Toffoli gate

A Toffoli gate with one (two) input(s) is also known as NOT (CNOT or Feynman) gate respectively.

Since fan-out is not permitted in reversible circuits, a 2×2 Toffoli gate (Feynman gate) can be used to provided it (see Fig. 2).

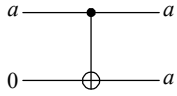


Fig. 2: Fan-out gate

3×3 Fredkin Gate: this gate passes the first input unaltered; second and third inputs are swapped, if the first input is '1'. A 3×3 Fredkin gate is presented in Fig. 3 [6].

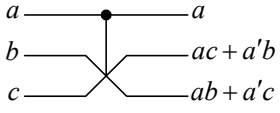


Fig. 3: 3×3 Fredkin gate

Peres Gate: the block diagram and input-output equations of this gate are shown in Fig. 4 [7].

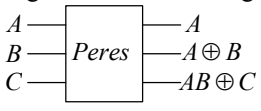


Fig. 4: Peres Gate

The quantum cost and depth of a reversible circuit is determined from its implementation with elementary quantum logic gates [8]. These gates are depicted in Fig. 5 [9].

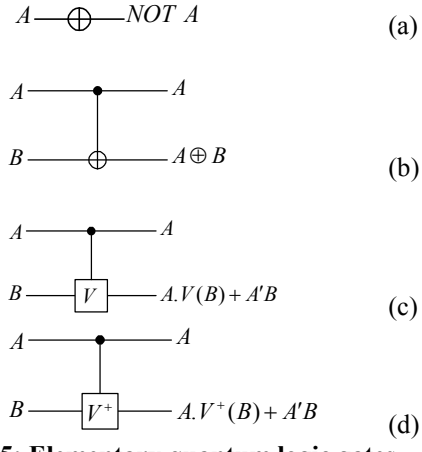


Fig. 5: Elementary quantum logic gates

Controlled-V gate depending on the value of control line, changes the value on the target line using the transformation given by the matrix $V = \frac{i+1}{2}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}$.

Controlled-V^+ gate depending on the value of control line, changes the value on the target line using the transformation $V^+ = V^{-1}$.

The following rules are postulated for the controlled V and V^+ gates [10]:

$$V.V^+ = 1 \quad , \quad V^2 = V^{+2} = NOT$$

Each 2-qubit gate has a quantum cost and depth of 1. Moreover, each symmetric gate pattern (as is shown in Fig. 6) has also a cost and depth of 1 [9].

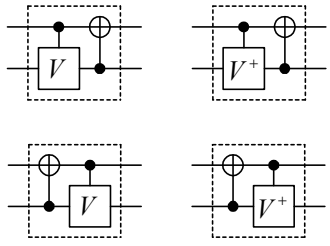


Fig. 6: Merged 2-qubit gates

In the next section, we obtain quantum cost and depth of the mentioned reversible gates according to their implementation using elementary quantum gates. Furthermore, since full-adder and half-adder are two main arithmetic cells which are used in the design of the proposed multiplier, the reversible implementation of these two cells by means of quantum gates is also discussed.

B. Reversible gates Implemented using elementary Quantum gates

Reversible implementations of 3×3 Toffoli, Peres and Fredkin gates using elementary quantum gates are shown in Fig. 7, Fig. 8, and Fig. 9, respectively.

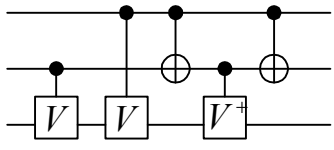


Fig. 7: Implementation of the 3×3 Toffoli gate [11]

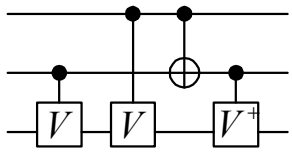


Fig. 8: Implementation of the Peres gate [12]

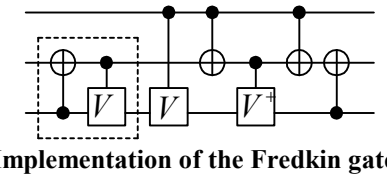


Fig. 9: Implementation of the Fredkin gate [11, 13]

Table I contains the values obtained for quantum cost and depth of each gate.

Table I: Quantum cost and Depth of gates

	TOF3		Peres		Fredkin
Depth	1^{st}, 2^{nd} outp.	3^{rd} outp.	2^{nd}, 3^{rd} outp.	1^{st} outp.	5
	5	4	4	3	
QC	5		4		5

Reversible implementation of full-adder, as a basic arithmetic cell, attracted the attention of reversible circuit designers. Many reversible implementations have been proposed by various designers for this cell. Fig. 10 shows the reversible implementation of full-adder, introduced in [10], while its quantum cost, depth, and the garbage outputs equal 6, 5, and 2, respectively.

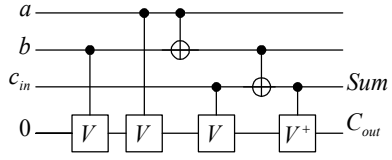


Fig. 10: Reversible implementation of full-adder

Furthermore the reversible implementation of a Half-adder is also shown in Fig. 11 with the quantum cost and depth of 4 and with one garbage output [9].

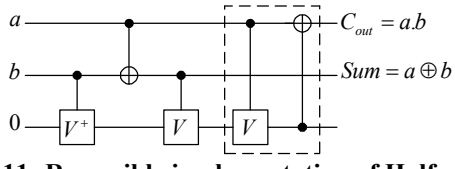


Fig. 11: Reversible implementation of Half-adder

III. Parallel Multipliers

There are two types of multipliers which are known as sequential and parallel multipliers. The first type iteratively computes the final product. It needs to use feedbacks and loops to compensate for the iterative portion. This design is too slow and not suitable for the reversible implementation. The second type (i.e., parallel multiplier), conventionally, consists of two main steps:

- Partial product generation
- Multi-operand addition

Algorithm 1 describes the $n \times n$ parallel multiplier.

Algorithm 1 (The $n \times n$ parallel multiplier):
Inputs: Two n-bit operands

$$X : x_{n-1} \ldots x_1 x_0 , \quad Y : y_{n-1} \ldots y_1 y_0$$

Output: A $2n$-bit product $P : p_{2n-1} \ldots p_1 p_0$.

I. Generate n partial products

a. $W^i : w^i_{n-1} \ldots w^i_1 w^i_0 \qquad 0 \le i \le n-1$

such that $w^i_j = x_j \times y_i$

II. Produce the final product $P = \sum_{i=0}^{n-1} W^i$.■

This paper is meant to propose a reversible implementation of the parallel multiplier following the approach described in Algorithm 1.

IV. The Proposed Reversible Multiplier

For the sake of simplicity the proposed reversible multiplier, likewise previous works is designed for the 4-bit input operands. However, it can be applied to any other $n \times n$ reversible multiplier. The operation of the 4×4 parallel multiplier is depicted in Fig. 12. We describe our proposed design in two subsections:

Part I: Partial Product Generation (PPG)
Part II: Multi-operand Addition (MA)

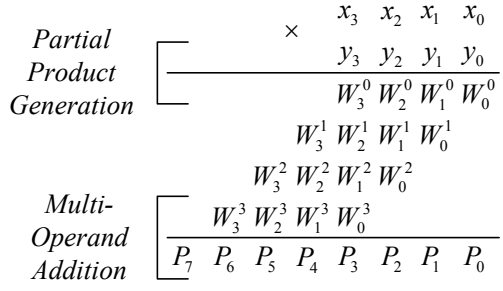


Fig. 12: The operation of the 4×4 parallel multiplier

A. Partial Product Generation

Partial products can be generated in parallel using 16 Peres gates (see Fig. 13).

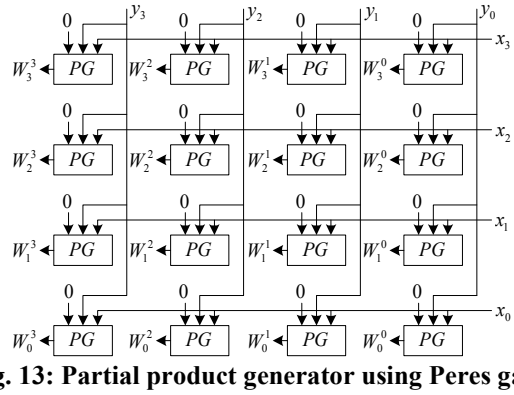


Fig. 13: Partial product generator using Peres gates

An important point that should be considered is that in an $n \times n$ parallel multiplier (in reversible logic) for generating partial products in parallel, n copies of each bit of the operands are needed. Therefore, some fan-out gates are needed. The number of fan-out gates needed for the reversible 4×4 multiplier is 24.

The quantum cost, depth and the number of garbage outputs for this part of circuit is shown in Table II.

B. Multi-operand Addition

As discussed in previous section, next step is an n-operand addition. To implement this part of circuit, we use carry save adder (CSA). As is shown in Fig. 12, in the

reversible 4×4 multiplier, four operands must be added to produce the final product. We use the CSA tree to reduce the four operands to two. Thereafter, a Carry Propagating Adder (CPA) adds these two operands and produces the final 8-bit product (see Fig. 14).

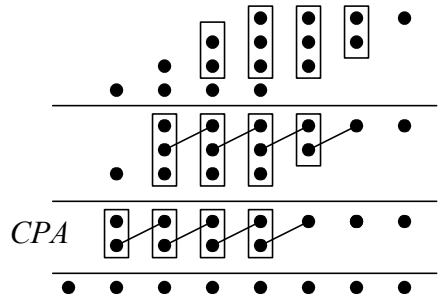


Fig. 14: Four-operand Addition (Dot notation)

The proposed four operand adder is shown in Fig. 15. We apply reversible FAs and HAs (see Figs. 10 and 11) to implement this part.

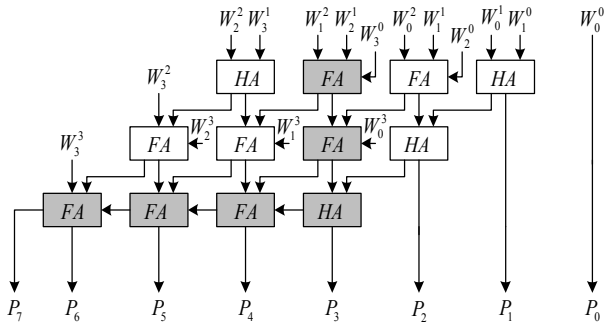


Fig. 15: Four-operand Addition (Block diagram)

Shaded blocks, in Fig. 15, indicate the critical path of this circuit. Third row in Table II shows the quantum cost, depth and the number of garbage outputs for this part of the circuit. However, in the last row of this table we obtain the total results of our proposed design.

Table II: Evaluation of the Proposed Design

	QC	Depth	GO
PPG	88	6	32
MA	64	23	20
Multiplier	152	29	52

V. Comparison with previous works

We have encountered three different designs for reversible multipliers in literatures where all of them, for the sake of simplicity, have implemented their design for a 4-bit multiplier. Therefore, here in this section, we compare our proposed multiplier with prior counterparts based on the 4-bit reversible multiplier. In order to have a reasonable comparison, first, we examine the detailed implementation of the previous works. Next, compare the proposed design based on the QC, depth, and the number of GOs with the previously mentioned cases as follows:

A. Reversible 4-bit multiplier of [14]

For the partial product generation phase of their multiplier, they used 24 gates of 2×2 Toffoli (TOF2), for preparing the essential fan-outs. Moreover, 16 Fredkin gates are used so as to generate the partial products.

For the multi-operand addition phase they used three 4-bit binary adders, where each of them is composed of 4 TSG, plus and extra TSG for the generation of the most significant bit of the final product.

By and large, the overall gate consumption of their reversible multiplier is equal to (24×TOF2) + (16×Fredkin) + (13×TSG). The overall critical path of their multiplier consists of two TOF2, a Fredkin gate, and seven TSG gates. Unfortunately, there is no reference for how the TSG can be implemented. Moreover, there is nothing mentioned in [14] about how a TSG can be built by means of elementary 2×2 reversible/quantum gates. For the sake of a fair comparison we assume the QC, Depth, and GO of a TSG gate as equal as that of a full-adder. Nevertheless, we believe that the QC, Depth, and GO of a TSG gate are much more than that of a FA.

B. Reversible 4-bit multiplier of [15]

For the partial product generation phase of their multiplier, like that of [14], they used 24 gates of TOF2 for preparing the necessary fan-outs. Moreover, 16 Peres gates are used in order to generate the partial products.

For the multi-operand addition phase they used 12 MKG gates where a MKG gate is a 4×4 reversible gate. Therefore, the overall gates used in their reversible multiplier is (24×TOF2) + (16×Peres) + (12×MKG). The overall critical path of their multiplier consists of two TOF2 gates, a Peres gate, and seven MKG gates. As the case for TSG, there is also no reference for the implementation of the MKG. Therefore, although we believe that the QC, Depth, and GO of a TSG gate is much more than that of a FA, we assume, for the sake of a fair comparison, the QC, Depth, and GO of a MKG gate the same as that of a full-adder.

C. Reversible 4-bit multiplier of [16]

This multiplier and that of [15] are somehow the same except for the multi-operand addition phase which is implemented in [16] by means of 8 HNG gates along with four Peres gates. This modification leads to the following critical path: (2×TOF2) + (2×Peres) + (6×HNG).

D. The proposed reversible 4-bit multiplier

In the proposed design for the partial product generation phase, like those of [15] and [16], we take advantage of the Peres gates in order to generate the partial products. For the multi-operand addition phase as is shown in Fig. 15, we use 8 full-adders and 4 half-adders. The critical path of this new design consists of two TOF2 plus a Peres gate for the partial product generation phase and 5 full-adders plus a half-adder for the multi-

operand addition phase. However, as is shown in Fig. 10, the implementation of the reversible FA by means of 2×2 quantum gates leads the designer to take advantage of fast inputs (e.g., C_{in}). This advantage gives rise to less depth of the overall design. For example, the effective depth in the critical path of the last three FAs of Fig. 15 is 3×3=9 instead of 3×5=15.

As a summery Table III compares the proposed reversible multiplier with those of the prior works. It is shown that the depth of the proposed reversible 4-bit multiplier is 27% less than that of the best of the previous works without increasing the quantum cost and the number of garbage outputs.

Table III. Comparison between different Reversible 4-bit Multipliers

	QC	GO	Depth
TSG [14]	182	58	42
MKG [15]	160	56	41
HNG [16]	152	52	40
Proposed	152	52	29

VI. Conclusions

Multiplier is a basic arithmetic cell in computer arithmetic units. Furthermore, reversible implementation of this unit is necessary for quantum computers. For this purpose, various designs can be found in the literature. We proposed in this paper a novel reversible multiplier with the less depth and no increase in quantum cost or the number of garbage outputs with respect to previous counterparts. In proposed design, partial products were generated using Peres gates. Next, the final product was obtained using a multi-operand adder including CSA tree and carry propagate addition.

The comparison between the proposed multiplier and those of the previous works showed that the depth of the new design is about 27% less than that of the best previous work.

The prospect for further research includes the reversible implementation of more complex arithmetic circuits such as Function Evaluation and multiplicative division circuits using this multiplier.

References

[1] R. Landauer, "Irreversibility and heat generation in the computing process", IBM J. Res. Develop., Vol. 5, pp. 183–191, July 1961.

[2] C. Bennett, "Logical reversibility of computation" IBM J.Res. Develop. Vol. 17, No. 6, pp. 525–532, November 1973.

[3] P. Gupta, A. Agrawal, N. K. Jha, "An Algorithm for Synthesis of Reversible Logic Circuits", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 25, No. 11, pp. 2317-2330, November 2006.

[4] P. Kaye, R. Laflamme, and M. Mosca, "An Introduction to Quantum Computing", *Oxford University Press*, January 2007.

[5] T. Toffoli, "Reversible computing", MIT, Tech. Rep., 1980.

[6] E. Fredkin and T. Toffoli, "Conservative logic", Int. J. Theoretical Physics, Vol. 21, No. 3/4, pp. 219–253, 1982.

[7] A. Peres, "Reversible logic and quantum computers", Physical Review A, Vol 32, pp. 3266-3276, 1985.

[8] D. Maslov, G. W. Dueck and D. M. Miller, "Simplification of Toffoli Networks via Templates". Proc. 16th Symposium on Integrated Circuits and Systems Design, pp. 53-58, September 2003.

[9] W. N. N. Hung, X. Song, G. Yang, J. Yang and M. A Perkowski, "Quantum Logic Synthesis by Symbolic Reachability Analysis", Proc. 41st annual conference on Design automation DAC, pp.838-841, January 2004.

[10] D. Maslov, C. Young, D. M. Miller, and G. W. Dueck, "Quantum Circuit Simplification Using Templates", Proc. Design Automation and Test in Europe (DATE), Vol 2, pp. 1208-1213, March 2005.

[11] X. Fei, D. Jiang-Feng, S. Ming-Jun, Z. Xian-Yi, H. Rong-Dian and W. Ji-Hui, "Realization of the Fredkin Gate by three Transition Pulses in a Nuclear Magnetic Resonance Quantum Information Processor", Chinese Physics Letters, Vol.19, pp. 1048-1050, 2002.

[12] D. Maslov and G.W. Dueck, "Improved Quantum Cost for n-bit Toffoli Gates", Electronic Letters, Vol. 39, No. 25, pp. 1790-1791, December 2003.

[13] J. A.Smolin and D. P.DiVincenzo, "Five Two-Bit Quantum Gates are Sufficient to Implement the Quantum Fredkin Gate", Physical Review A (Atomic, Molecular, and Optical Physics), Vol. 53, No. 4, pp. 2855-2856, April 1996.

[14] H. Thapliyal and M.B. Srinivas, "Novel Reversible Multiplier Architecture Using Reversible TSG Gate", Proc. IEEE International Conference on Computer Systems and Applications, pp. 100-103, March 2006.

[15] M. Shams, M. Haghparast and K. Navi, "Novel Reversible Multiplier Circuit in Nanotechnology", World Applied Science Journal Vol. 3, No. 5, pp. 806-810, 2008.

[16] M. Haghparast, S. Jafarali Jassbi, K. Navi and O. Hashemipour, "Design of a Novel Reversible Multiplier Circuit Using HNG Gate in Nanotechnology", World Applied Science Journal Vol. 3 No. 6, pp. 974-978, 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Improved Method of Highly Accurate Supply Detection using Bandgap Reference Circuit and Its Implementation in a Pseudo BiCMOS Process

Mustafa Ryadh[1], Khondker Zakir[2], Mohiuddin Hafiz[3], and A. B. M. H. Rashid[4]

[1, 2] Design Dept., Power IC Ltd. Dhaka, Bangladesh, [3] Dept of EECS, Hiroshima University, Hiroshima, Japan, [4] Dept of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology.
E-mail: s_ryadh@hotmail.com , abmhrashid@eee.buet.ac.bd

Abstract- This paper presents an improved method of highly accurate supply detection by using a bandgap circuit and its implementation in a comparatively inexpensive BiCMOS process. The process is the vanilla N-well complementary metal oxide semiconductor process technology with added deep N-well and P-well layers. The circuit consisting of a simple bandgap core with a resistance divider produces a supply detection method with high accuracy. In this method the low supply threshold can be set at any voltage according to the system requirement and the process, supply and temperature variation of this threshold voltage is very tight which produce an accurate low supply sense as a result. This circuit is implemented in 0.5 µm technology. The layout of the circuit also presented in this paper and the performance improved in layout are discussed. Simulation results and test data are given to prove the functionality and performance of the presented method.

I. Introduction

In an electronic application board all the circuit blocks do have an operating range for supply depending of the process mainly and the circuit design topology. When supply goes below the minimum operating voltage then circuit operation might not work. The digital blocks might produce wrong logics or in analog blocks any device can saturate and produce erroneous outputs. That is why the accurate low supply detection is so important for reliable operation of electronic applications. This paper presents an improved method of supply detection by using bandgap reference with produces a very accurate low supply threshold. In the modern electronics today reference circuits are present in the application ranging from purely analog, mixed-mode to purely digital circuits [1-2]. When the supply goes lower than minimum operating supply voltage then this detection circuit produces a logic which can be used to disable other circuit blocks to prohibit them from producing incorrect output.

II. Bandgap Reference

The bandgap reference block constitutes a precise band-gap core. The principle of the band-gap core relies on two groups of transistors running at different emitter current densities [3-4]. This difference in the current densities cause a difference between the base-emitter voltages, ΔV_{BE} From the large signal behavior of the bipolar transistor and neglecting the base current (for high β_F) of it, we have the following relationship

$$V_{BE1} = \left(\frac{KT}{q}\right)\ln\left(\frac{I_1}{I_{S1}}\right) \quad (1)$$

$$V_{BE2} = \left(\frac{KT}{q}\right)\ln\left(\frac{I_2}{I_{S2}}\right) \quad (2)$$

and $I_{S1} \alpha A_1 \quad \& \quad I_{S2} \alpha A_2$

where, $I_1 = I_{C1} \cong I_{E1}$
$\& I_2 = I_{C2} \cong I_{E2}$,
$I_S => saturation\ current$,
$A => base-emitter\ junction\ area\ of\ a\ bipolar\ transistor$. (3)

$$\therefore \Delta V_{BE} = V_{BE2} - V_{BE1}$$
$$= \left(\frac{KT}{q}\right)\ln\left(\frac{A_1}{A_2}\right)$$
$$= V_T \ln\left(\frac{A_1}{A_2}\right) \quad (4)$$

for $I_1=I_2$, i.e in the balanced condition of the band-gap core.

$V_T=\left(\frac{KT}{q}\right)$,is called thermal voltage.

The core, used in the design, has the emitter area ratio of 8:1, i.e. $A_1:A_2=8:1$

III. Circuit Realization

The circuit shown in Figure 1 is basically a bandgap circuit used for supply detection. Here first the supply is taken through a resistance divider to a band gap circuit in such a way that when the supply surpasses the threshold then the band gap tap point crosses bandgap voltage 1.23V. We know that from the theory of band gap reference when the bandgap voltage is less than the reference settling voltage 1.23V then collector current in 8x device is more than the current in 1x bipolar device.

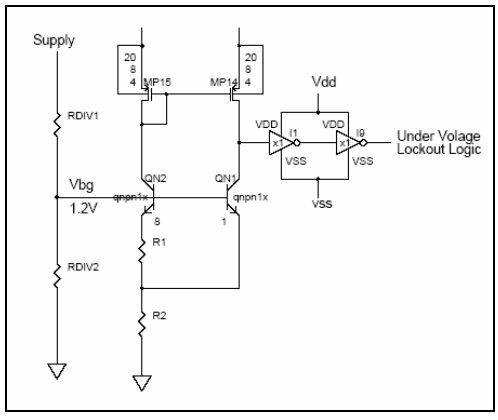


Fig. 1 Circuit implementation of supply detection method using bandgap circuit

In the circuit shown in the Fig. 1 we can see that how the bipolar transistor currents are compared. When the bandgap core bipolar devices' base voltage are lower than regulation voltage or settling voltage then the 8x device current is higher than 1x device. This is because in the 8x device the collector current becomes divided into 8 devices which makes the base-emitter junction voltage of 8x device more lower than 1x device base-emitter voltage. So the voltage drop just below the emitter node resistance in 8x device will be higher that of the 1x device. These two currents are compared by mirroring the 8x device collector current which makes the bipolar device QN2 and PMOS MP14 current in Fig. 1 equal to each other. So the output logic from this current comparison gives a high logic at output. As soon as the reference tap point voltage crosses 1.23V the collector current of 1x device gets higher than the collector current of the 8x bipolar device which makes the output logic low.

IV. Accuracy

The accuracy of the threshold voltage is a very important issue to be considered. In this method the accuracy of the low supply detection mainly depends on the variation of the resistance divider and the variation of the band gap circuit. In the IC any performance parameter can vary mostly for the process variation of die to die, for supply variation, temperature variation in the die and part to part variation in the die because of oxide thickness mismatch, crystal defect or scattering of doping atoms. The process and temperature effect can be minimized for the resistance divider circuit by using same type of resistance above and under the divided point. In the circuit if the ROV1 is Rnwell, then ROV2 should be Rnwell type resistance. In that case the temperature coefficient and process variation of both the resistances will match which will minimize the variation. To reduce the part to part random mismatch we adopt inter-digitization matching technique in layout where we matched the resistance segments very closely. Then we need to focus on the variation of the reference voltage which consists of the same kinds of variations we saw in case of resistance divider. The process variation of the band gap reference voltage is minimized by using the same type of resistances in the band gap core and by keeping the bipolar devices ratio high [5] which we kept here 8. We have used the PMOS with large L to reduce the channel length modulation effect to reduce the supply variation of the band gap core. Another major issue is the temperature variation of the reference circuit.

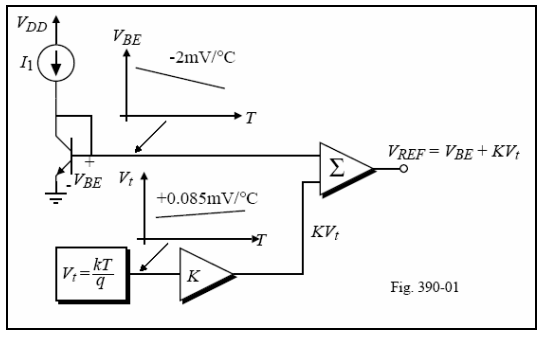


Fig. 2 Temperature coefficient of bandgap reference

The temperature coefficient of the bandgap reference is combination of V_{BE} and KVt. As we know that, base emitter forward voltage of bipolar device decreases with temperature increase where as the KVt voltage will increase with temperature. So we add V_{BE} and KVt voltage to create bandgap reference such a way that the amount of voltage decreases because of VBE the same amount of voltage we will increase with KVt increase.

V. Layout

First of all we consider the layout of resistance divider very carefully. Inter- digitized matching of the resistance segments ROV1 and ROV2 are done in the layout. This matching will help to reduce the part to part variation of the supply voltage detection threshold voltage in the die.

Fig. 3 Inter-digitization of resistance divider in layout to reduce offset in the circuit

For the reference voltage accuracy several works are done. To reduce the offset the bipolar devices are matched very well by applying common centroid matching technique in layout of band gap reference circuit. We can see the matching technique in Fig. 4 where we have places the QN1 bipolar device in the centre and place QN2 which is a 8x device surrounding QN1 so that they match very closely there exists no offset in bipolar devices.

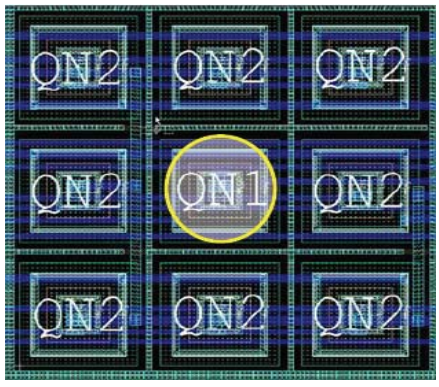


Fig. 4 Common centroid matching technique applied in bandgap core to reduce offset

The whole layout of the highly accurate supply detection circuit is shown here. The total area has been optimized by grouping and doing the layout of all the NMOS and PMOS of the circuit together. The cross coupling matching of the MOSFETS reduce the offset in the circuit.

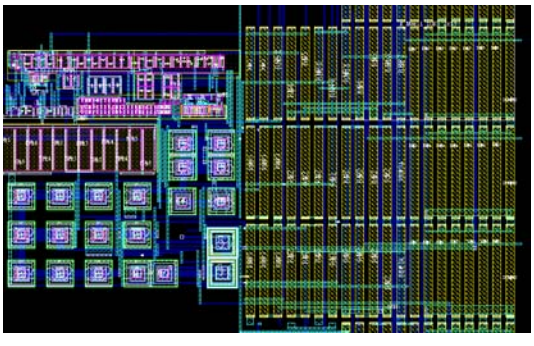


Fig. 5 Layout of the supply detection circuit with bandgap core and resistance divider

VI. Results and Discussion

The method presented in this paper is implemented in 0.5 µm technology BiCMOS process. We set the low supply threshold in 1.8V which is shown in the simulation result in Fig. 6. in this simulation we ramp the supply voltage from zero volt to 3V. We can see in this simulation that the supply detection logic gives a logic high logic output when the supply is 1.8V. we have given the test data in Fig. 7. we can see that in the test result the output logic becomes high when supply voltage Vsupply becomes

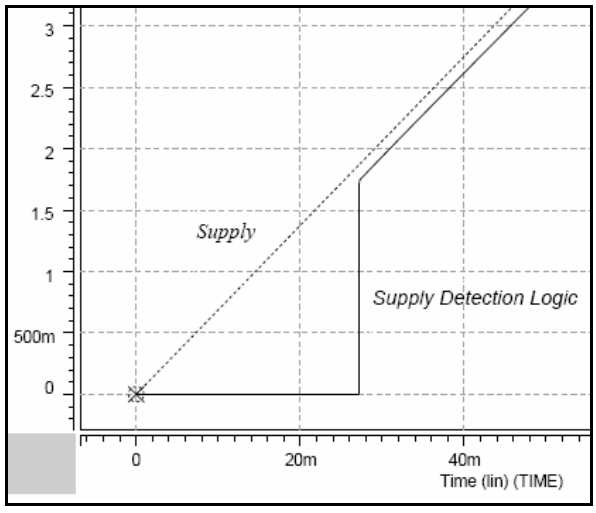


Fig. 6 The simulation result of the supply detection circuit

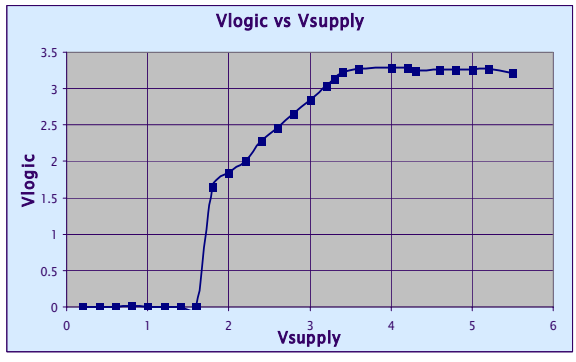


Fig. 7 Test data of supply detection circuit

1.8V. After that the level of the high logic voltage will follow the supply which we have seen in the Fig. 7. As we said earlier the variation of the detection threshold voltage will be same as the variation of the bandgap reference voltage which we can see in the Fig. 8. Here we can see that the variation of the threshold voltage is only 1.4% of the threshold voltage level over the temperature range from -40 to 120˚C.

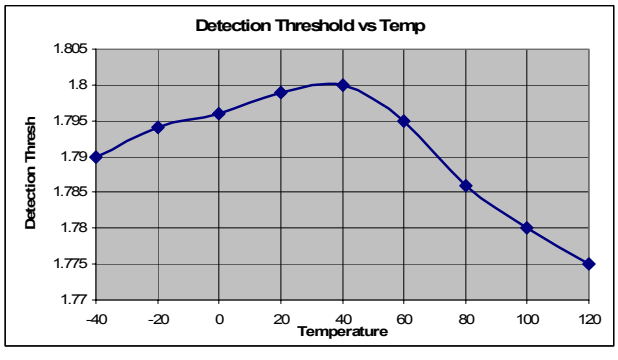


Fig. 8 Test data of supply detection threshold voltage with temperature

From the test results and the simulation results we can see that the test result in Fig. 7 correlates the simulation result in Fig. 8 very closely. The added test result in Fig. 8

shows that the threshold variation is very small over wide temperature range which helps the precision of the supply detection circuit.

VII. Conclusion

An improved method of low supply detection is presented in this paper by using the bandgap reference core and resistance divider which is very easy to implement. The advantage of using this method for supply detection is that the threshold variation over process, supply and temperature is very low which makes the supply detection very accurate. The layout of the circuit is also shown and the performance like offset is also improved in the layout. The circuit implemented in 0.5 µm technology process and the test data are given which proves the functionality of the method presented in this paper.

References

[1] B. Razavi, *Design of Analog CMOS Integrated Circuits.* New york, NY: 2001, ch 11.

[2] D. Hilbiber, "A New Semicondictor Voltage Standard," *IEEE J. of Solid-State Circuits,* vol. 8, pp. 222-226, June 1973.

[3] Robert A. Pease, "The Design of Band-Gap Reference Circuits: Trials and Tribulations," *IEEE Proc. of the 1990 Bipolar Circuits and Technology Meeting*, Minneapolis, Minnesota , Sept. 1990

[4] J. Michejda and S.K. Kim, "*A Precision CMOS Bandgap Reference*," IEEE J. Solid-State Circuits, vol. SC-19, pp. 1014-1021, December 1984.

[5] T. Brooks and A.C Westwisk, "*A low power differential CMOS bandgap Reference*", ISSCC Dig. Of Tech. Papers, pp 248-249, Feb 1994

[6] K.E. Kujik, "*A Precision Reference Voltage Source*," IEEE J. Solid-State Circuits, vol. SC-8, pp. 222-226, June 1973.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

RECONFIGURABLE MONOCYCLE PULSE BASED UWB TRANSMITTER IN 0.18 μM CMOS FOR INTRA/INTERCHIP WIRELESS INTER CONNECT

S. M. Salahuddin[1], Salahuddin Raju[2], P. K. Saha[3]

[1]Department of Electrical and Electronic Engineering.
East West University, Bangladesh.
[2]Department of Electrical and Electronic Engineering.
American International University-Bangladesh, Bangladesh.
[3]Department of Electrical and Electronic Engineering.
Bangladesh University of Engineering and Technology, Bangladesh.
Email: [1]sms@ewubd.edu, [2]raju@aiub.edu, [3]sahapk@eee.buet.ac.bd

Abstract - Ultra wide band (UWB) communication has got special attention as a promising radio technology for networks delivering extremely high data rate with relatively low power consumption at short distance. In this paper we are reporting a new Monocycle Pulse (MCP) generation technique which can be used in a single chip implementation of Impulse Radio based ultra wideband (UWB) transceiver architecture for high speed intra/interchip communication. The proposed MCP based UWB transmitter is implemented in 0.18μm CMOS process. The performance and simulation results of the circuit is presented in this paper. The MCP is found to be reconfigurable; therefore it is possible to control the data rate.

I. INTRODUCTION

To provide communication among distant chips with minimum latency, proper interconnection is required. During the past few decades, with the advancement of transistor scaling, interconnect scaling has augmented the distributed resistance-capacitance product which engenders larger latency for a given interconnect length. Scaling of interconnects serves to reduce cost but increases latency and energy dissipation. These increase result from relatively larger average interconnect lengths and larger die sizes for successive generation. So, the chief obstacle in the route of good performance and reduced energy dissipation in future ULSI is undoubtedly the interconnect problem.

The time delay in case of global interconnects increases by a square of the factor *m* where devices and interconnects are scaled down by a factor *m*. But the local interconnect delay reduced only by a factor of m which again proves that global interconnect is the most dominating one [1]. Presently 60%-70% of the clock cycle is consumed by interconnect delays which is rising with the ever decreasing device size. Sometimes, to reduce the interconnect delay repeater is inserted between the route. This brings increased power consumption and somewhat reduction of delay. Using thicker and wider metal wires, delay can be also reduced but the problem is due to high performance System-on-Chip, optimally buffered minimum size of global wire is decreasing sharply with technology scaling. Furthermore, wider wires create signal integrity problem, cross talk, delay instability and introduce inductance due to enormous speed. On the contrary, when we increase the spacing to ameliorate the performance, it causes high capacitance and consequently high parasitic effects. To solve the future interconnect problem wireless interconnect technique using integrated antenna was first proposed in [2].

Ultra wide band (UWB) communication systems use signal with bandwidth that is larger than 20 % of center frequency or wider than 500 MHz [3]. Of the different techniques for UWB, carrier free impulse radio based UWB (IR-UWB) technology uses very short low duty cycle pulse which has wide bandwidth and high center frequency [4]. Conventional IR-UWB system use Gaussian monocycle pulse (GMP) which has no dc component. GMP shape is changed due to the differentiation effect of antenna [4]**.** As a result the received signal will be second derivative of the GMP. Thus the receiver needs the second derivative of GMP as template signal which is difficult to generate. In this paper we propose a new monocycle pulse (MCP) which can be generated from a sinusoidal signal train. Twice differentiation of the sinusoidal signal due to antennas keeps the same wave shape at the receiver side. The proposed MCP shows UWB characteristics and could be reconfigured for the desired MCP center frequency and pulse repetition rate.

II. System Level Simulation of MCP

The schematic of proposed MCP based UWB transceiver is shown in Fig 1. and system level simulation of the proposed transceiver is found in [5]. In this paper MCP generation process and circuit is well described.

A schematic block diagram of the proposed MCP generation technique is shown in Fig 2. Here voltage controlled oscillator (VCO) is used to generate sinusoidal signal of desired frequency. Output of the VCO is then fed into a comparator, which output works as a clock of a 3 bit

ring counter. Ring counter generates square wave which duty cycle is equal to the desired MCP's width. One of the ring counter's outputs is then multiplied by the VCO's output to generate MCP. The only complexity of proposed MCP generation is the synchronization of ring counter and VCO due to the inherent circuit delay. Thus VCO output is required to be delayed prior to the multiplication with the ring counter's output. VCO output can be delayed by using a voltage controlled delay circuit or a properly designed transmission line section.

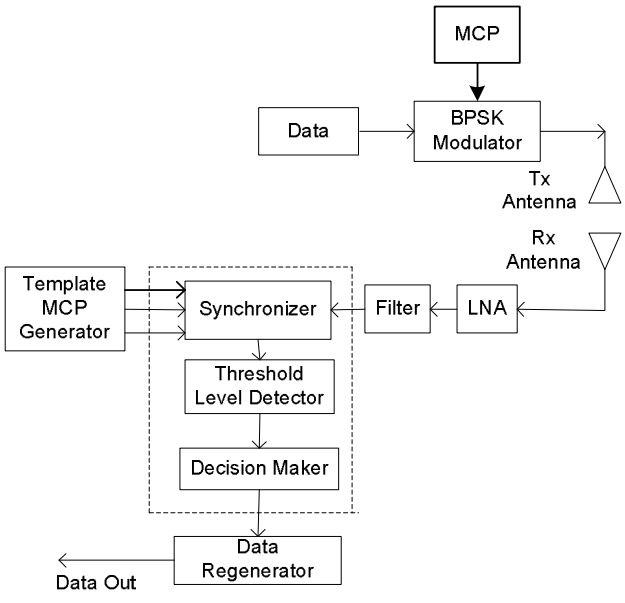


Fig. 1 Block diagram of MCP based Transceiver

The simulation result is shown in Fig. 3. Fig 3(a) shows the 3 GHz sinusoidal VCO output. The ring counter output is given in Fig 3(b). Fig 3(c) shows new MCP which has width of 330 ps and pulse repetition rate of 1 GHz. Thus a maximum data transmission rate of 1 Gbps could be achieved. The generated MCP has also a good symmetry and no ringing. The Fourier transform of the MCP as shown in Fig. 4 depicts that the new MCP has a center frequency of 3 GHz and a -3 dB bandwidth of 2.2 GHz, 73.33% of center frequency which confirms the ultra wideband characteristic of the proposed MCP. The center frequency which is reciprocal of the MCP's width could be changed by changing the VCO's control voltage. This makes the reconfigurable MCP generation.

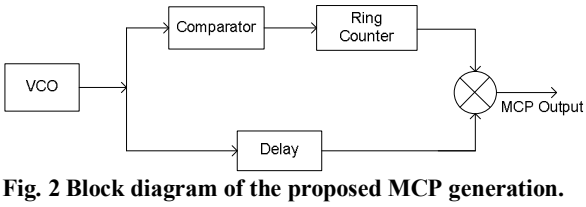


Fig. 2 Block diagram of the proposed MCP generation.

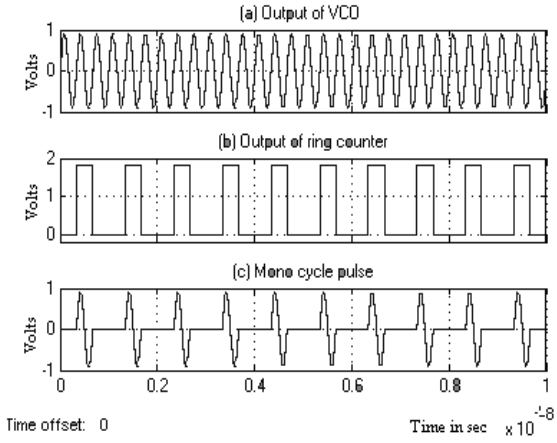


Fig. 3 MATLAB output at different stages of MCP generator circuit

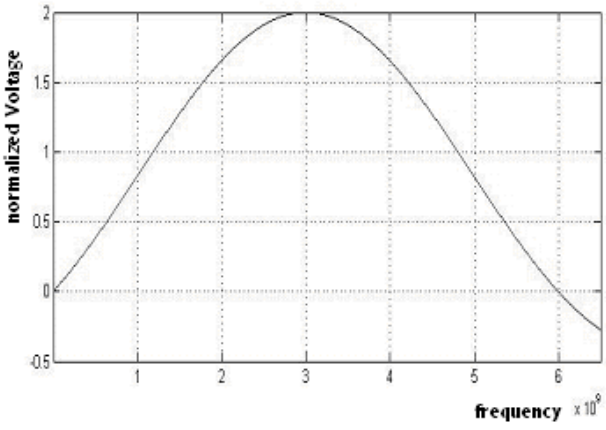


Fig 4. New MCP in frequency domain

III. Circuit Level Simulation

A. Voltage Controlled Oscillator:

The main block of the designed MCP is the voltage controlled oscillator (VCO) shown in Fig 5 . The VCO is fully differential to minimize the role of noise associated with the system inherently. A varactor diode is used as voltage controlled capacitor to change the operating frequency. Gate length of both nMOS and pMOS set to 0.18µm and width 5.4µm for nMOS and 10.8µm for pMOS. TSPICE simulation Output of VCO is shown in Fig 6.

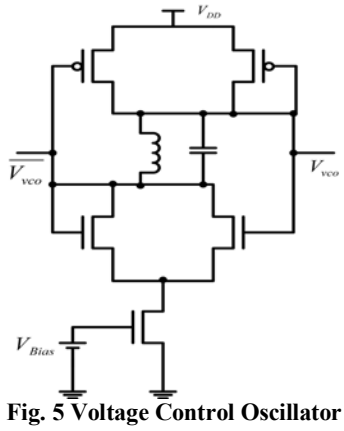


Fig. 5 Voltage Control Oscillator

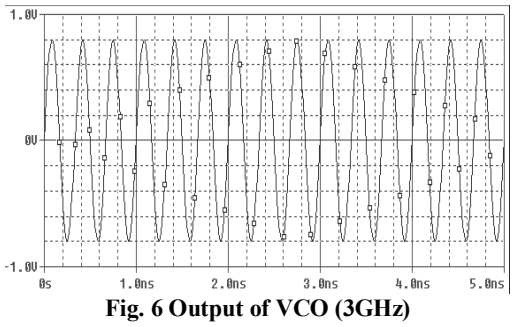


Fig. 6 Output of VCO (3GHz)

B. Comparator:

In this design, a single ended single stage CMOS comparator is used to convert the sinusoidal signal to digital signal which is used as clock for ring counter. Fig. 7 shows the comparator circuit. For buffering and minimize the rise time and fall time output of comparator is directly connected to a buffer.

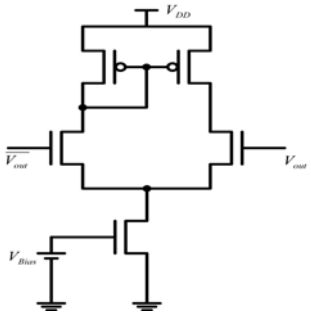


Fig. 7 Comparator

C. Ring Counter:

The buffered signal is used as clock in ring counter. A three bit ring counter is designed using D flip flop which having the provision of set and preset and these options are active low. Fig 8 shows the ring counter's building block.

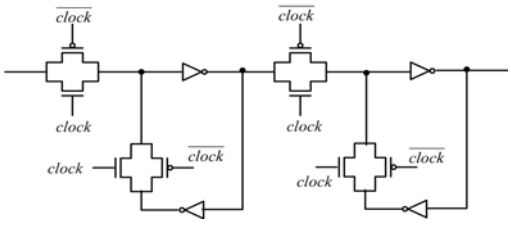


Fig 8. D Flip-flop used in ring counter

Gate pulse generated from the ring counter is used to activate the transmission gate to control the delayed version of sinusoidal signal through transmission gate. Fig 9, Fig 10 and Fig 11 shows the output of transmission gate having different center frequency. For each case peak to peak voltage swing is 1.6 volts. These figures depict that the generated MCP has also a good symmetry and 6.75% ringing due to the following edge of gate pulse used for transmission gate.

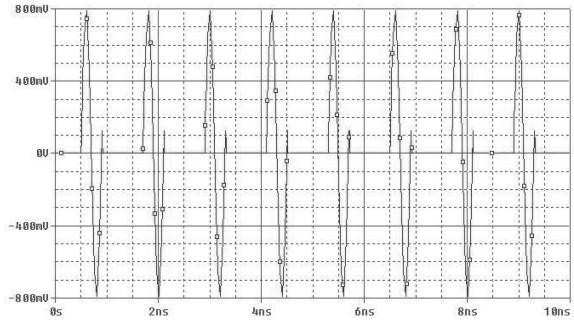


Fig 9 Center frequency 2.5 GHz

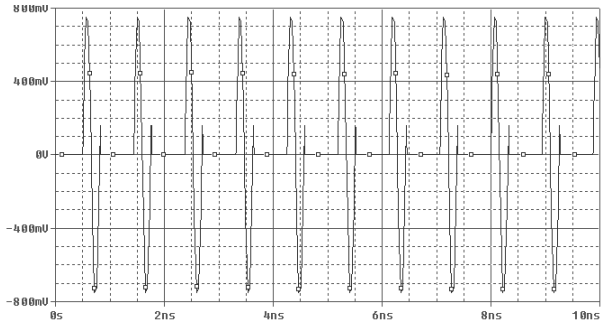


Fig 10 Center frequency 3.5 GHz

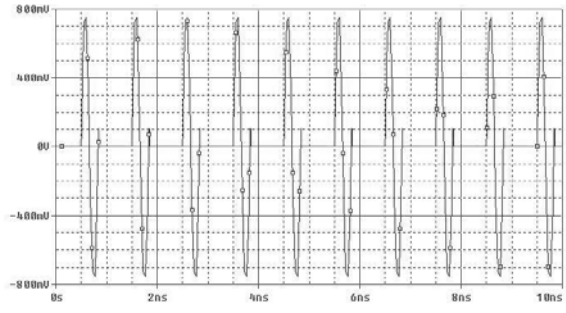


Fig 11 Center frequency 3 GHz

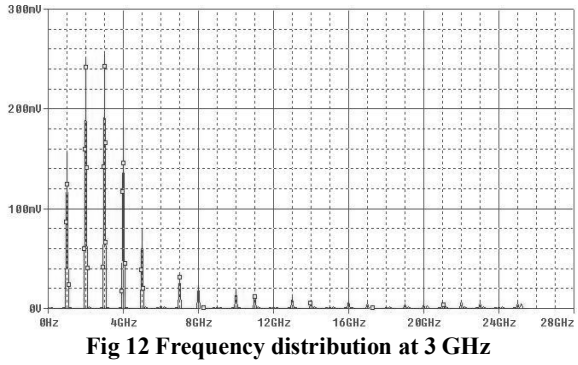


Fig 12 Frequency distribution at 3 GHz

IV. MCP Performance Comparison

Comparison of MCP performance obtained from system level simulation by MATLAB with that obtained from TSPICE simulation is given in Table 1. The table shows a very good agreement of MCP performance obtained from both level of simulation.

Comparison of overall performance of MCP, Gaussian monocycle pulse (GMP) and step recovery diode based GMP is given in table 2. The table 2 depicts that the new MCP shows better performance as compared with that of the other two techniques [6],[7].

Table 1: MCP Performance Comparison

	System Level	Circuit Level
Ringing level	No ringing	-24dB
Symmetry	100%	100%
Pulse width	Reconfigurable	Reconfigurable
Center frequency	Reconfigurable	Reconfigurable
Peak to Peak amplitude	1.8V	1.6V
3 dB bandwidth	73.33%	56%

Table 2: MCP and GMP Performance Comparison

	GMP	SRD	MCP
Ringing level	-20.2dB	-17dB	-24dB
Symmetry	Not well balanced	Not well balanced	100%
Pulse width	0.28ns	0.30ns	Reconfigurable
Center frequency	3.6 GHz	--------	Reconfigurable
Peak to Peak amplitude	0.123V	0.2V	1.6V

V. Conclusion

The MCP circuit is implemented using TSMC 0.18µm CMOS process. It shows well symmetry and low ringing level. As the proposed MCP is reconfigurable and generated from sinusoidal signal, data rate can be varied by changing VCO's reference voltage.

References

[1] W. J. Dally, "Interconnect Scaling - the real limiter to high performance ULSI" in IEDM Tech. Dig., 1995 pp. 15-17

[2] B A Floyd, C.M. Hung and Kenneth K. O ." A 15 GHz Wireless interconnect implemented in a 0.18µm CMOS technology using integrated transmitters, receivers and antenna ," *IEEE J. Solid-State Circuits*, vol. 37, no. 5, pp. 543-552, May 2002.

[3] Federal Communication Commission (FCC), Ultra – Wideband First report and Order, Feb. 2002.

[4] Win M. Z. and Scholtz R. A, "Impulse radio: how it works", IEEE Communication Letter, 1998,2, pp. 36-38

[5] Salauddin Raju, S. M. Salahuddin, Md Shahidul Islam , P.K. Saha and A. H. M. Zahirul Alam "DSSS IR UWB Transceiver for Intra/ Inter Chip Wireless Interconnect in Future ULSI using Reconfigurable Monocycle Pulse" *Proceedings of the International Conference on Computer and Communication Engineering 2008*, May 13-15, 2008 Kuala Lumpur, Malaysia pp. 306-309.

[6] Pran Kanai Saha, Nobuo Sasaki and Takamaro Kikkawa, " A Single-chip Gaussian Monocycle Pulse Transmitter using 0.18 µm CMOS Technology for Intra/Interchip UWB communication", *in Digest of Technical Papers, IEEE Symposium on VLSI Circuits*, Hondula, HI, USA, 11-15 June, 2006, pp. 252-253.

[7] Jeongwoo Han and Cam Nguyen, "A New Ultra-Wideband, Ultra-Short Monocycle Pulse Generator With Reduced Ringing", *IEEE Microwave and wireless components letters*, vol. 12, no 6, 2002, pp 206-208.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Fully Digital Nonlinear, High-speed Rank Order Filter in 0.18µm CMOS Technology

George John Toscano[1] and Pran Kanai Saha[2]
[1] American International University Bangladesh
[2] Bangladesh University of Engineering and Technology
ggtoscano@gmail.com, sahapk@eee.buet.ac.bd

Abstract - Some efficient techniques to realize modular, high-speed digital Rank Order Filter (ROF) are presented in this paper. Using the proposed digital filters, it is possible to find the element of a certain rank (Maximum, Minimum and Median) in a given sequence of *N* elements in each window in *M* steps, where *M* is the number of bits used in binary representation for the elements of the sequence. A bit-level algorithm by Kar and Pradhan has been modified in this work to implement the proposed Filters. The modified algorithm is implemented using *N* number of identical Bit Update Circuit (BUC) along with other logic gates in 0.18µm CMOS technology. The proposed ROF circuits are simulated using HSPICE. Simulation results depict the good performance of the filters in terms of speed and power. The ROF algorithm is also tested in FPGA. The post fit simulation output using Xilinx is presented in this paper. In both simulations the ROF's performance is gratifying.

I. INTRODUCTION

Rank Order Filter is a nonlinear filter widely used in digital signal and image processing. It involves selecting a sample with the specified rank from a one dimension or multidimensional window of samples. Special cases of Rank Order Filters are median, minimum and maximum filters, where the outputs are the median, the minimum and the maximum value.

Median filtering is used to filter impulse noise without blurring the edges of images and for background signal level estimation. Maximum and Minimum filters are used for application such as robot vision and pattern recognition. To realize median filters array architectures [1,2] have a limited circuit complexity, but often have a large sampling period due to the sequential execution of several operations in the same sampling period. Sorting network-based architectures [3,4] are inherently pipelined, but they require a large number of compare-swap units. The proposed rank order filters are very fast in operation and requires less Si area. The proposed Rank Order Filters are also very simple and modular in architecture and function of each part of the filter is well defined.

II. BIT-LEVEL SELECTION ALGORITHM

A modified Rank selection algorithm based on the bit-level algorithm reported in [5] is used to design the proposed Rank Order Filter. Let W is a window of size N. W can be expressed as $W = \{x_1, x_2, \ldots\ldots, x_N\}$. Let M is the size of a sample in the sequence in bits. Then each sample x_j can be represented as:

$$x_j = b_j^{M-1} b_j^{M-2} \ldots\ldots b_j^0$$

Where b_j^k is the $(k+1)$-th bit of x_j with a weight of 2^k.
An element of rank r is the rth smallest element in the window of N elements. Let $Y = y^{M-1} y^{M-2} \ldots\ldots y^0$ be the rth rank-order element in W. Based on [5] a modified bit level algorithm to find Y is given below.

1. Set $k=M-1$
2. Set $B_j^k = b_j^k$ for $j=1,2,\ldots\ldots N$
3. Find sum $S^k = \sum_{j=1}^{N} B_j^k$
4. If $S \geq N-r+1$ then
   Set bit $y^k=1$
   Else Set bit $y^k=0$
5. For j=1 to N
   If $y^k=1$ and $B_j^k=0$ then
   Reset bits $B_j^{k-1}, B_j^{k-2}, \ldots\ldots B_j^0$ to 0.
   If $y^k=0$ and $B_j^k=1$ then
   Reset bits $B_j^{k-1}, B_j^{k-2}, \ldots\ldots B_j^0$ to 1.
   Else Set $B_j^{k-1} = b_j^{k-1}$
6. Set $k=k-1$
7. Repeat steps 3 to 6 until $k < 0$
8. Output $Y = y^{M-1} y^{M-2} \ldots\ldots y^0$ as rank-order element.

To implement the following algorithm N number of identical Bit Update Circuit, (Here N is the number of samples in each window), a Parallel Counter Circuit and a Comparator Circuit is required as shown in figure 1 and described in [6].

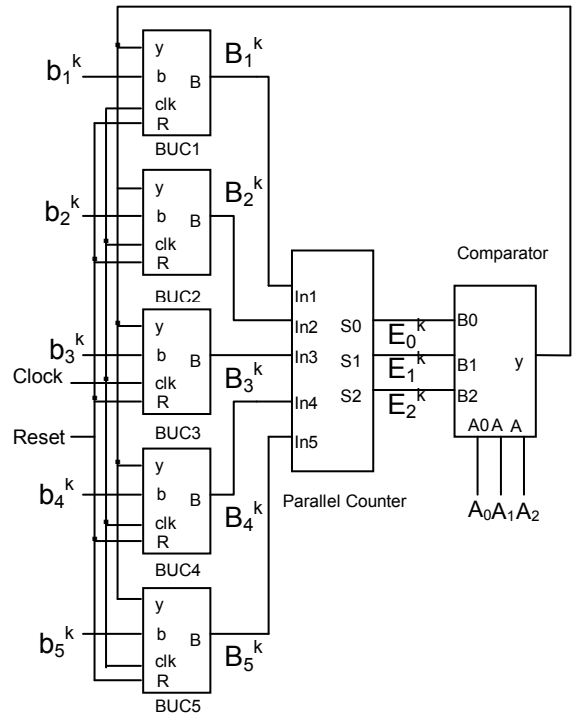


Figure 1. Reconfigurable Rank Selection Circuit For window size $N = 5$.

Here in figure 1, b_1^k, b_2^k, b_3^k, b_4^k, b_5^k are the input samples and B_1^k, B_2^k, B_3^k, B_4^k, B_5^k are the output from the BUCs. E_0^k, E_1^k, E_2^k are the output of the Parallel Counter Circuit, A_0,A_1,A_2 are the external input to the Comparator Circuit on which the output of the Rank Selection Circuit y^k depends.

III. MAXIMUM FILTER:

The parallel counter and comparator stage of the reconfigurable rank order filter as shown in figure 1, can be eliminated by an N input OR gate to implement the maximum filter. The block diagram of the Maximum filter is shown in figure 2.

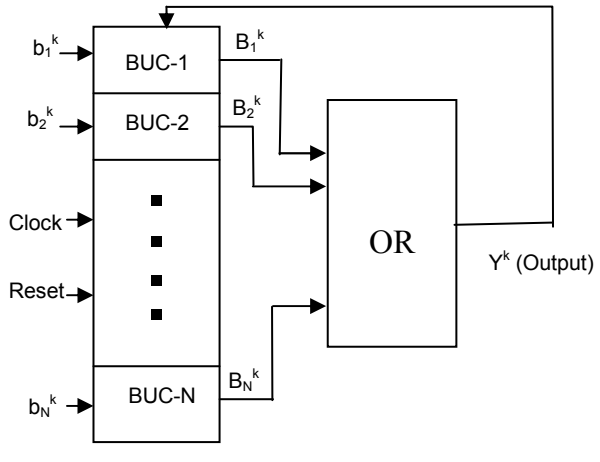


Figure 2. Block diagram of the Maximum Filter

So the architecture of the maximum filter becomes very simple. A reconfigurable rank order filter with window size N acts as a maximum filter when the value of r is N or the input to the comparator circuit is $A = N\text{-}r+1 = 1$. That means if the number of 1's at the output of the BUCs (S^k) is greater than or equal to 1, the output of a Maximum filter will be '1'. This is the logic for OR gate. So the parallel counter and comparator stage of the rank order filter can be eliminated in a maximum filter by an N input OR gate. The output of the Maximum filter can be written as:

$$Y^k = B_1^k \, OR \, B_2^k \, OR \, B_3^k \, OR \ldots\ldots OR \, B_N^k \qquad (1)$$

The internal structure of BUC is shown in figure 3. The total number of gates required to implement the BUC is 9. N number of BUC operates in parallel to modify the incoming bits. N number of MSBs from N samples are input to the BUCs parallely during the first clock cycle. In next clock cycle, N bits next to MSB, from N samples enters the BUCs. In M steps all the bits from all the samples from a window enters the BUCs. For each input b_j^k the BUC gives modified output B_j^k.

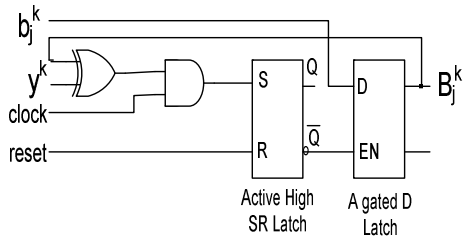


Figure 3. Bit Update Circuit (BUC).

The output of the BUC can be written as:

$$B_j^k = b_j^k \overline{Q}_j + B_j^k Q_j \qquad (2)$$

where, $$Q_j = Q_j + (b_j^k \oplus y^k) \qquad (3)$$

An illustration of the algorithm for a Maximum filter with window size N=5, W= { 107,22,110,81,62}, and M = 8 is given in figure 4.

k	b_1^k (B_1^k)	b_2^k (B_2^k)	b_3^k (B_3^k)	b_4^k (B_4^k)	b_5^k (B_5^k)	y^k
7	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
6	1 (1)	0 (0)	1 (1)	1 (1)	0 (0)	1
5	1 (1)	0 (0)	1 (1)	0 (0)	1 (0)	1
4	0 (0)	1 (0)	0 (0)	1 (0)	1 (0)	0
3	1 (1)	0 (0)	1 (1)	0 (0)	1 (0)	1
2	0 (0)	1 (0)	1 (1)	0 (0)	1 (0)	1
1	1 (0)	1 (0)	1 (1)	0 (0)	1 (0)	1
0	1 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0
Decimal	107	22	110	81	62	110

Figure 4. Computation of rank-order element for Maximum Filter.

The output of the BUCs are given inside the bracket as shown in figure 4. The output of the Maximum filter is found to be $(110)_{10}$. Bits modified by the Bit Update Circuit is shown in shaded cells of the figure 4.

The output of the post-fit VHDL model of the Maximum filter with target device Spartan2E-xc2S50e is shown in figure 5. The circuit is found to operate smoothly at a clock speed of 10 MHz.

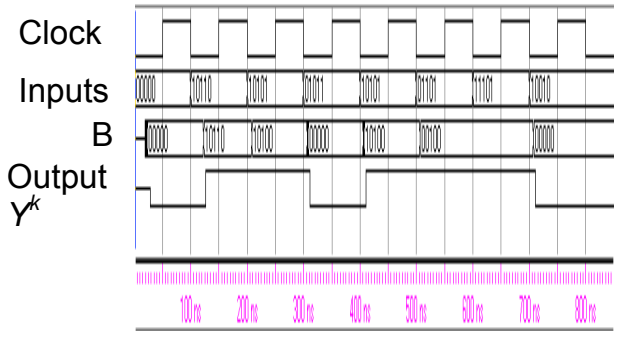


Figure 5. Post-fit simulation output of the Maximum Filter for the inputs as shown in figure 4.

In figure 5, '*B*' indicates the output of the BUCs.

IV. MINIMUM FILTER

The parallel counter and comparator stage of the reconfigurable rank order filter as shown in figure 1, can be eliminated by an *N* input AND gate to implement the Minimum filter. The block diagram of the Minimum filter is shown in figure 6.

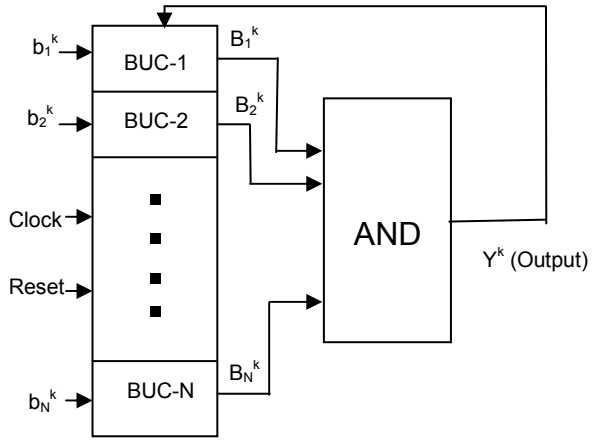


Figure 6. Block diagram of the Minimum Filter

A reconfigurable rank order filter with window size *N* acts as a minimum filter when the value of *r* is 1 or the input to the comparator circuit is $A = N\text{-}r+1 = N$. That means if all the output of the BUCs are HIGH, the output of the minimum filter will be '1'. This is the logic for AND gate. So the parallel counter and comparator stage of the rank order filter can be eliminated in a minimum filter by an *N* input AND gate. The output of the Minimum filter can be written as:

$$Y^k = B_1^k \; AND \; B_2^k AND \; B_3^k \; AND \ldots\ldots\ldots AND \; B_N^k \quad (4)$$

An illustration of the algorithm for a Minimum filter with window size N=5, W= { 107,22,110,81,62}, and M = 8 is given in figure 7. The output of the BUCs are shown inside the bracket. The output of the Minimum filter is found to be $(22)_{10}$. Bits modified by the Bit Update Circuit is shown in shaded cells.

k	b_1^k (B_1^k)	b_2^k (B_2^k)	b_3^k (B_3^k)	b_4^k (B_4^k)	b_5^k (B_5^k)	y^k
7	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
6	1 (1)	0 (0)	1 (1)	1 (1)	0 (0)	0
5	1 (1)	0 (0)	1 (1)	0 (1)	1 (1)	0
4	0 (1)	1 (1)	0 (1)	1 (1)	1 (1)	1
3	1 (1)	0 (0)	1 (1)	0 (1)	1 (1)	0
2	0 (1)	1 (1)	1 (1)	0 (1)	1 (1)	1
1	1 (1)	1 (1)	1 (1)	0 (1)	1 (1)	1
0	1 (1)	0 (0)	0 (1)	1 (1)	0 (1)	0
Decimal value	107	22	110	81	62	22

Figure 7. Computation of rank-order element for Minimum Filter.

The output of the post-fit VHDL model of the Minimum filter with target device Spartan2E-xc2S50e is shown in figure 8. The circuit is found to operate smoothly at a clock speed of 10 MHz.

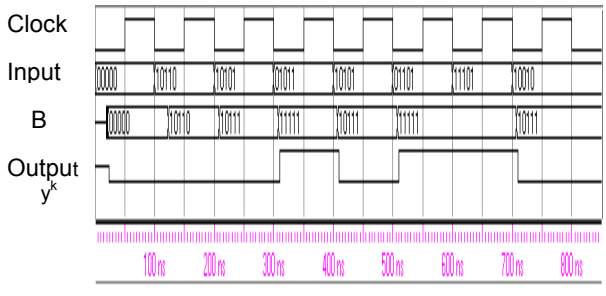


Figure 8. Post-fit simulation output of the Minimum Filter for the inputs as shown in figure 7.

In figure 8, '*B*' indicates the output of the BUCs.

V. MEDIAN FILTER

To obtain median as output from a reconfigurable Rank Order Filter the input to the comparator circuit of the Reconfigurable Rank Order Filter should be $A = (N+1)/2$. But implementing a Median filter in this way is not so efficient. Some portion of the reconfigurable Rank Order Filter can be eliminated in median filter and hence less Si area and power will be needed to operate the circuit. The circuit minimization of Median Filter depends on the number of Filter inputs. Filter input dependency on the circuit simplicity is discussed in details below.

a) Five-input Median Filter:

To design a 5-input Median Filter the 3-bit Comparator Circuit of the Reconfigurable Rank Order Filter can be replaced by a 2-input OR and a 2-input AND gate. The output of a 5-input Median Filter will be '1' if number of 1's out of the BUCs is greater than or equal to three or the output of the Parallel Counter Circuit is 011, 100 or 101. These three states can be decoded by the circuit as shown in figure 9.

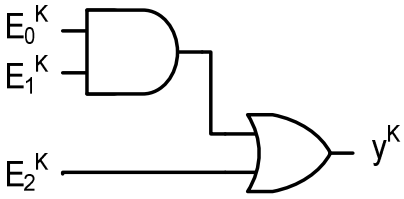


Figure 9: The Comparator Circuit of 5-input reconfigurable ROF can be replaced by this circuit in 5-input median filter.

In figure 9, E_2^K represents the MSB of the output of the Parallel Counter Circuit and E_0^K represents the LSB. The logic function of the combined logic gate circuit of the Median filter is expressed as.

$$Y^k = E_2^K \; OR \; (E_1^K \; AND \; E_0^K) \qquad (5)$$

b) Seven-input Median Filter:

To design a 7-input Median Filter, the Comparator Circuit of the Reconfigurable Rank Order Filter can be eliminated. The output of the 7-input Median Filter will be '1' if the number of the number of 1's out of the BUC is greater than or equal to '4' or the output of the Parallel Counter Circuit is 100, 101, 110 and 111. In all these cases the MSB of the output of the Parallel Counter Circuit is '1'. So the MSB of the output of the Parallel Counter Circuit can be treated as the output of the Median Filter. In 7-input Median Filter the LSB and the Middle bit of the Parallel Counter Circuit need not to be determined. So two Full Adder of the Parallel Counter Circuit can be replaced by two Carry Out Generator circuit. In figure 10, a 7-bit Parallel Counter Circuit is shown. The shaded Full Adders can be replaced by two simple Carry Out Generator circuit. To implement a conventional Full Adder it needs 28 transistors while to make a Carry Out Generator circuit using NAND gates number of transistor needed is only 18. The Carry Out Generator circuit is shown in figure 11.

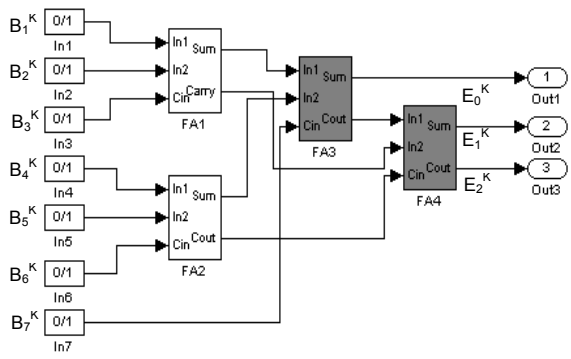


Figure 10: The Full Adders represented by ash colored cells can be replaced by Carry Generator in a 7-input Median Filter.

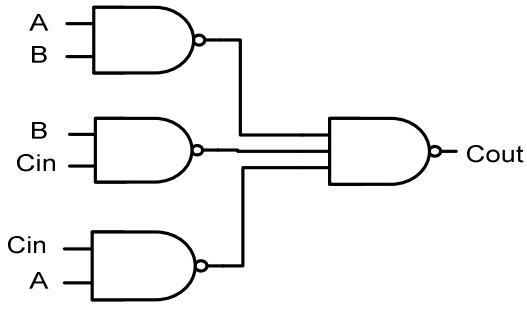


Figure 11. Carry Out Generator Circuit.

The output of the VHDL model of the 7-input Median filter with target device Spartan2E-xc2S50e is shown in figure 12.

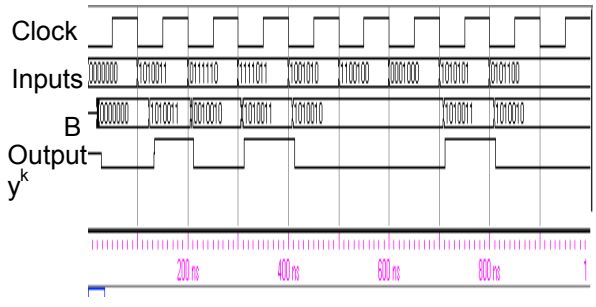


Figure 12. Post-fit simulation output of a 7-input Median Filter.

Here in figure 12, 'B' indicates the output of the BUCs.

c) Nine-input Median Filter:

To design a 9-input Median Filter the 4-bit Comparator Circuit of the Reconfigurable Rank Order Filter can be replaced by two 2-input OR and a 2-input AND gate. The output of a 9-input Median Filter will be '1' if the number of 1's out of the BUCs is greater than or equal to five or the output of the Parallel Counter Circuit is 0101, 0110, 0111, 1000 or 1001. These five states can be decoded by the circuit as shown in figure 13.

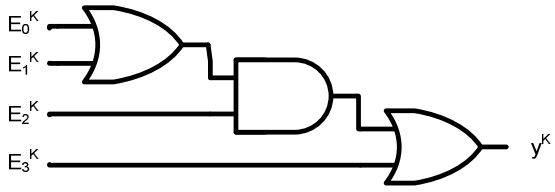


Figure 13: The Comparator Circuit of 9-input reconfigurable ROF can be replaced by this circuit in 9-input median filter.

In figure 13, E_3^K represents the MSB of the output of the Parallel Counter Circuit and E_0^K represents the LSB. The logic function of the combined logic gate circuit of the Median filter is expressed as.

$$Y^k = E_3^K \, OR \, (E_2^K \, AND(E_1^K \, OR \, E_0^K)) \qquad (6)$$

d) Fifteen-input Median Filter:

To design a 15-input Median Filter, the Comparator Circuit of the Reconfigurable Rank Order Filter can be eliminated. The output of the 15-input Median Filter will be '1' if the number of 1's out of the BUC is greater than or equal to 8 or the output of the Parallel Counter Circuit is from 1000 to 1111. In all these cases the MSB of the output of the Parallel Counter Circuit is '1'. So the MSB of the output of the Parallel Counter Circuit can be treated as the output of the Median Filter as in case of 7-input Median Filter. In 15-input Median Filter except the MSB the other three bits of the Parallel Counter Circuit need not to be determined. So three Full Adder of the Parallel Counter Circuit can be replaced by three Carry Out Generator circuit as shown in figure 11.

VI. HSPICE SIMULATION USING 0.18 µM CMOS PROCESS

The proposed Rank Order Filters were simulated in HSPICE using 0.18um CMOS process for a window size $N = 5$. The circuits were found to operate at a clock speed of 500 MHz. To implement the Maximum Filter number of transistor required is only 204. To implement the Minimum Filter also 204 transistors are required. To implement the Median Filter number of transistor required is 300. The transistor level schematic diagram of the BUC is shown in figure 14. The number of transistor required to implement the BUC is 38. The number of transistor required to implement the Full Adder is 28 and number of transistor required to implement the Half Adder is 18. The simulation output for Maximum, Minimum and Median Filters are shown in figure 15.

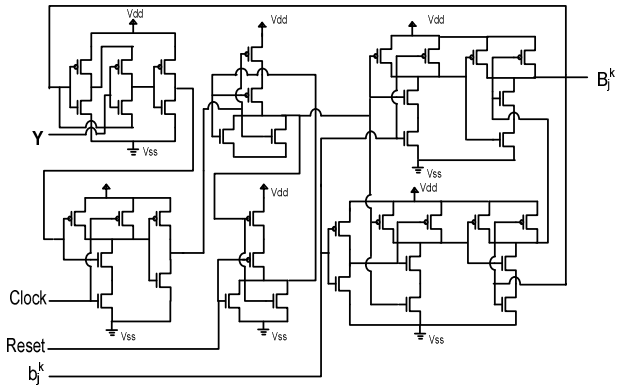


Figure 14. Schematic diagram of the Bit Update Circuit

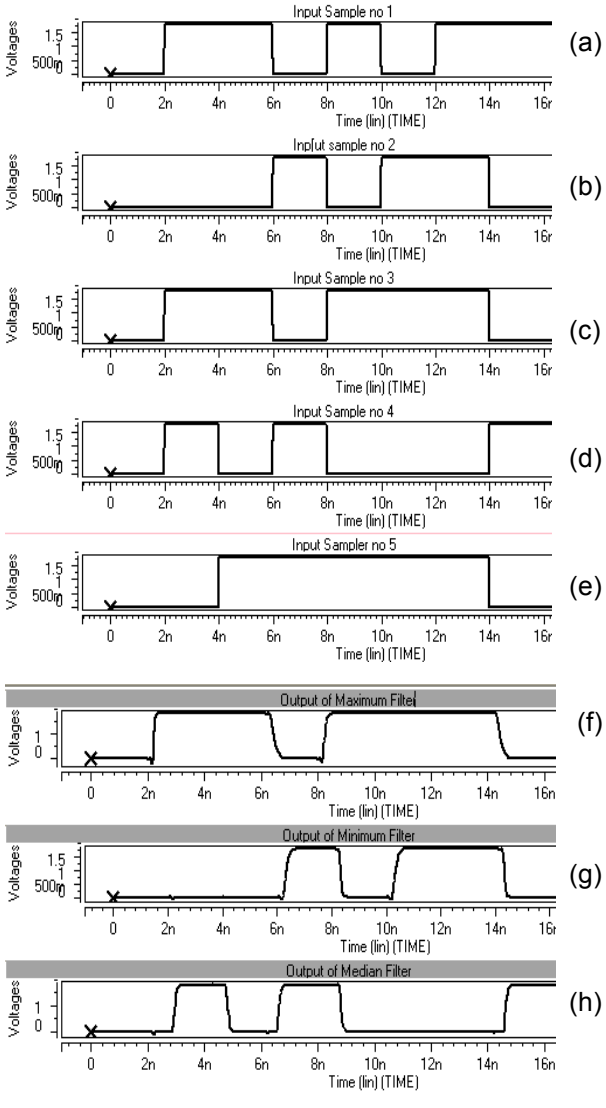


Figure 15: HSPICE simulation results for window size $N = 5$.

In figure 15, (a)-(e) represents the inputs to the 5 input ROFs and (f), (g) and (h) are the output of the Maximum, Minimum and Median filter respectively.

The proposed Maximum, Minimum and Median Filters are also implemented in transistor level for a window size of $N = 7$. The simulation output for Maximum, Minimum and Median Filter are shown in figure 16.

In figure 16, (a)-(g) represents the inputs to the 7 input ROF and (h), (i) and (j) are the output of the Maximum, Minimum and Median filter respectively.

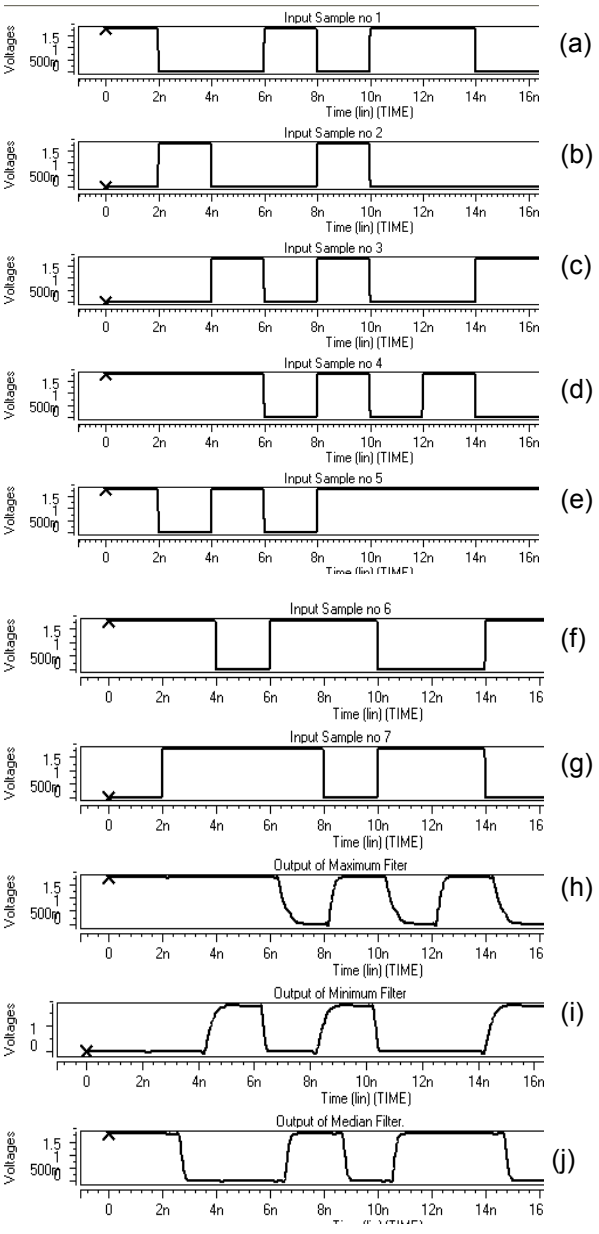


Figure 16: HSPICE simulation results for window size $N = 7$.

The simulation output for both the 5 input and 7 input ROF shows that the circuit can operate correctly at a very high clock rate of 500 MHz.

VII. CONCLUDING REMARKS

The architecture of the reconfigurable Rank Order Filter [6] has been modified to find the element of a certain rank in a given sequence of N element in each window in this work. In Minimum and Maximum Filter the Parallel Counter and Comparator Circuit of the reconfigurable ROF has been replaced by an AND and an OR gate respectively. In case of Median Filter Circuit if the number of input is 2^n-1, (where n is a positive integer) then the comparator circuit of the reconfigurable ROF can be totally eliminated and only the MSB of the Parallel Counter Circuit is to be determined (as shown for 7 and 15 input Median Filter). Thus the proposed circuit requires less number of transistors than that of required by a reconfigurable ROF. It results in low power, less area and high speed ROF. Rank selection by the proposed circuit from HSPICE simulation shows a very good agreement with that obtained by FPGA.

REFERENCES

[1] L. E. Lucke, K. K. Parhi, "A New VLSI Architecture for Rank Order and Stack Filters", in proceedings of the IEEE International Symposium on Circuits and Systems, 1992

[2] C. L. Lee, C. W. Jen, "Bit-Sliced median filter design based on majority gate", IEE Proceedings Vol. 1391, pp 63-71, 1992

[3] L. E. Lucke, K. K. Parhi, "Parallel Structures for Rank Order and and Stack Filters", in proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing 1992.

[4] R. Roncella, R. Salemi and P. Terreni, "70-MHz 2um CMOS Bit-Level Systolic Array Median Filter", IEEE Journal on of Solid State Circuits, vol. 28, no. 5, 1993, pp. 530-536.

[5] B. K. Kar, D.K. Pradhan, "A new algorithm for order statistics and sorting," IEEE Transaction on Signal Processing Vol. 41, No. 8, pp. 2688-2694, August 1993.

[6] George J. Toscano, Pran K. Saha, A.H.M. Zahirul Alam, "A New VLSI Architecture for a Reconfigurable, High-Speed, Digital Rank Order Filter", Proceedings of the International Conference on Computer and Communication Engineering, pp 764-767, 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Scalable Arithmetic Cells for Iterative Logic Array

Bo-Yuan YE, Po-Yu YEH, Sy-Yen KUO, and Shyue-Kung LU

Graduate Institute of Electronics Engineering
Department of Electrical Engineering, BL-522
National Taiwan University
Taipei, Taiwan 106
E-mail: sykuo@cc.ee.ntu.edu.tw

Abstract –In this paper, we propose a novel test technique to achieve both acceptable number of test patterns (NTP) and hardware overhead (HO) by finding a balance between them. Novel bijective and scalable cells are proposed to apply the technique on ILA-based (Iterative-Logic-Array) architectures. A scalable cell consists of *n* bit-level cells and has both hardware and bijective scalability. These simple scalable cells establish the relationship between HO and NTP. Both HO and NTP change as *n* varies. By adjusting the value of *n*, we can obtain an optimal balance between HO and NTP. The ILA based on these scalable cells will be still *C-testable*. We have *C-testable* design for MAC (Multiplier-Accumulator) and the (HO, NTP) for *n = 2* are only (4.43%, 74). With bijective cells, all different ILA meshes can be connected into a bigger hybrid mesh for saving lots of test pins and BIST (Build-In Self Test) area. In addition, the proposed scalable cells induce a simple and systematic way to have balanced results. The proposed technique makes ILA-based DFT schemes more practical, systematic and useful for real complicated applications.

I. Introduction

ATPG (Automatic Test Pattern Generation) for scan-chain DFT (Design-For-Test) is very popular and useful for general applications in VLSI industry. Besides extra scan multiplexers, the number of test-patterns and test-pins may grow as chip gate-count increases. It is well-known that the complexity of general logic testing problem is NP-complete [1]. To test a circuit with repeated cells such as a multiplier, the well-known Iterative-Logic-Array (ILA) DFT scheme should be a better solution. In the past, an ILA is *C-testable* with a constant number of test patterns regardless of the number of the cells [3-15]. The most important condition for a *C-testable* ILA is that the single cell function is bijective, where the bijective property means that the input/output (I/O) function of the cell is one-to-one mapped. However, the traditional stuck-at and the single cell fault models (SCFM) may not be sufficient to achieve the required test level in VLSI industry.

Thus more comprehensive sequential fault models are established [16-19]. In [19], a novel Realistic Sequential Cell Fault Model (RS-CFM) is proposed, and it is a more comprehensive, cell-level model suitable for ILA testing. In [14], it has been shown that a constant number of test vectors are sufficient for fully testing a *k*-dimensional ILA for sequential faults if the cell function is bijective. Furthermore, this concept is extended under RS-CFM fault model [21]. Based on these researches, we can only focus on how to meet the conditions of *C-testable* ILA under SCFM. Then this ILA will still be *C-testable* for the sequential faults under RS-CFM.

There are several approaches for the testability of the adder and multiplier without using bijective property. In [13], reorganized *C-testable* test pattern sequences are proposed for testing a single *n*-bit adder at bit-level. In [22-27], *C-testable* test patterns are generated for a single multiplier. In general, most of applications have hybrid-cascade situation such as *N*-tap FIR (Finite Impulse Response) filter cascading lots of adders and multipliers alternately. Thus there will be a problem that the input and output pins of each arithmetic logics should be controllable and observable respectively. If all arithmetic logics are modified to be bijective with same I/O pins, the original circuit can be taken as a *C-testable* ILA with cascaded bijective cells. It will reduce lots of test pins and BIST area. In order to meet the requirements of *C-testable* ILA under SCFM, each module (cell) should be modified to be bijective. Conventional ILA test schemes at module-level or bit-level usually leads to large number of test patterns (NTP) or significant hardware overhead (HO), respectively. All we only can do is to pick up the best test scheme at bit-level or module-level after the modification. However, the HO or NTP may not be small enough at either bit-level or module-level.

In order to avoid extreme HO/NTP, novel scalable cells with bijective property are proposed here to trade off between HO and NTP. A scalable cell is composed of *n* bit-level cells, where the value of *n* is decided by balancing between HO and NTP. These simple scalable cells establish the relationship between HO and NTP. With the scalable cells, The *C-testable* multiplier and MAC are proposed with scalable cells, and the (HO, NTP) pairs for them are only (4.16%, 74) and (4.43%, 74) respectively. By adjusting the value of *n*, we can obtain an optimal balance between HO and NTP. Thus both acceptable HO and NTP can be achieved.

The organization of this paper is described as follows. Section II reviews the bijective characteristics of ILA architectures. Section III shows how to derive the proposed scalable arithmetic cells. It also demonstrates what advantages the proposed technique contributes. Performance analysis is shown in section IV. Finally, conclusions are given in Section V.

II. Review of ILA Architectures

Based on SCFM, we assume that a module's behaviour is invariant over time, even if it is faulty. A faulty module's function may deviate from the correct one in any manner, as long as it remains combinational, i.e. we are testing for permanent combinational faults only. As long as each module in the ILA has the same number of I/O pins and conforms to bijective property, the NTP of an ILA will be a constant same as testing a single module. N bijective cells with n-bit I/O pins are cascaded one after another to form a 1-D ILA as shown in Fig. 1. For a given input pattern I_0 to the 1^{st} cell, it will be propagated cell by cell as I_1, I_2 ...I_{N-1} and I_N, i.e. $Cell_2$, $Cell_3$, …$Cell_{N-1}$ and $Cell_N$ get input patterns I_1, I_2 ...I_{N-2} and I_{N-1}, respectively. The bijective property makes all cells be fully controlled by propagated patterns. As 2^n exhaustive test patterns (ETPs) are fed into the 1^{st} cell, every cell will get and generate 2^n ETPs due to the bijective property. If there is a faulty cell in an ILA, the faulty pattern will be propagated cell by cell to the last (boundary) cell (i.e. $Cell_N$). Therefore, the bijective property guarantees not only the testability of each cell but also the *C-testability* of ILA.

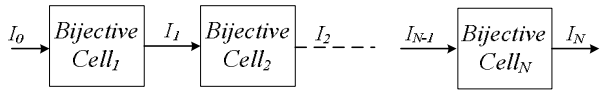


Fig. 1 A one-dimensional iterative logic array.

III. Scalable Cell Architectures

In this section, the goal is to figure out how to achieve the balance between HO and NTP for a testable ILA design. The key to the solution is to obtain the relationship between HO and NTP. Thus, scalable cells are proposed to establish the relationship. A scalable cell is composed of n bit-level cells. It can be said that a scalable cell is a cell with bijective property in between bit-level and module-level. The balance between HO and NTP for ILA is achieved by adjusting the parameter n in a scalable cell.

A. Scalable Word-Level Adder/ Subtractor

In Fig. 2, the word-level adder/subtractor (FA_w/FS_w) cells perform simple w-bit addition and subtraction, respectively. The FA_w/FS_w cells can be modified to become bijective cells tFA_w/tFS_w by simply bypassing the input bus Bi to the output bus Bo. Obviously, for each fixed value of Bi, the mapping from Ai to Ao is a permutation. Thus the function $\{Ao, Bo\}=\{(Ai \pm Bi)_{mod}\ 2^w, Bi\}$ is still bijective, where the comma in {…} expression is to combine several bits or buses into an integrated bus and the expression $(X)_{mod}\ 2^w$ will return the remainder when the bus X divides 2^w. Take a 1-bit $tFA_w|_{w=1}$ as an example, if the value pairs *0/0*, *0/1*, *1/0* and *1/1* are input to *Ai/Bi* buses, the output buses *Ao/Bo* will output *0/0*, *1/1*, *1/0* and *0/1* respectively.

The NTP and HO for the tFA_w (or tFS_w) cell are 2^{2w} and 0% respectively. Even though the NTP is pretty large and no balance can be done between NTP and HO (due to that HO always stays 0%), this bijective property is still very important because it is the fundamental of all scalable cells in this paper.

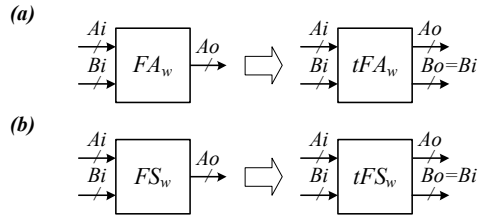


Fig. 2 (a) FA_w/ scalable tFA_w cells. (b) FS_w/ scalable tFS_w cells.

B. Scalable n-bit FA/FS

In Fig. 3, the bit-level FA_n/FS_n performs a slice of the operation in FA_w/FS_w, respectively. First, the input bus Bi directly is bypassed to the output bus Bo. Then, the input pin ci is switched to the output pin co with an extra 2-to-1 multiplexer. Thus, the FA_n/FS_n cells are modified to become bijective cells tFA_n/tFS_n in test mode. Based on tFA_w, $\{Ao,Bo\}=\{(Ai+Bi+ci)_{mod}\ 2^n, Bi\}$ is also bijective no matter what value ci is. Thus, the tFA_n with $\{Ao,Bo,co\}=\{(Ai+Bi+ci)_{mod}\ 2^n,Bi,ci\}$ is still bijective in test mode (tm=1). In the same manner, tFS_n is also bijective. Take a 1-bit $tFA_n|_{n=1}$ as an example, if the test patterns *0/0/0*, *0/0/1*, *0/1/0*, *0/1/1*, *1/0/0*, *1/0/1*, *1/1/0* and *1/1/1* are input to *Ai/Bi/ci* buses in test mode (tm=*1*), the output buses *Ao/Bo/co* will output *0/0/0*, *1/0/1*, *1/1/0*, *0/1/1*, *1/0/0*, *0/0/1*, *0/1/0*, and *1/1/1* respectively.

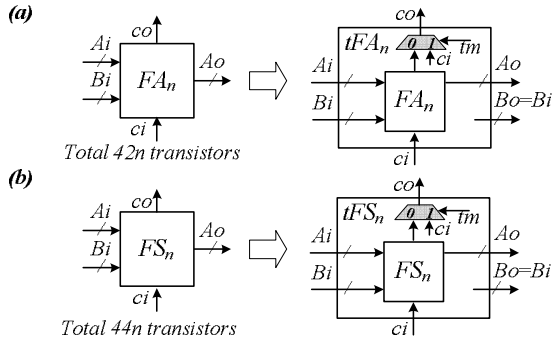


Fig. 3 (a) FA_n/ scalable tFA_n cells. (b) FS_n/ scalable tFS_n cells.

There are 8 extra test patterns for testing the untested paths (i.e. the co path of FA_n/FS_n) in normal mode (tm=0). When n is large, the number of extra test patterns is only a small percentage of the total NTP. Notice that the NTP and HO of the tFA_n cell are only $2^{2n+1}+8$ and $T(MUX)/T(FA_n)=4/(42n)$ respectively, where $T(X)$ stands for transistor count for X. When n increases, the NTP becomes larger and the HO is reduced. When a FA_w is replaced by w/n tFA_n cells, there are much smaller NTP compared with that for a single tFA_w. In this way, FA_w can be viewed as an ILA using tFA_n cells and is *C-testable* (i.e. NTP is independent of w). Through circuit implementation, the delay overhead (DO) of the extra multiplexer can be hidden in the non-critical path of tFA_n/tFS_n.

C. Scalable n-bit MUL

Fig. 4(a) shows the circuits of the MUL_n cell, which is the basic cell in a multiplier. Note that the AND gates perform $(Bi,\ di)_{and}$, which stands for bit-wise AND operation on each bit of bus Bi with the bit di. In Fig.

4(b), the FA_n cell in the MUL_n cell is replaced by tFA_n cell. *Bi* and *di* are also bypassed to *Bo* and *do*, respectively. Based on tFA_n, thus the function $\{Ao,Bo,co\}=\{(Ai+(Bi, di)_{and})_{mod}\ 2^n, Bi, ci\}$ is still bijective no matter what value *di* is. The $tMUL_n$ cell becomes a bijective cell in test mode (*tm*=1). Take 1-bit $tMUL_n|_{n=1}$ as an example. When *di* is equal to 1 in test mode (*tm*=1), $tMUL_n|_{n=1}$ acts as the bijective mapping of the $tFA_n|_{n=1}$ cell. When *di* is equal to 0, $tMUL_n|_{n=1}$ acts as the bijective mapping of the $tFA_n|_{n=1}$ cell with $Bi = 0$. Notice that the NTP and HO of the $tMUL_n$ cell are only 2^{2n+2} and *4/(48n)* respectively. In addition, it also needs 8 extra test patterns in normal mode to test the untested paths in tFA_n.

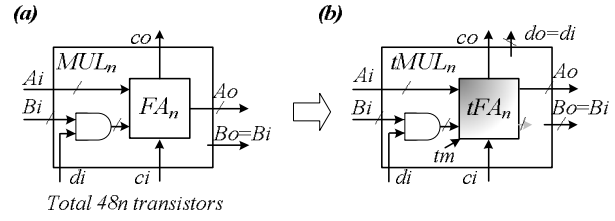


Fig. 4 (a) MUL_n cell. (b) Scalable $tMUL_n$ cell.

D. Scalable $m\times m$ -bit unsigned Multiplier

A 2's complement multiplier can be implemented by unsigned one with slight modification. In order to simplify our discussion, an $m\times m$ unsigned multiplier is taken as the target design. At first, the symbol of MUL_m is redrawn from Fig. 4(a) to Fig. 5(a) in order to illustrate the stair-like structure (1-bit shift) in a multiplier, where Ai_0^{m-2} and Ao_1^{m-1} stand for $\{ai_0,..., ai_{m-2}\}$ and $\{ao_1,..., ao_{m-1}\}$, respectively. *m* MUL_m cells can be cascaded one by one to implement an unsigned multiplier $X_{m\times m}$ as shown in Fig. 5(b). When the inputs *Ai* and *Ci* are 0, the multiplier acts as $Ao=\{Ao_H, Ao_L\}=Bi\times Di$. Notice that *Ao* and *Bi* (or *Di*) have *2m* and *m* bits, respectively.

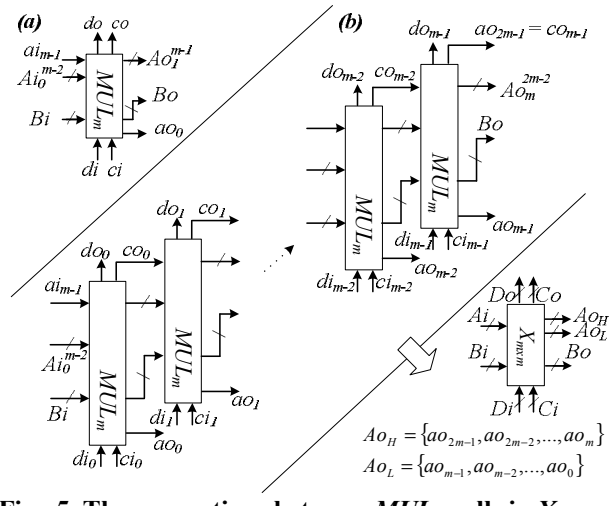


Fig. 5 The connections between MUL_m cells in $X_{m\times m}$.

In order to achieve a testable multiplier $tX^n_{m\times m}$ as illustrated in Fig. 6, each MUL_m cell is separated into a MUL_n cell and *p-1* $tMUL_n$ cells, where *p=m/n*. Because the *co* and *do* of MUL_m are used for observation, the top cell (MUL_n) is not replaced by $tMUL_n$. Second, extra multiplexers is added between MUL_n cells in $tX^n_{m\times m}$. Since the extra multiplexers are located in the non-critical paths, there is no delay overhead for $tX^n_{m\times m}$. In normal mode (*tm=0, Ai=0*, and *Ci=0*), all $tMUL_n$ cells act as MUL_n cells and therefore $tX^n_{m\times m}$ acts as $X_{m\times m}$. In test mode (*tm=1*), the bijective property of the $tMUL_n$ cell makes $tX^n_{m\times m}$ become *C-testable*. Next, let us investigate in detail how $tX^n_{m\times m}$ becomes *C-testable* with the help of the extra multiplexers and $tMUL_n$ cells.

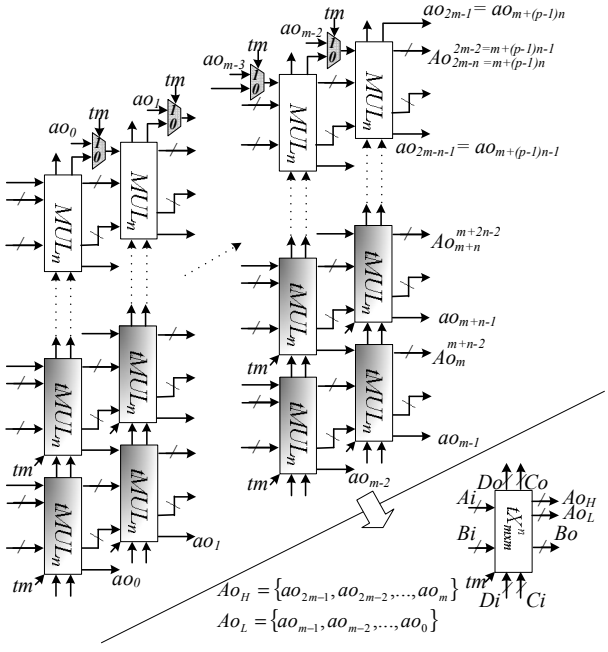


Fig. 6 *C-testable* multiplier $tX^n_{m\times m}$ using $tMUL_n$ cells.

The one part of Fig. 6 is enlarged in Fig. 7. In Fig. 7, each column is taken as a bijective stage in test mode (*tm=1*). $\{Ai, Bi, (ci,di)\}^i$ and $\{Ao, Bo, (co,do)\}^i$ stand for inputs and outputs for the MUL_n or $tMUL_n$ cell in the i^{th} stage, respectively. In the 1^{st} stage, we let $\{Ai, Bi\}^{i=1}$ of each cell (MUL_n or $tMUL_n$ cell) get the same input pattern *{I1,J1}*. Due to *{ci, di}* are bypassed to *{co,do}* in each $tMUL_n$ cell, $\{(ci,di)\}^{i=1}$ of each cell in the 1^{st} stage will gets the same pattern *K1*. Except the 1^{st} stage, the input patterns for other stages in $tX^n_{m\times m}$ are propagated from the previous stage.

Let's examine carefully the connections between the 1^{st} and the 2^{nd} stages as shown in Fig. 7. Due to the stair-like structure, the *Ai* of the $tMUL_n$ cell in the 2^{nd} stage is connected with the *Ao* of the two $tMUL_n$ cells in the 1^{st} stage. Since all 1^{st}-stage cells get the same input patterns, $tMUL_n$ cells in the 2^{nd} stage get the input pattern as $ROR(\{Ao\}^{i=1})= ROR(I2)$, where *ROR(X)* stands for 1-bit right-rotation operation on *X*. Due to the extra multiplexer between the 1^{st} and 2^{nd} stages, the top cell (MUL_n) in the 2^{nd} stage also get the pattern as $ROR(\{Ao\}^{i=1})= ROR(I2)$. Therefore, all the cells (one MUL_n and *p-1* $tMUL_n$ cells) in the 2^{nd} stage also get the same input pattern *{I2,J2,K2}*. This shows that all the cells in the i^{th} stage will get the same input pattern *{Ii,Ji,Ki}*.

In other words, all $tMUL_n$ cells in the same stage get the same input and output patterns. Since the $tMUL_n$ cell is bijective, the ETPs for $\{Ai, Bi, (ci,di)\}^{i=1}$ will be

propagated stage by stage. All the cells in $tX^{n}_{m\times m}$ can be controlled by propagating test patterns. This shows the controllability for $tX^{n}_{m\times m}$. If there is a faulty cell in $tX^{n}_{m\times m}$, the incorrect output pattern will be propagated cell by cell. Eventually, the incorrect pattern can be observed from the $\{Ao,Bo\}^{m}$ of the cells in the last stage, the *{(co,do)}* of all top cells (MUL_n) and the output ao_{m-1} pin. Except the output ao_{m-1} pin, each pins of Ao_L is already tested as the input of the top cell in each stage. This shows the observability for $tX^{n}_{m\times m}$ multiplier. Therefore, $tX^{n}_{m\times m}$ is an ILA with *C-testability* and can be used as a bijective cell in another ILA.

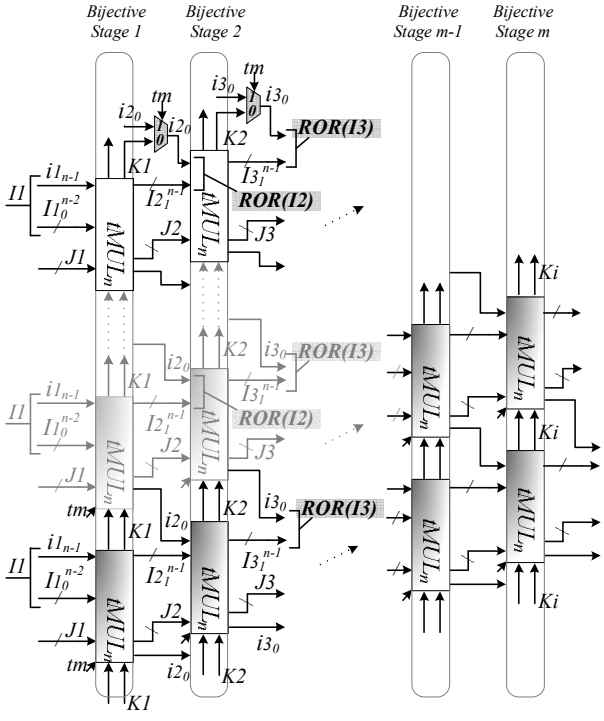


Fig. 7 Detail connections between neighbor stages in $tX^{n}_{m\times m}$.

Notice that the NTP of $tX^{n}_{m\times m}$ is equal to 2^{2n+2}. It still needs extra 10 test patterns, which can be found by ATPG tools, to test the untested paths in all tFA_n cells and *m-1* extra multiplexers between MUL_n cells in normal mode. There are $m\times(p\text{-}1)$ $tMUL_n$ in $tX^{n}_{m\times m}$ and then we have $OT(tX^{n}_{m\times m}) = (m\text{-}1)\times 4 + m\times(p\text{-}1)\times 4$, where *OT(X)* stands for the overhead transistor count of the *X* cell and *T(MUX)=4*. The NTP and HO are summarized as equation (1) below.

$$NTP\left(tX^{n}_{m\times m}\right) = 2^{2n+2}+10 \quad where \quad p = m/n$$
$$HO\left(tX^{n}_{m\times m}\right) = \frac{(m-1)\times 4 + m\times(p-1)\times 4}{m\times m\times 48} \qquad (1)$$

In fact, the *ci* of the bottom cells in all stages can be connected together and this will release the routing resource. With the $tMUL_n$ cells, $tX^{n}_{m\times m}$ is also scalable and can balance between NTP and HO like other scalable cells do.

E. Scalable Multiplier-Accumulator

Now, we make an adder (FA_m) cascade with a multiplier ($X_{m\times m}$) to form a MAC (Multiplier-Accumulator) cell ($MAC_{m\times m}$) in Fig. 8(a). In order to simplify the explanation, all operations are limited in *m* bits. Therefore, Ao_H is taken as the result from the $X_{m\times m}$ cell. The $MAC_{m\times m}$ cell acts as *Ao=Ei+[(Bi×Di)>>m]* with *Ai=0* and *Ci=0*. *Di* and *Do* both are formed from $X_{m\times m}$. *Ci* and *Co* are formed from the vertical inputs and outputs of two cells (FA_m and $X_{m\times m}$), respectively. Moreover, we follow the structure of $tX^{n}_{m\times m}$ to form the $tMAC^{n}_{m\times m}$ cell easily. Therefore, FA_m in $MAC_{m\times m}$ is separated into a FA_n cell and *p-1* tFA_n cells. Of course, an extra multiplexer is also added for the top FA_n cell. The result forms the tFA^{n}_{m} cell as shown in the right side of Fig. 8(c). Fig. 8(c) also shows the detail connections between the *Ao* of $tX^{n}_{m\times m}$ and the *Ai* of tFA^{n}_{m}.

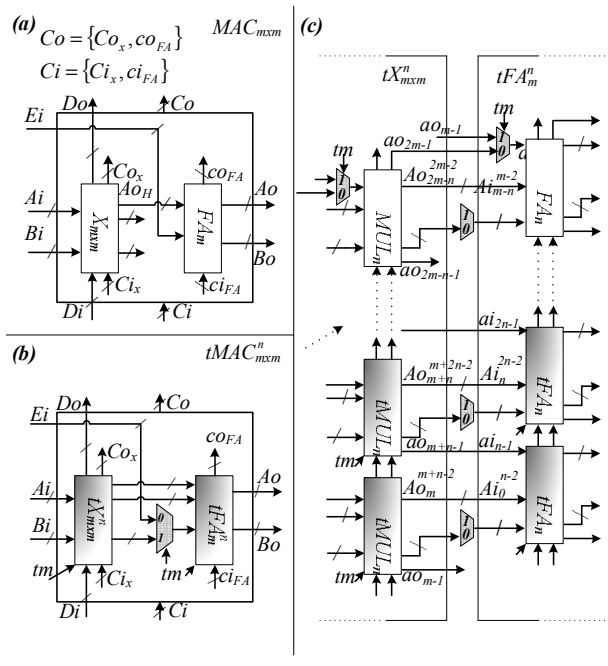


Fig. 8 (a) MAC_{mxm}. (b) *C-testable* $tMAC^{n}_{m\times m}$. (c) The connections between the *Ao* of $tX^{n}_{m\times m}$ and the *Ai* of tFA^{n}_{m}.

In normal mode (*tm=0*, *Ai=0* and *Ci=0*), the *Ai* of tFA^{n}_{m} is only connected to the Ao_H of $tX^{n}_{m\times m}$. In test mode (*tm=1*), the *Ai* of tFA^{n}_{m} is connected to a sub-bus of *Ao* in $tX^{n}_{m\times m}$, i.e. $\{Ai_0^{m-1}\}$ input pins for tFA^{n}_{m} are connected to $\{ao_{m-1}$ (the MSB of Ao_L), $Ao_m^{2m-2}\}$ output pins from $tX^{n}_{m\times m}$. Besides, *m* 2-to-1 multiplexers are added between the *Bo* of $tX^{n}_{m\times m}$ and the *Bi* of tFA^{n}_{m} as shown in Fig. 8(b). Then, we can switch between the *Bo* of $tX^{n}_{m\times m}$ and *Ei* by the test-mode-enable signal *tm*. It is easy to see that the $tMAC^{n}_{m\times m}$ cell is scalable and *C-testable* because of $tX^{n}_{m\times m}$ and tFA_n cells. The NTP and HO are summarized as equation (2) below.

$$NTP\left(tMAC^{\ n}_{m\times m}\right)=2^{2n+2}+10 \ \ where \ \ p=m/n$$
$$HO\left(tMAC^{\ n}_{m\times m}\right)=\frac{m\times 8+(m+1)\times(p-1)\times 4}{m\times m\times 48+m\times 42} \qquad (2)$$

F. Summary on Scalable Cells

Based on the discussions above, scalable arithmetic cells with bijective property for any *n* are proposed. As the parameter *n* of a scalable cell increases, NTP and HO will become larger and lower, respectively. Fig. 9 shows a curve depicting the typical relationship between HO and NTP for a scalable cell. tFA_w and tFS_w are special cases with *HO=0*.

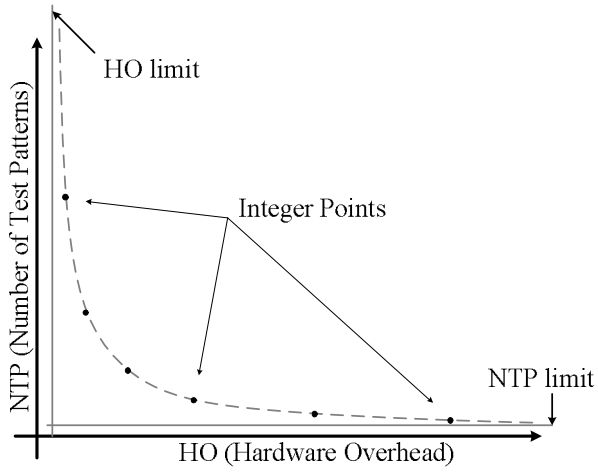


Fig. 9. The HO-NTP curve of a scalable cell.

The proposed cells demonstrate that scalability indeed is of great help. Besides, scalability also induces a systematic way for ILA DFT design summarized as follows:

a). Find scalable cells for the target circuit and get the equations for HO and NTP.
b). Plot the curve of the relationship between HO and NTP.
c). Finally, to balance between HO and NTP along the relationship trajectory based on the requirement of the application.

IV. Performance Analysis

Under SCFM, novel scalable and testable cells have been proposed and discussed in detail. The NTP and HO equations for the proposed scalable cells are list in Table 1.

Table 1 The NTP/HO list for the proposed scalable cells.

w,m: word-length *n*: *n*-bit group	NTP	HO
tFA_w/ tFS_w	2^{2w}	0%
tFA_n	$2^{2n+1}+8$	*4/(42n)*
tFS_n	$2^{2n+1}+8$	*4/(44n)*
$tMUL_n$	$2^{2n+2}+8$	*4/(48n)*
$tX^n_{m\times m}$	$2^{2n+2}+10$	Equation (1)
$tMAC^n_{m\times m+m}$	$2^{2n+2}+10$	Equation (2)

$tX^n_{m\times m}|_{n=m}$ and $tX^n_{m\times m}|_{n=1}$ (i.e. $tX^n_{m\times m}$ with *n=1*) can be taken as the DFT at module-level and at bit-level, respectively. (HO, NTP) for $tX^n_{m\times m}|_{n=m}$ and $tX^n_{m\times m}|_{n=1}$ are (0.25%, 2^{66}+10) and (8.33%, 26) respectively, where *m* is equal to 32. Without scalability, we only can pick one of them as result. It is easy to see that both the NTP of $tX^n_{m\times m}|_{n=m}$ and the HO of $tX^n_{m\times m}|_{n=1}$ are not small enough to be acceptable. However, the scalable $tX^n_{m\times m}$ cell can solve such a situation with *n = 2*. The $tX^n_{m\times m}|_{n=2}$ module has more acceptable NTP and HO ((HO, NTP) = (4.16%, 74)) compared with $tX^n_{m\times m}|_{n=m}$ and $tX^n_{m\times m}|_{n=1}$. This is the advantage of the scalable cells. The HO and NTP might not the best one due to the trade-off, but they are both more acceptable.

There are only four (HO, NTP) pairs for $tMAC^n_{m\times m}$ from *n = 1* to *4* denoted by black circles in Fig. 10. We can investigate the plot to decide which value for *n* is acceptable. According to the HO-NTP plot, it is easy to select acceptable solutions among the pairs, e.g. we take *n = 2* for $tMAC^n_{m\times m}$ and (HO, NTP) is (4.43%, 74) as shown in Fig. 10.

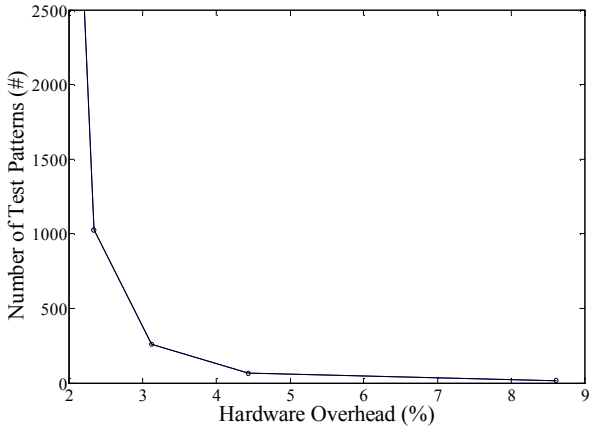


Fig. 10 The HO-NTP curve for $tMAC^n_{m\times m}|_{m=32}$.

Due to the scalable cells establish the relationship of HO and NTP, acceptable results can be achieved by balancing between HO and NTP. There are more choices than the two solutions, i.e. test schemes at module-level and bit-level. The more acceptable solution might exist.

Due to the scalable arithmetic cells with bijective property, we can cascade with homogenous or non-homogenous scalable cells to form a hybrid ILA. Thus, we can cascade many $tMUL_n$ (many homogenous cells) cell by cell to form a *C-testable* multiplier ($tX^n_{m\times m}$). We also can cascade tFA_n and $tX^n_{m\times m}$ (two non-homogenous cells) to form a *C-testable* MAC ($tMAC^n_{m\times m}$). After proposing $tX^n_{m\times m}$ and $tMAC^n_{m\times m}$ cells, it is much easier to get a *C-testable N*-tap-FIR filter or matrix multiplication, which cascade lots of adders and multipliers alternately. Scalable cells are reusable for similar structure design. By testing many different cells in a mesh, we can save routing resources, test pins and BISTs area.

V. Conclusions and Future works

In this paper, we proposed a novel test idea to achieve both acceptable HO and NTP by trading off between them. This novel idea makes ILA-based DFT techniques more practical, systematic and useful for real complicated applications. In summary, we have three main contributions. First, novel scalable and bijective cells were proposed to establish the relationship between HO and NTP. The scalable cells provide a simple and systematic way to obtain both acceptable HO and NTP at the same time. It is much attractive that a scalable cell has not only hardware scalability but also bijective scalability.

Second, we demonstrate that the method is feasible and useful by applying it to well-known arithmetic logics. It is shown that the scalability and *C-testability* can be inherited from the composing scalable cells. The new scalable cell could be easily and quickly formed from the existing ones due to similar architectures. It also shows that different cells can be joined or cascaded together due to bijective and scalable properties. Moreover, it is easy to design the BIST for the scalable cells.

Last but not least, all results in our work show that the proposed scalable architectures are useful. It indicates that it is very important to find versatile scalable cells for any application. In addition to the feasibility of the novel idea, it is shown that the proposed scalable test scheme is superior to the conventional ILA method without scalability. For the future work, we will try to explore more applications such as DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), ME (Motion Estimation) and even more complicated designs, to derive more novel cells with scalability and show the idea can be comprehensively applied.

References

[1]. H. Fujiwara and S. Toida, "The complexity of fault detection problems for combinational logic circuits," *IEEE Trans. Computers*, Vol. C-31, No. 6, pp. 555-560, June 1982.

[2]. W. H. Kautz, "Testing for faults in combinational cellular logic arrays," *Proc. 8th Annu. Symp. Switching, Automata Theory*, 1967, pp. 161-174.

[3]. P. R. Menon and A. D. Friedman, "Fault detection in iterative arrays," *IEEE Trans. Computers*, Vol. C-20, pp. 524-535, May 1971.

[4]. A. D. Friedman, "Easily testable iterative systems," *IEEE Trans. Computers*, Vol. C-22, pp. 1061-1064, Dec. 1973.

[5]. T. Sridhar and J. P. Hayes, "Design of easily testable bit-sliced systems," IEEE Trans. Computers, Vol.C-30, No. 11, pp. 842-854, November 1981.

[6]. R. Parthasarathy and S. M. Reddy, "A testable design of iterative logic arrays," IEEE Trans. Computers, Vol.C-30, No. 11, pp. 833-841, November 1981.

[7]. E. M. Aboulhamid and E. Cerny, "Built-in testing of one-dimensional unilateral iterative arrays," IEEE Trans. Computers, Vol. C-33, No. 6, pp. 560-564, June 1984.

[8]. A. Vergis and K. Steiglitz, "Testability conditions for bilateral arrays of combinational cells," IEEE Trans. Computers, Vol. C-35, No. 1, pp. 13-26, January 1986.

[9]. W. T. Cheng and J. H. Patel, "Testing in two-dimensional iterative logic arrays," Proc. Int. Symp. Fault Tolerant Computing, pp. 76-81, 1986.

[10]. H. Elhuni, A. Vergis, and L. Kinney, "C-testability of two-dimensional iterative arrays," IEEE Trans. Computers-Aided Des. Circuits Syst., Vol. CAD-30, No. 4, pp. 573-581, October 1986.

[11]. W. K. Huang and F. Lombardi, "On an improved design approach for C-testable orthogonal iterative arrays," IEEE Trans. Computers-Aided Des. Circuits Syst., Vol. 7, No. 5, pp. 609-615, May 1988.

[12]. F. Lombardi, "On a new class of C-testable systolic arrays," Integration, Vol. 8, pp. 269-283, 1989.

[13]. C. W. Wu and P. R. Cappello, "Easily testable iterative logic arrays," *IEEE Trans. Computers*, Vol. C-31, No. 6, pp. 640-652, May 1990.

[14]. C.Y. Su and C.W. Wu, "Testing Iterative Logic Arrays for Sequential Faults with a Constant Number of Patterns," *IEEE Trans. Computers*, Vol. 43, No. 4 pp. 495-501, April 1994.

[15]. S. K. Lu, C. W. Wu and S.-Y. Kuo, "Enhancing testability of VLSI arrays for fast Fourier transform," *IEE Proc. Part E*, Vol. 140, No. 3, pp. 161-166, May. 1993.

[16]. J. Galiay, Y. Crouzet, and M. Vergiault, "Physical versus logical fault models in MOS LSI circuits, impact on their testability," IEEE Trans. Computers, Vol. 29, No. 6, pp. 527-531, June 1980.

[17]. A. K. Pramanick and S. M. Reddy, "On detection of delay faults," Proc. IEEE Int'l Test Conf., pp. 845-856, 1988.

[18]. G. L. Smith, "Model for delay faults based upon path," Proc. IEEE Int. Test Conf., pp. 342-349, 1985.

[19]. M. Psarakis, D. Gizopoulos, A. Paschalis and Y. Zorian, "Sequential Fault Modeling and Test Pattern Generation for CMOS Iterative Logic Arrays," *IEEE Trans. Computers*, Vol. 49, No. 10, pp. 1083-1098, Oct. 2000.

[20]. S.-K. Lu, C.-W. Wu, and R.-Z. Hwang, "Cell Delay Fault Testing for Iterative Logic Arrays, " J. Electronic Testing: Theory and Applications, vol. 9, no. 3, pp. 311-316, Dec. 1996.

[21]. S. K. Lu, "Delay Fault Testing for CMOS Iterative Logic Arrays with a Constant Number of Patterns," IEICE TRANS. INF. &SYST., Vol. E86-D, No.12 December 2003.

[22]. S. M. Aziz, "C-testable modified Booth's array multiplier," *Proc. 8th Int. Conf. VLSI Design*, 4-7, pp. 278 -282, Jan. 1995.

[23]. D. Gizopoulos, A. Paschalis; D. Nikolos and C. Halatsis, "Linear-testable and C-testable Nx×Ny modified Booth multipliers, "*IEE Proc. Computers and Digital Techniques*, Vol. 143, Issue 1, pp. 44-48, Jan. 1996.

[24]. W. A. J. Waller and S. M. Aziz, "C-testable parallel multiplier using differential cascade voltage switch (DCVS) logic," IFIP Trans. A, A-42, pp. 133-142, 1994.

[25]. J. Shen and F. Ferguson, "The Design of Easily Testable VLSI Array Multipliers," *IEEE Trans. Computers*, Vol. c-33, No. 6, pp. 554-560 June 1984.

[26]. D. Gizopoulos, A. Paschalis and Y. Zorian, "Effective Built-In Self-Test for Booth Multipliers," *IEEE Design & Test of Computers*, pp. 105-111, 1998.

[27]. S.K. Lu, J.C. Wang, and C.W. Wu, "C-Testable Design Techniques for Iterative Logic Arrays," *IEEE Trans. VLSI*, Vol. 3, No. 1, pp. 146-152, March 1995.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Measured Impedance for Inter Phase Faults on Next Line and Second Circuit of a Double Circuit Line

H. Shateri and S. Jamali

Centre of Excellence for Power Systems Automation and Operation Department of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak 16846, Tehran, Iran
shateri@iust.ac.ir and sjamali@iust.ac.ir

Abstract - Distance relays are one of the mostly used protective devices in the transmission systems. These relays are used as the main, the first zone and a part of the second zone, and the backup, the rest of the second zone and the third zone, protective devices. Fault resistance is a source of error for distance relays in the form of the measured impedance deviation from its actual value. This paper compares the measured impedance a by distance relay on a double circuit line for the faults on the next line and the second circuit of the line, in the case of inter phase, phase to phase and three phase, faults. This is done by presenting the measured impedance at the relaying point and distance relay ideal tripping characteristic. The variation of the ideal tripping characteristic due to the changes in the power system parameters is investigated.

I. Introduction

Frequency variation, ground fault resistance in the single-phase to ground faults (and inter phase faults), and power swing are some of the phenomena adversely affect distance protection performance. Among these, frequency variation and power swing relate to dynamic states of power systems. The ground fault resistance depends on the state of the fault arc creation between the ground and the faulted phase of the transmission line and the ground path [1]. More than 70% of the occurred faults on the overhead lines are of the single phase to ground type. However, phase to phase faults are the most common of the fault type after the single phase to ground faults. In the case of phase to phase faults, inter phase fault resistance depends on the quality of the fault arc creation between the faulted phases. The other types of the faults (double phase to ground and three phase faults) are less common in the transmission systems. The three phase fault is the least common fault type.

Since more than 70% of the faults in the transmission systems are of the single phase to ground type, many efforts studied the measured impedance at the relaying point for single phase to ground faults [2]-[5]. But, no serious attention has been paid to the other types of the faults. Only an effort [6] studied the measured impedance in the case of the double phase to ground faults.

Ground fault resistance is an unknown phenomenon causing distance relay mal-operation. Reference [2] presents an adaptive distance protection method to overcome this problem. Due to unknown magnitude of the ground fault resistance, this method defines an operational characteristic, which is robust against the variations of the ground fault resistance and does not mal-operate. References [3]-[5] discuss different aspects of adaptive distance protection.

In today's power systems, there are some difficulties for constructing new transmission lines, because of the limits in the rights for their paths. This leads to introduction of double-circuit lines into power systems. In the case of double-circuit lines, because of the low distance between the conductors of two circuits, the two circuits affect each other mutually.

Usually, distance relays, at least, have three protective zones. First or fast zone covers 80% to 95% of the line, dependent on the accuracy of the metering equipments. Second zone covers the rest of the line (as the main protective device) and a part, e.g. 50%, of the lines connected to the far end of the line (as the backup protection). Third zone covers the rest of the adjacent lines and a part of the line connected to them as well as a part of the lines connected to relaying point. Unlike the first zone, second and third zones are delayed zones.

As mentioned, the second and third zones cover a part of the lines connected to the far end of the protected line. In the case of a double circuit line, the second circuit of the line and the other lines connected to the far end of the line are placed in this category.

The fault resistance not only is an unknown quantity from protective system point of view, but this resistance could be measured as an inductive or capacitive impedance, depending on the far end infeed and the line pre-fault loading. Variation of the imposed impedance due to the presence of the ground fault resistance from its resistive feature depends on both the structural and operational conditions of the system.

This paper compares the measured impedance by a distance relay on a double circuit line for the faults on the next line and the faults on the second circuit, for inter phase, phase to phase and three phase, faults. This is done by presenting the measured impedance in the two cases. In addition, the variation of the measured impedance due to changes in power system conditions is investigated.

II. Measured Impedance at Relaying Point

Distance relays operate based on the measured impedance at the relaying point. In the case of zero fault

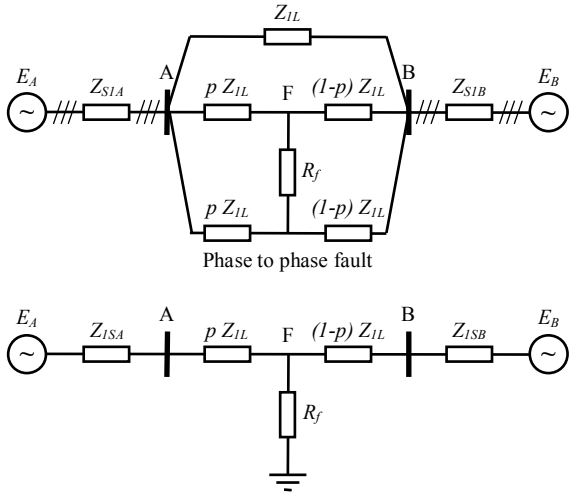


Fig. 1. Inter phase faults equivalent circuit, single circuit line

resistance, the measured impedance at the relaying point only depends on the length of the line section between the fault and the relaying points. According to Fig. 1 this impedance is equal to pZ_{1L} where p is the per-unit length of the line section between the relaying and the fault points, and Z_{1L} is the line positive sequence impedance in ohms.

In the case of a non-zero fault resistance, the measured impedance is not equal to the mentioned value. In this case, the structural and operational conditions of the power system affect the measured impedance at the relaying point. The structural conditions are evaluated by short circuit levels at the line ends, S_{SA} and S_{SB}. The operational conditions prior to the fault instance can be represented by the load angle of the line, δ, and the ratio of the magnitude of the line end voltages, h, or totally $E_B / E_A = he^{-j\delta}$. The measured impedance can be expressed by the following equations. These equations are achieved by following the same procedure presented in [2].

In the case of a phase to phase fault, the measured impedance at the relaying point is:

$$Z_A^{bc} = pZ_{1L} + \frac{R_f}{C_{ld} + 2C_1} \tag{1}$$

On the other hand, in the case of a three phase fault, the measured impedance is:

$$Z_A = pZ_{1L} + \frac{R_f}{C_{ld} + C_1} \tag{2}$$

The above equations show that in the case of zero fault resistance, the measured impedance is equal to its actual value and otherwise it deviates from its actual value depending on the fault resistance and the system conditions. As the fault resistance increases, the measured impedance deviates more.

In the case of a double circuit line, see Fig. 2, the measured impedance is the same as (1) and (2), but the elements of the measured impedance are different in spite of the same general form.

As mentioned, (1) and (2) presents the measured impedance for the faults on the protected line, for a single circuit or double circuit line. But in the case of the second

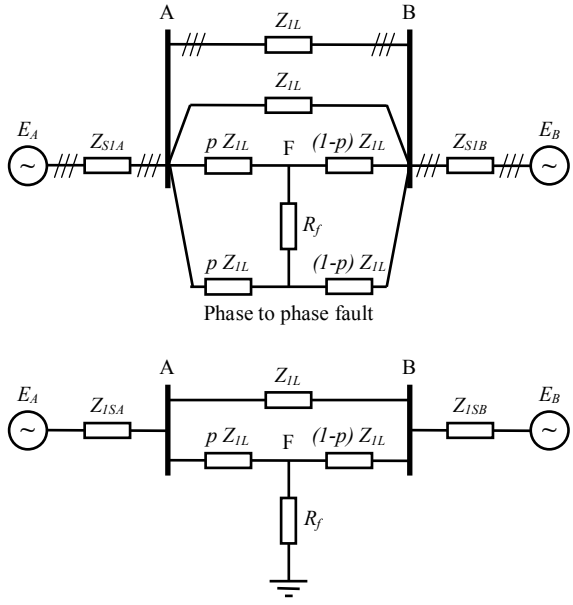


Fig. 2. Inter phase faults equivalent circuit, double circuit line

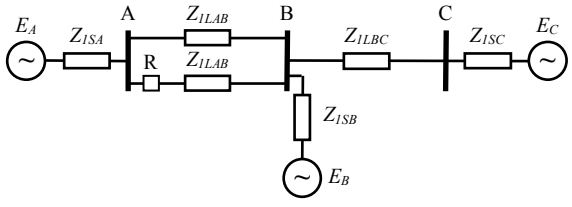


Fig. 3. Double line model

and third zones, the second circuit and the next line also should be considered. Therefore, it is necessary to modify the power system model. Fig. 3 shows a power system model consists of two lines, a single circuit and a double circuit line, three buses and three sources.

According to Fig. 3, the measured impedance for the faults on the next line and the second circuit can be evaluated. Here, also it is necessary to determine the power system conditions for impedance evaluation. Similar to the case of the single line model, structural conditions are represented by short circuit levels at three buses, S_{SA}, S_{SB}, and S_{SC}. The pre-fault operational condition is expressed by the lines load angles, δ_B and δ_C, and the ratios of the voltage of the buses, h_B and h_C, or totally $E_B / E_A = h_B e^{-j\delta_B}$ and $E_C / E_A = h_C e^{-j\delta_C}$.

Here, independent of the fault location, Z_{1AB}, Z_{1BC} and Z'_{1BC} are defined as:

$$Z_{1AB} = Z_{1SA} + Z_{1LAB} \tag{3}$$

$$Z_{1BC} = Z_{1SC} + Z_{1LBC} / 2 \tag{4}$$

$$Z'_{1BC} = \frac{Z_{1SB} Z_{1BC}}{Z_{1SB} + Z_{1BC}} \tag{5}$$

The measured impedance by the relay placed on Line AB at Bus A for the faults on Line BC can be evaluated as:

$$Z_{1AF} = Z_{1SA} + Z_{1LAB} / 2 + pZ_{1LBC} \tag{6}$$

$$Z_{1BF} = pZ_{1LBC} \tag{7}$$

$$Z_{1CF} = Z_{1SC} + (1-p)Z_{1LBC} \tag{8}$$

$$Z_{1A} = Z_{1BF} + \frac{Z_{1SB} Z_{1AB}}{Z_{1SB} + Z_{1AB}} \quad (9)$$

$$Z_{1C} = Z_{1CF} \quad (10)$$

$$Z_1 = \frac{Z_{1A} Z_{1C}}{Z_{1A} + Z_{1C}} \quad (11)$$

$$C_{1B} = \frac{Z_{1C}}{Z_{1A} + Z_{1C}} \quad (12)$$

$$C_{1A} = \frac{Z_{1SB}}{2(Z_{1SB} + Z_{1AB})} \quad (13)$$

$$Den = Z_{1AB}[Z_{1BF} h_C e^{-j\delta_C} + Z_{1CF} h_B e^{-j\delta_B}] + Z_{1SB}[Z_{1AF} h_C e^{-j\delta_C} + Z_{1CF}] \quad (14)$$

$$K_{ld} = Z_{1BC}[1 - h_B e^{-j\delta_B}] + Z_{1SB}[1 - h_C e^{-j\delta_C}] \quad (15)$$

$$C_{ld} = K_{ld}(Z_\Sigma + R_f)/2Den \quad (16)$$

$$K_{ld_\Delta} = 2Z_{1AB}[h_B e^{-j\delta_B} - h_C e^{-j\delta_C}] - Z_{1BC}[1 - h_B e^{-j\delta_B}] + Z_{1SB}[1 - h_C e^{-j\delta_C}] \quad (17)$$

$$C_{ld_\Delta} = K_{ld_\Delta}(Z_\Sigma + R_f)/2Den \quad (18)$$

In the case of a phase to phase fault, the measured impedance at the relaying point is:

$$Z_\Sigma = 2Z_1 \quad (19)$$

$$C_{Sh} = Z_{1BF}[C_{ld_\Delta} + 2C_{1B}(1 - C_{1A})] \quad (20)$$

$$Z_{A_1}^{BC,bc} = Z_{1LAB} + pZ_{1LBC} + \frac{C_{Sh} + R_f}{C_{ld} + 2C_{1A}C_{1B}} \quad (21)$$

Otherwise, in the case of a three phase fault, the measured impedance is:

$$Z_\Sigma = Z_1 \quad (22)$$

$$C_{Sh} = Z_{1BF}[C_{ld_\Delta} + C_{1B}(1 - C_{1A})] \quad (23)$$

$$Z_{A_1}^{BC} = Z_{1LAB} + pZ_{1LBC} + \frac{C_{Sh} + R_f}{C_{ld} + C_{1A}C_{1B}} \quad (24)$$

On the other hand, the measured impedance by the relay placed on Line AB at Bus A for the faults on the second circuit of Line AB (AB_2) can be evaluated as:

$$Z_{1AF} = Z_{1SA} + pZ_{1LAB}/2 \quad (25)$$

$$Z_{1BF} = (1 - p)Z_{1LAB}/2 \quad (26)$$

$$Z_{1CF} = Z_{1SC} + Z_{1LBC} + (1 - p)Z_{1LAB}/2 \quad (27)$$

$$Z_{1A} = Z_{1AF} \quad (28)$$

$$Z_{1C} = Z_{1BF} + Z'_{1BC} \quad (29)$$

$$Z_1 = \frac{Z_{1A} Z_{1C}}{Z_{1A} + Z_{1C}} + \frac{p(1-p)}{2} Z_{1LAB} \quad (30)$$

$$C_{1A} = \frac{Z_{1C}}{Z_{1A} + Z_{1C}} \quad (31)$$

$$C_1 = \frac{(1-p)(Z_{1SA} + Z_{1LAB}) + (2-p)Z'_{1BC}}{2(Z_{1A} + Z_{1C})} \quad (32)$$

$$Den = Z_{1BC}[Z_{1AF} h_B e^{-j\delta_B} + Z_{1BF}] + Z_{1SB}[Z_{1AF} h_C e^{-j\delta_C} + Z_{1CF}] \quad (33)$$

$$K_{ld} = Z_{1BC}[1 - h_B e^{-j\delta_B}] + Z_{1SB}[1 - h_C e^{-j\delta_C}] \quad (34)$$

$$C_{ld} = K_{ld}(Z_\Sigma + R_f)/2Den \quad (35)$$

In the case of a phase to phase fault, the measured impedance at the relaying point is:

$$Z_\Sigma = 2Z_1 \quad (36)$$

$$C_{Sh} = pZ_{1LAB}[2(C_1 - C_{1A})] \quad (37)$$

$$Z_{A_1}^{AB_2,bc} = pZ_{1LAB} + \frac{C_{Sh} + R_f}{C_{ld} + 2(C_{1A} - C_1)} \quad (38)$$

Otherwise, in the case of a three phase fault, the measured impedance is:

$$Z_\Sigma = Z_1 \quad (39)$$

$$C_{Sh} = pZ_{1LAB}[(C_1 - C_{1A})] \quad (40)$$

$$Z_{A_1}^{AB_2} = pZ_{1LAB} + \frac{C_{Sh} + R_f}{C_{ld} + (C_{1A} - C_1)} \quad (41)$$

III. Ideal Tripping Characteristic

Knowing the power system conditions, the measured impedance at the relaying point, ideal tripping characteristic, could be introduced. The tripping characteristic shows the measured impedance for various fault locations, on both Lines BC and AB_2, while the fault resistance varies from 0 to 200 ohms.

The tripping characteristic for the faults on Lines BC and AB_2 are presented for a test system. Two 400 kV transmission lines, AB and BC, with the lengths of 200 and 300 km have been used in this study. By utilizing the Electro-Magnetic Transient Program (EMTP) [10] various sequence impedances of these lines are evaluated according to their physical dimensions. The calculated impedances are:

Z_{1LAB} = 0.0134 + j 0.3114 Ω/km
Z_{1LBC} = 0.0113 + j 0.3037 Ω/km

The structural conditions of the system are:

Z_{1SA} = 1.3945 + j 15.9391 Ω
Z_{1SB} = 13.945 + j 159.391 Ω
Z_{1SC} = 0.6972 + j 7.9696 Ω

The ideal tripping characteristic for the faults on the next line, Line BC, and the second circuit, Line AB_2, are presented in the following sections.

A. Faults on Line BC

Fig. 4 shows the ideal tripping characteristic for phase to phase faults on the next line, Line BC. Here, the ratio of voltage magnitudes, h_B and h_C, are equal to 0.98 and 0.96; and the load angles, δ_B and δ_C, are 16° and 32°, respectively. The quadrilateral characteristic set to 80% of Line AB and the impedances of Lines AB and BC are shown in Fig. 4.

On the other hand, Fig. 5 shows the tripping characteristic in the case of negative load angles for phase to phase faults on the next line. Here, the ratios of voltage magnitudes are 1.02 and 1.04; and the load angles are – 16° and – 32°.

It can be seen that in the case of positive load angles, the tripping characteristic for the faults on Line BC has a quasi-quadrilateral shape. On the other hand, in the case of negative load angles, the tripping characteristic for the faults on Line BC has a crescent shape. Here, the deviation of the measured impedance for zero fault resistance could be observed vividly.

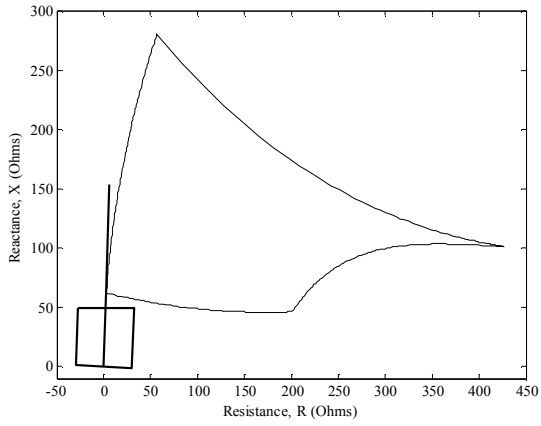


Fig. 4. Tripping characteristic, phase to phase faults on Line BC

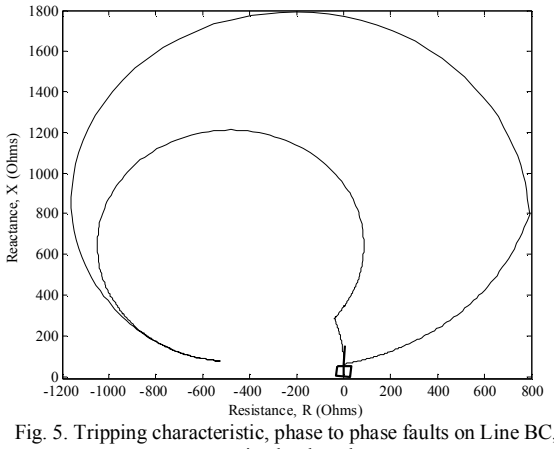


Fig. 5. Tripping characteristic, phase to phase faults on Line BC, negative load angles

On the other hand, Fig. 6 shows the ideal tripping characteristic for three phase faults on the next line. Here, the operational conditions are the same as Fig. 4. In addition, the quadrilateral characteristic set to 80% of Line AB and the impedances of Lines AB and BC are shown in Fig. 6.

On the other hand, Fig. 7 shows the tripping characteristic in the case of the negative load angles for three phase faults on Line BC. Here, the power conditions are the same as Fig. 5.

It can be seen that the ideal tripping characteristic in the case of three phase faults has the same general shape to that of in the case of phase to phase faults. In spite of general similarities, the ideal tripping characteristics in the two cases are somehow different.

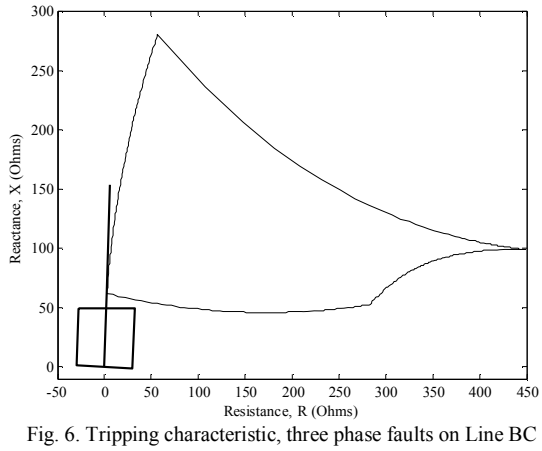


Fig. 6. Tripping characteristic, three phase faults on Line BC

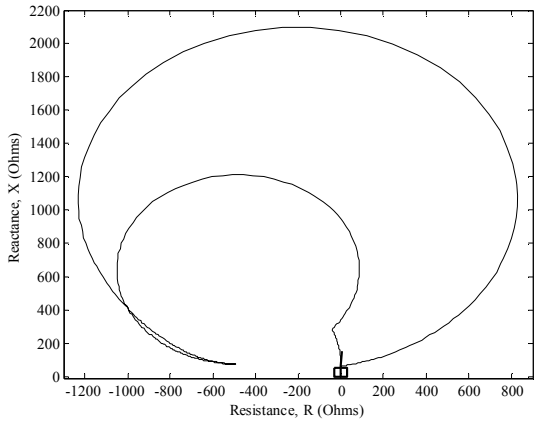


Fig. 7. Tripping characteristic, three phase faults on Line BC, negative load angles

B. Faults on Line AB_2

Fig. 8 shows the ideal tripping characteristic for phase to phase faults on the second circuit. Here, the ratio of voltage magnitudes, h_B and h_C, are 0.98 and 0.96; and the load angles, δ_B and δ_C, are 16° and 32°, respectively. The quadrilateral characteristic set to 80% of Line AB and the impedance of Line AB are shown in Fig. 8. In addition, the dotted region in the figure is the ideal tripping characteristic.

The close look of Fig. 8 around the quadrilateral characteristic is shown in Fig. 9. The measured impedance in the case of zero fault resistance could be seen here, the quasi-oval curve.

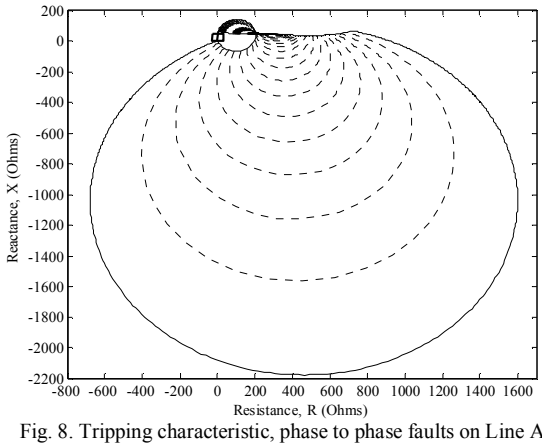


Fig. 8. Tripping characteristic, phase to phase faults on Line AB_2

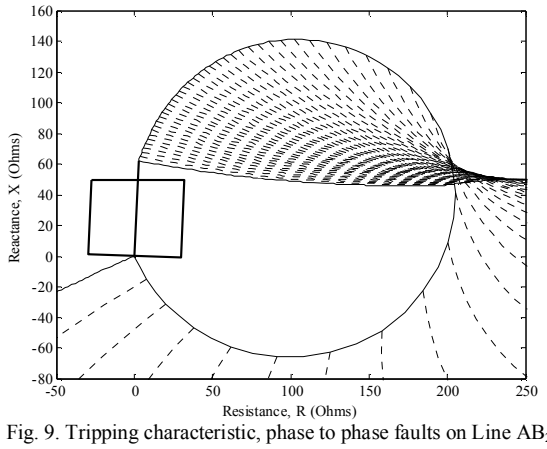


Fig. 9. Tripping characteristic, phase to phase faults on Line AB_2, close look

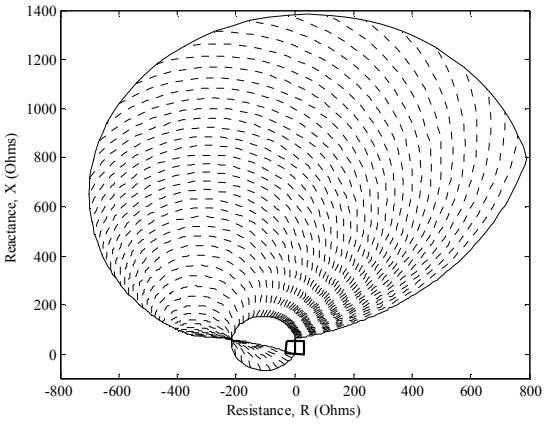


Fig. 10. Tripping characteristic, phase to phase faults on Line AB_2, negative load angle

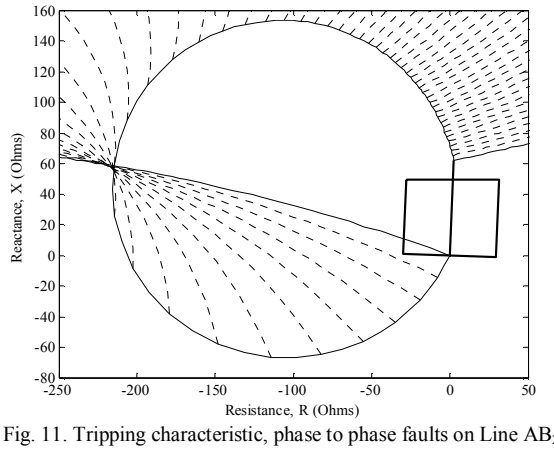


Fig. 11. Tripping characteristic, phase to phase faults on Line AB_2, negative load angle, close look

Otherwise, Fig. 10 shows the tripping characteristic in the case of negative load angles for phase to phase faults on the second circuit, Line AB_2. Here, the ratios of voltage magnitudes are 1.02 and 1.04; and the load angles are – 16° and – 32°.

The close look of Fig. 10 around the quadrilateral characteristic is shown in Fig. 11. The measured impedance in the case of zero fault resistance could be seen here, the quasi-oval curve.

It can be seen that, in the case of negative load angles, the tripping characteristic is changed considerably, but it is still somehow similar to the tripping characteristic in the case of positive load angles. It can be said that the tripping characteristic turns 180°.

On the other hand, Fig. 12 shows the ideal tripping characteristic for three phase faults on the second circuit. Here, the operational conditions are the same as Fig. 8. The quadrilateral characteristic set to 80% of Line AB is plotted. The close look of Fig. 12 is shown in Fig. 13.

Fig. 14 shows the tripping characteristic in the case of negative load angles for three phase faults on the second circuit. Here, the ratios of voltage magnitudes are 1.02 and 1.04; the load angles are – 16° and – 32°, the same as Fig. 11. The close look of Fig. 14 is shown in Fig. 15.

It can be seen that the ideal tripping characteristic in the case of three phase faults has the same general shape to that of in the case of phase to phase faults. In spite of general similarities, the ideal tripping characteristics in the two cases are somehow different.

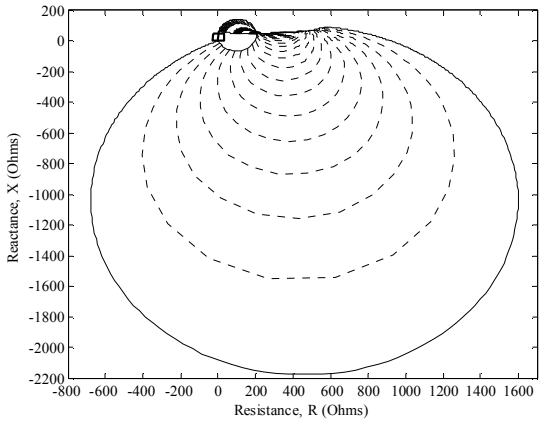


Fig. 12. Tripping characteristic, three phase faults on Line AB_2

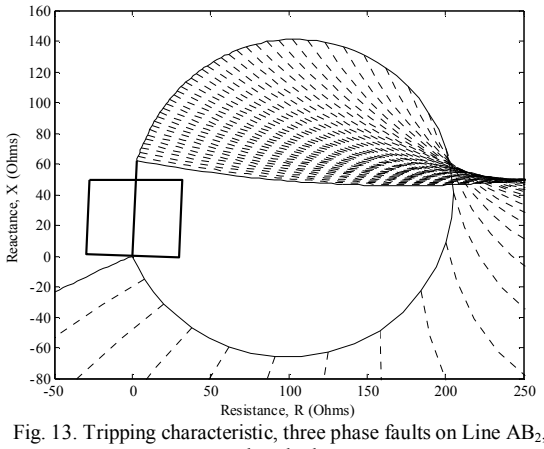


Fig. 13. Tripping characteristic, three phase faults on Line AB_2, close look

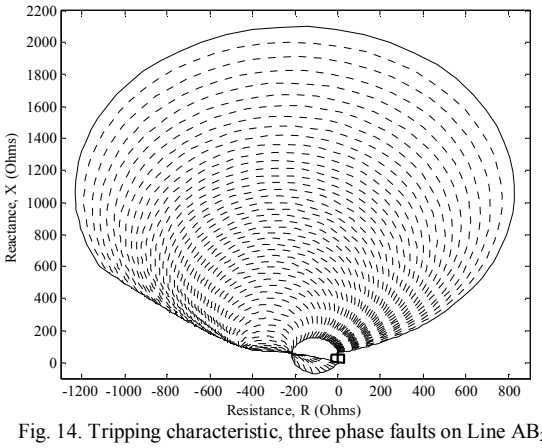


Fig. 14. Tripping characteristic, three phase faults on Line AB_2, negative load angle

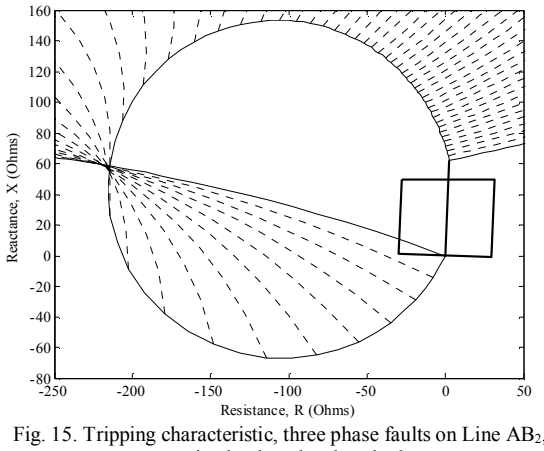


Fig. 15. Tripping characteristic, three phase faults on Line AB_2, negative load angle, close look

IV. Conclusion

This paper presents the measured impedance at the relaying point by a distance relay on a double circuit line for the faults on the next line and the faults on the second circuit, in the case of the inter phase, phase to phase and three, faults. The ideal tripping characteristic is presented for the two cases. In addition, the variation of the measured impedance due to changes in power system conditions is investigated.

In the case of faults on the next line, the measured impedance deviates considerably, for a fault at Bus C, it is not equal to the summation of the impedances of Lines AB and BC. The tripping characteristic depends greatly to the operational conditions of the power system.

In the case of faults on the second circuit, the measured impedance deviates more considerably, compared with the previous case. Unlike the previous case, the tripping characteristic has the same general form for positive and negative load angels.

The next line and the second circuit are both connected to the far end of the protected line, but the measured impedance in these tow cases is significantly different. In spite of the measured impedance great deviation, there is a very small overlapping between the ideal tripping and quadrilateral characteristics. In the case of zero fault resistance, the probability of the distance relay mal-operation is zero for both cases.

It can be concluded that the ideal tripping characteristic in the case of three phase faults has the same general shape to that of in the case of phase to phase faults. In spite of general similarities, the ideal tripping characteristics in the two cases are somehow different. Usually, the close look of the tripping characteristic in the two cases of phase to phase and three phase faults are the same, but the whole tripping characteristics are somehow different.

References

[1] Zhang Zhizhe and C. Deshu, "An adaptive approach in digital distance protection", *IEEE Trans. Power Delivery*, vol. 6, no. 1, pp. 135–142, Jan. 1991.

[2] Y. Q. Xia, K. K. Li, and A. K. David, "Adaptive relay setting for stand-alone digital distance protection", *IEEE Trans. Power Delivery*, vol. 9, no. 1, pp. 480–491, Jan. 1994.

[3] S. Jamali, "A fast adaptive digital distance protection", in *Proc. 2001 IEE 7th International Conference on Developments in Power System Protection, DPSP2001*, pp. 149–152.

[4] Chang-Ho Jung, Dong-Joon Shin, and Jin-O Kim, "Adaptive setting of digital relay for transmission line protection", in *Proc. 2000 IEEE International Conference on Power System Technology, PowerCon2000*, vol. 3, pp. 1465–1468.

[5] K. K. Li, L. L. Lai, and A. K. David, "Stand alone intelligent digital distance relay", *IEEE Trans. Power Systems*, vol. 15, no. 1, pp. 137–142, Feb. 2000.

[6] P. J. Moore, R. K. Aggarwal, H. Jiang, and A. T. Johns, "New approach to distance protection for resistive double-phase to earth faults using adaptive techniques", *IEE Proceedings Generation, Transmission and Distribution*, vol. 141, no. 4, pp. 369–376, July 1994.

[7] Yi Hu, D. Novosel, M. M. Saha, and V. Leitloff, "An adaptive scheme for parallel-line distance protection", *IEEE Trans. Power Delivery*, vol. 17, no. 1, pp. 105–110, Jan. 2002.

[8] Yi Hu, D. Novosel, M. M. Saha, and V. Leitloff, "Improving parallel line distance protection with adaptive techniques", in *Proc. 2000 Power Engineering Society Winter Meeting*, vol. 3, pp. 1973–1978, Jan. 2000.

[9] A. G. Jongepier and L. van der Sluis, "Adaptive distance protection of a double-circuit line", *IEEE Trans. Power Delivery*, vol. 9, no. 3, pp. 1289–1297, July. 1994.

[10] H. W. Dommel, "EMPT reference manual", Microtran Power System Analysis Corporation, Vancouver, British Columbia, Canada, August 1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Distance Relay Ideal Tripping Characteristic for Inter Phase Faults in Presence of SSSC on Next Line

A. Kazemi, S. Jamali, and H. Shateri

Centre of Excellence for Power Systems Automation and Operation Department of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak 16846, Tehran, Iran
kazemi@iust.ac.ir, sjamali@iust.ac.ir, and shateri@iust.ac.ir

Abstract - This paper discusses the distance relay ideal tripping characteristic in the presence of Static Synchronous Series Compensator (SSSC), one of the series connected Flexible Alternating Current Transmission System (FACTS) devices, on the next line, in the case of inter phase, phase to phase and three phase, faults. Especial attention is paid to the over-reaching problem. This is done by presenting the measured impedance at the relaying point in the presence of SSSC at the near end of the next line. The measured impedance at relaying point is greatly influenced in the presence of SSSC on the protected line or even on the near end of the next line. The measured impedance at relating point depends on many factors including the power system structural conditions, pre-fault loading, fault resistance, and SSSC structural and controlling parameters.

I. Introduction

The distance protection operates based on the measured impedance at the relaying point. There are several factors affecting the measured impedance at the relaying point in the case of first or other zones. Some of these factors are related to the power system parameters prior to the fault instance [1]-[3], which can be categorized into two groups. In the case of the first zone, the first group is the structural conditions of the line, represented by the short circuit levels at the transmission line ends, whereas the second group is the pre-fault operational conditions, represented by the line load angle and the voltage magnitude ratio at the line ends. In the case of second zone, two lines are included in the measured impedance evaluation. Therefore, the structural conditions should be presented in the form of short circuit levels at the three buses, and the operational conditions prior the fault instance are represented in the form of the lines load angles and the voltage magnitude ratios at the line ends.

In addition to the power system parameters, the fault resistance could greatly influence the measured impedance. In the case of first zone, when the fault resistance is equal to zero, the power system parameters do not affect the measured impedance, i.e. in the absence of the fault resistance the measured impedance is equal to the actual impedance of the line section between the relaying and the fault points. But for the second zone, even in the absence of the fault resistance, the measured impedance could greatly be deviated from its actual value. Therefore, in the case of the second zone and for the faults on the next line, the evaluation of the measured impedance is more complicated and more factors are included, in comparison with the case of first zone.

More than 70% of the occurred faults on the overhead transmission lines are of the single phase to ground type. However, phase to phase faults are the most common fault type after the single phase to ground faults. In the case of phase to phase faults, inter phase fault resistance depends on the quality of the fault arc creation between the faulted phases. The other types of the faults (double phase to ground and three phase faults) are less common in the transmission systems. The three phase fault is the least common fault type.

In the recent years, FACTS devices are introduced to enhance the stability of power systems by means of increasing the transmission capacity of the lines and provide the optimum utilization of the system capability. This is done by pushing the power systems to their thermal limits [4]. It is well documented in the literature that the introduction of FACTS devices into power systems has a great influence on their dynamics. Therefore, it is essential to study effects of FACTS devices on protective systems, especially the distance protection, which is the main protective device at HV and EHV levels.

It is well-known that the presence of series capacitors on the transmission lines would lead to the over-reaching of the distance relays located on the adjacent lines. This is true for both fixed and controllable series capacitors. The probability of over-reaching due to presence of TCSC at the near end of the next line is mentioned in [5]-[6]. In [5]-[6] only the probability of over-reaching is discussed by means of the simulations and no equations are presented. On the other hand, series compensation could be performed via a series voltage source, the case of SSSC or UPFC. Therefore, the measured impedance in the presence of SSSC at the near end of the next line is studied. Here it is assumed that the protective system operates before SSSC control system.

This paper presents the measured impedance at the relaying point in the presence of SSSC at the near end of the next line in the case of inter phase, phase to phase and three phase, faults; and discusses the probability of the distance relay over-reaching. In addition to SSSC structural and controlling parameters, the measured impedance depends on the power system structural and operational conditions and especially the fault resistance.

II. SSSC and its Modelling

As mentioned, Static Synchronous Series Compensator (SSSC) is placed in the group of series connected FACTS devices. As shown in Fig. 1, SSSC consists of a voltage source inverter connected in series through a coupling transformer to the transmission line. A source of energy is required for providing and maintaining the dc voltage across the dc capacitor and SSSC losses compensation [7].

Fig. 2 shows the equivalent circuit of SSSC which consists of a series connected voltage source in series with an impedance. This impedance represents the impedance of SSSC coupling transformer.

When energy source only has the ability of maintaining the dc voltage and supplying the losses, SSSC only could compensate the reactive power. In this case the magnitude of injected voltage can be controlled due to compensation strategy, but the phase angle of the injected voltage would be perpendicular to the line current. The injected voltage could either lead or lag the line current by 90°.

III. Measured Impedance at Relaying Point

Distance relay operates based on the measured impedance at the relaying point. In the case of faults in the first zone and in the absence of the fault resistance, the measured impedance by a distance relay is the actual impedance of the line section between the fault and the relaying points. According to Fig. 3, this impedance is equal to pZ_{1L}, where p is per unit length of the line section between the fault and the relaying points, and Z_{1L} is the line positive sequence impedance in ohms.

In the case of a non-zero fault resistance, the measured impedance at the relaying point is not equal to the mentioned value. In this case, the structural and operational conditions of the power system affect the measured impedance at the relaying point. The structural conditions are evaluated by short circuit levels at the line ends, S_{SA} and S_{SB}. The operational conditions prior to the fault instance can be represented by the load angle of the line, δ, and the ratio of the magnitude of the line end voltages, h, totally $E_B / E_A = he^{-j\delta}$.

The measured impedance could be expressed by the following equations. These equations are driven by following the presented procedure in [2].

In the case of a phase to phase fault, the measured impedance at the relaying point is:

$$Z_A^{bc} = p Z_{1L} + \frac{R_f}{C_{ld} + 2C_1} \quad (1)$$

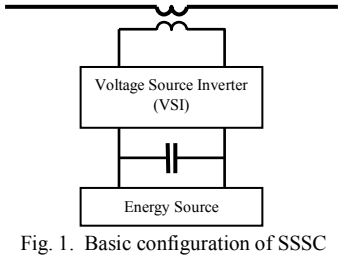


Fig. 1. Basic configuration of SSSC

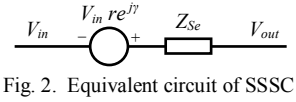


Fig. 2. Equivalent circuit of SSSC

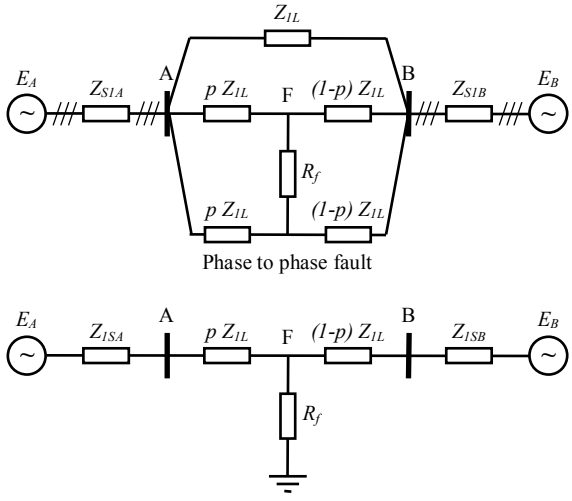


Fig. 3. Inter phase faults equivalent circuit

On the other hand, in the case of a three phase fault, the measured impedance is:

$$Z_A = p Z_{1L} + \frac{R_f}{C_{ld} + C_1} \quad (2)$$

The above equations show that in the case of zero fault resistance, the measured impedance at the relaying point is equal to its actual value and otherwise it deviates from its actual value depending on the fault resistance and the power system conditions. As the fault resistance increases, the measured impedance deviates more.

As mentioned, (1)-(2) present the measured impedance for the faults in the first zone. But in the case of the second zone, the next line should be considered, since this zone covers a part of the adjacent lines. Therefore, it is necessary to modify the power system of Fig. 3 to the system of Fig. 4, which shows a power system consists of two lines, three buses and three sources.

According to Fig. 4, the measured impedance for the faults on the next line can be evaluated. Here, also it is necessary to determine the power system conditions for impedance evaluation. Similar to the case of single line model, structural conditions are represented by short circuit levels at three buses, S_{SA}, S_{SB}, and S_{SC}. The pre-fault operational condition are expressed by the lines load angles, δ_B and δ_C, and the ratios of the voltage magnitudes of the buses, h_B and h_C, or generally $E_B / E_A = h_B e^{-j\delta_B}$ and $E_C / E_A = h_C e^{-j\delta_C}$. The measured impedance at Bus A for the faults on Line BC can be expressed by the following equations:

$$Z_{1AB} = Z_{1SA} + Z_{1LAB} \quad (3)$$

$$Z_{1BC} = Z_{1SC} + Z_{1LBC} \quad (4)$$

$$Z_{1AF} = Z_{1SA} + Z_{1LAB} + p Z_{1LBC} \quad (5)$$

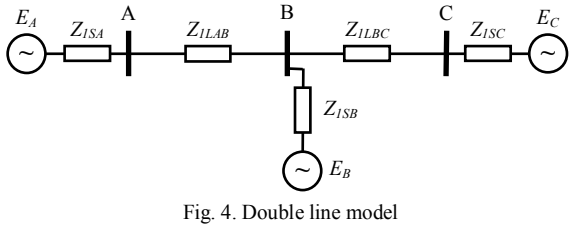


Fig. 4. Double line model

$$Z_{1BF} = p\,Z_{1LBC} \quad (6)$$

$$Z_{1CF} = Z_{1SC} + (1-p)\,Z_{1LBC} \quad (7)$$

$$Z_{1A} = Z_{1BF} + \frac{Z_{1SB}\,Z_{1AB}}{Z_{1SB} + Z_{1AB}} \quad (8)$$

$$Z_{1C} = Z_{1CF} \quad (9)$$

$$Z_1 = \frac{Z_{1A}Z_{1C}}{Z_{1A} + Z_{1C}} \quad (10)$$

$$C_{1B} = \frac{Z_{1C}}{Z_{1A} + Z_{1C}} \quad (11)$$

$$C_{1A} = \frac{Z_{1SB}}{Z_{1SB} + Z_{1AB}} \quad (12)$$

$$Den = Z_{1AB}[\,Z_{1BF}\,h_C e^{-j\delta_C} + Z_{1CF}\,h_B e^{-j\delta_B}\,] + Z_{1SB}[\,Z_{1AF}\,h_C e^{-j\delta_C} + Z_{1CF}\,] \quad (13)$$

$$K_{ld} = Z_{1BC}[\,1 - h_B e^{-j\delta_B}\,] + Z_{1SB}[\,1 - h_C e^{-j\delta_C}\,] \quad (14)$$

$$C_{ld} = (Z_\Sigma + R_f)K_{ld} / Den \quad (15)$$

$$K_{ld_\Delta} = Z_{1AB}[\,h_B e^{-j\delta_B} - h_C e^{-j\delta_C}\,] - Z_{1BC}[\,1 - h_B e^{-j\delta_B}\,] \quad (16)$$

$$C_{ld_\Delta} = (Z_\Sigma + R_f)K_{ld_\Delta} / Den \quad (17)$$

In the case of a phase to phase fault, the measured impedance at the relaying point is:

$$Z_\Sigma = 2Z_1 \quad (18)$$

$$C_{Sh} = Z_{1BF}[\,C_{ld_\Delta} + 2C_{1B}(1 - C_{1A})] \quad (19)$$

$$Z_A^{BC,bc} = Z_{1LAB} + p\,Z_{1LBC} + \frac{C_{Sh} + R_f}{C_{ld} + 2C_{1B}C_{1A}} \quad (20)$$

Otherwise, in the case of a three phase fault, the measured impedance is:

$$Z_\Sigma = Z_1 \quad (21)$$

$$C_{Sh} = Z_{1BF}[\,C_{ld_\Delta} + C_{1B}(1 - C_{1A})] \quad (22)$$

$$Z_A^{BC} = Z_{1LAB} + p\,Z_{1LBC} + \frac{C_{Sh} + R_f}{C_{ld} + C_{1B}C_{1A}} \quad (23)$$

In (20)-(23), for zero fault resistance the measured impedance is not equal to the impedance of the line section between the fault and the relaying points.

Once SSSC is installed at the near end of the next line, (4), (6), (13)-(14), and (16) should be modified and some new equations are introduced:

$$C_{1S} = Z_{Se} / Z_{1LBC} \quad (24)$$

$$Z_{1BC} = Z_{1SC} + (1 + C_{1S})Z_{1LBC} \quad (25)$$

$$Z_{1BF} = (p + C_{1S})Z_{1LBC} \quad (26)$$

$$Den = Z_{1AB}[\,Z_{1BF}h_C e^{-j\delta_C} + Z_{1CF}(1 + re^{j\gamma})h_B e^{-j\delta_B}\,] + Z_{1SB}\begin{bmatrix}[\,Z_{1AB}(1 + re^{j\gamma}) + Z_{1BF}\,]h_C e^{-j\delta_C} \\ + Z_{1CF}(1 + re^{j\gamma})\end{bmatrix} \quad (27)$$

$$K_{ld} = Z_{1BC}[\,1 - h_B e^{-j\delta_B}\,] + Z_{1SB}[\,1 + re^{j\gamma} - h_C e^{-j\delta_C}\,] \quad (28)$$

$$K_{ld_\Delta} = Z_{1AB}[(1 + re^{j\gamma})h_B e^{-j\delta_B} - h_C e^{-j\delta_C}\,] - Z_{1BC}[\,1 - h_B e^{-j\delta_B}\,] \quad (29)$$

$$K_{V_{Se}} = Z_{1AB}Z_{1BC}h_B e^{-j\delta_B} + Z_{1SB}[\,Z_{1AB}h_C e^{-j\delta_C} + Z_{1BC}\,] \quad (30)$$

$$C_{V_{Se}} = -(Z_\Sigma + 3R_f)K_{V_{Se}}re^{j\gamma} / Den \quad (31)$$

In the case of a phase to phase fault, (20) should be modified as:

$$Z_A^{BC,bc} = Z_{1LAB} + (p + C_{1S})Z_{1LBC} + \frac{C_{Sh} + C_{V_{Se}} + R_f}{C_{ld} + 2C_{1B}C_{1A}} \quad (32)$$

Otherwise, in the case of a three phase fault, (23) should be modified as:

$$Z_A^{BC,bc} = Z_{1LAB} + (p + C_{1S})Z_{1LBC} + \frac{C_{Sh} + C_{V_{Se}} + R_f}{C_{ld} + C_{1B}C_{1A}} \quad (33)$$

In (32)-(33), in the case of zero fault resistance, the measured impedance at the relaying point is not equal to the impedance of the line section between the relaying and the fault points. Here, the measured impedance deviates from its actual value more than that of the pervious case in such a way that the distance relay might be subjected to over-reaching.

IV. Effects of SSSC on Distance Relay Ideal Tripping Characteristic and Over-Reaching Problem

The impacts of the presence of SSSC on Line BC at Bus B, which might lead to mal-operation, have been tested for a test system. Two 400 kV transmission lines, AB and BC, with the lengths of 200 and 300 km have been used in this study. By utilizing the Electro-Magnetic Transient Program (EMTP) [11] various sequence impedances of these lines are evaluated according to their physical dimensions. The calculated impedances of various sequences of the lines and the other parameters of the test system are:

Z_{1LAB} = 0.0201 + j 0.2868 Ω/km
Z_{1LBC} = 0.0113 + j 0.3037 Ω/km
Z_{1SA} = 1.3945 + j 15.939 Ω
Z_{1SB} = 13.945 + j 159.39 Ω
Z_{1SC} = 0.6973 + j 7.9696 Ω
h_B = 0.98
h_C = 0.96
δ_B = 16°
δ_C = 32°

In the absence of SSSC, Fig. 5 shows the measured impedance at the relaying point for the fault resistances between 0 and 200 ohms and the fault point variation from Bus B up to Bus C for phase to phase faults. The impedances of Lines AB and BC are shown in Fig. 5 to make it possible to observe the impedance deviation, especially in the case of zero fault resistance. The quadrilateral characteristic, set to 80% of Line AB, is also plotted in Fig. 5, in order to show that in the absence of SSSC, in spite of measured impedance deviation, distance relay would not over-reach its setting point, in other words there is no overlapping between the tripping and quadrilateral characteristics.

In the absence of SSSC, Fig. 6 shows the measured impedance at the relaying point for the fault resistances from 0 to 200 ohms and for the faults between Bus B and Bus C in the case of three phase faults. The impedances of Lines AB and BC are shown in Fig. 6. The quadrilateral characteristic, set to 80% of Line AB, is also plotted in Fig. 6.

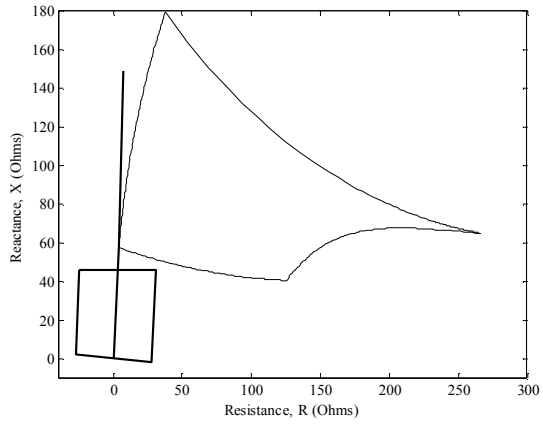


Fig. 5. Tripping characteristic, phase to phase faults, without SSSC

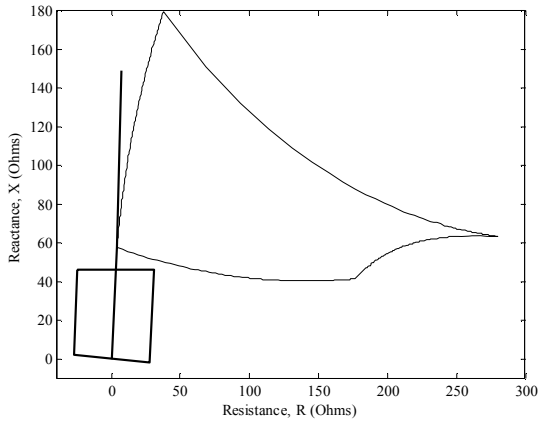


Fig. 6. Tripping characteristic, three phase faults, without SSSC

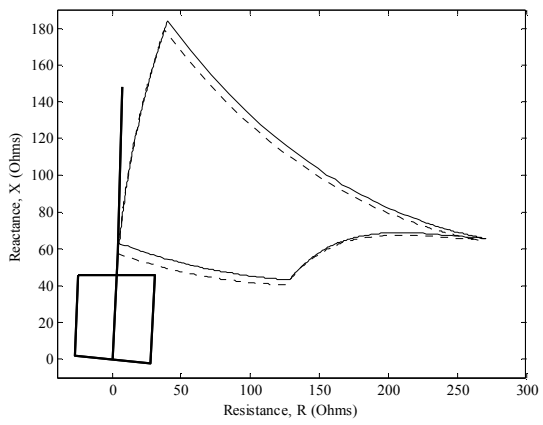


Fig. 7. Tripping characteristic, phase to phase faults, inactive SSSC

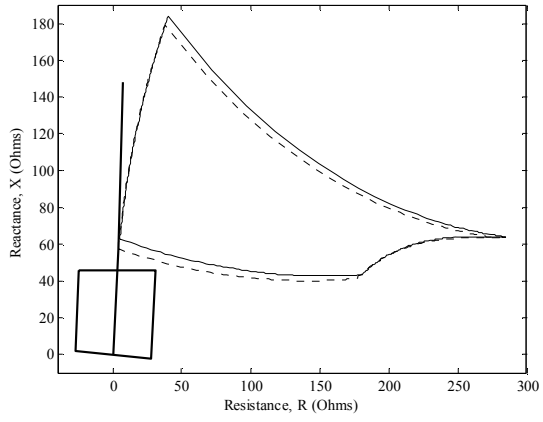


Fig. 8. Tripping characteristic, three phase faults, inactive SSSC

It can be seen that in Fig. 5 and Fig. 6 there is no overlapping region between the ideal tripping and the quadrilateral characteristics.

Usually SSSC injected voltage changes according to its controlling strategy. Therefore, SSSC injected voltage would vary as the loading conditions of the power system changes. But in this study the operational conditions of the power system are taken to be constant and it is assumed that these conditions are achieved by different SSSC injected voltages.

The tripping characteristic in the presence of an inactive SSSC, for phase to phase faults, is shown in Fig. 7, where the tripping characteristic without SSSC is also shown in the dotted form for comparison.

It can be seen that an inactive SSSC would affect the measured impedance at the relaying point. This is due to the presence of the coupling transformer. The tripping characteristic transfers upward. The measured resistance increases slightly as well as the measured reactance. There is no overlapping region between the ideal tripping and quadrilateral characteristics.

The tripping characteristic in the presence of an inactive SSSC, for three phase faults, is shown in Fig. 8, where the tripping characteristic without SSSC is also shown in the dotted form.

The measured impedance deviation due to presence of an inactive SSSC for three phase faults has a same pattern with that of in the case of phase to phase faults. But in spite of general similarities, there are some differences between the two cases.

Fig. 9 shows the effect of SSSC injected voltage amplitude variation on the measured impedance at the relaying point, in the leading mode in the case of phase to phase faults. Here, the amplitude of the injected voltage takes the values 0.00, 0.05, 0.10, and 0.15.

It can be seen that as the injected voltage amplitude increases in the leading mode, the measured resistance decreases for high fault resistances, while it increases in the case of low fault resistances. On the other hand, as the injected voltage amplitude increases, the measured reactance decreases for high fault resistances, while it increases in the case of low fault resistances. Generally, it can be said that in the presence of SSSC in the leading mode at the near end of the next line, the tripping characteristic shrinks and turns in clockwise direction.

It can be seen that as the injected voltage amplitude increases in the leading mode, the ideal tripping characteristic varies considerably, but the distance between it and the quadrilateral characteristic does not change considerably. This distance is more than the distance between two characteristics in the absence of SSSC.

Fig. 10 shows the effect of SSSC injected voltage amplitude variation on the measured impedance at the relaying point, in the leading mode in the case of three phase faults. Here, the amplitude of the injected voltage takes the values 0.00, 0.05, 0.10, and 0.15.

The measured impedance deviation due to SSSC injected voltage amplitude in the leading mode for three phase faults has a same pattern with that of in the case of phase to phase faults.

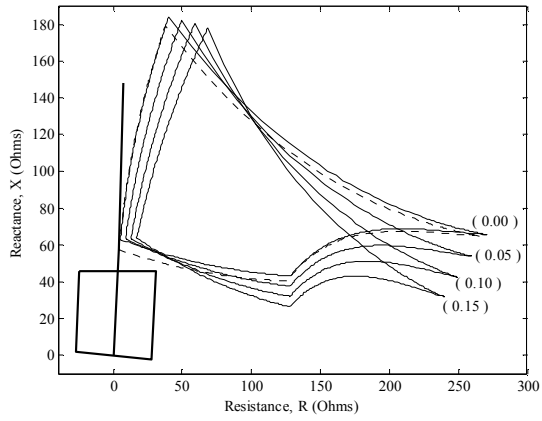


Fig. 9. Tripping characteristic, phase to phase faults, leading SSSC

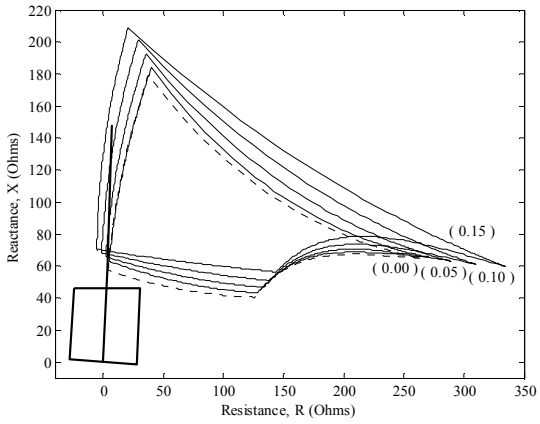


Fig. 11. Tripping characteristic, phase to phase faults, lagging SSSC

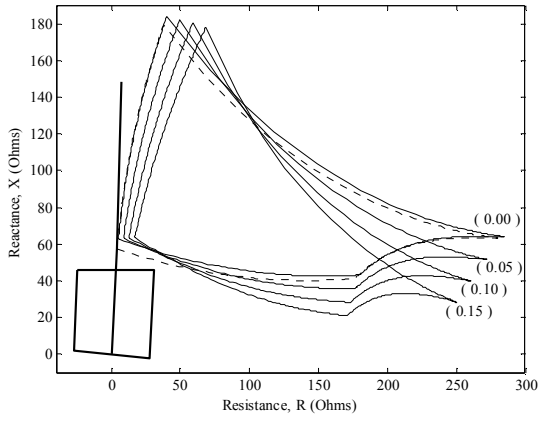


Fig. 10. Tripping characteristic, three phase faults, leading SSSC

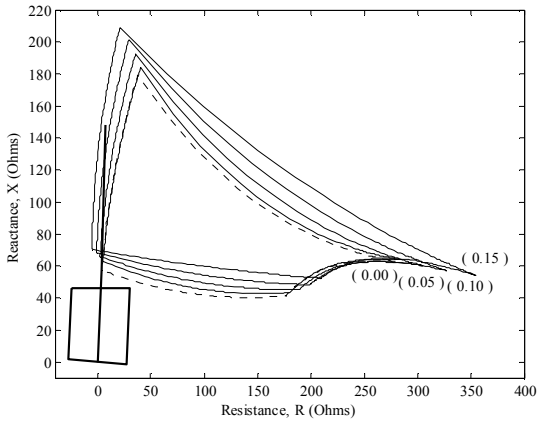


Fig. 12. Tripping characteristic, three phase faults, lagging SSSC

Fig. 11 shows the effect of SSSC injected voltage amplitude variation on the measured impedance at the relaying point, in the lagging mode in the case of phase to phase faults. Here the amplitude of the injected voltage takes the values 0.00, 0.05, 0.10, and 0.15.

It can be seen that as the injected voltage amplitude increases in the lagging mode, the measured resistance increases for the high fault resistances, while it decreases in the case of the low fault resistances; and in the case of zero fault resistance and the high amplitude of injected voltage, the measured resistance becomes negative. On the other hand, as the injected voltage amplitude increases, the measured reactance decreases for high fault resistances, and increases in the case of low fault resistances. Generally, it can be said that in the presence of SSSC in the lagging mode at the near end of the next line, the tripping characteristic expands and turns in anticlockwise direction.

It can be seen that as the injected voltage amplitude increases in the lagging mode, the ideal tripping characteristic varies considerably, and the distance between the tripping and the quadrilateral characteristics increases considerably. This distance is more than the distance between two characteristics in the absence of SSSC.

Fig. 12 shows the effect of SSSC injected voltage amplitude variation on the measured impedance at the relaying point, in the lagging mode in the case of three phase faults. Here the amplitude of the injected voltage takes the values 0.00, 0.05, 0.10, and 0.15.

The measured impedance deviation due to SSSC injected voltage amplitude in the lagging mode for three phase faults has a same pattern with that of in the case of phase to phase faults. But in spite of general similarities, there are some differences between the two cases.

V. Conclusion

This paper evaluates the measured impedance at the relaying point in the presence of SSSC at the near end of the next line in the case of inter phase, phase to phase and three phase, faults. The distance relay ideal tripping characteristic is presented for the mentioned cases. The variation of the tripping characteristic due to the changes in SSSC injected voltage and the probability of over-reaching due to presence of SSSC are investigated.

It can be concluded that the measured impedances for phase to phase and three phase faults vary in a similar pattern due to changes in SSSC injected voltage. This pattern is somehow the same as the pattern of the measured impedance variation in the case of single phase to ground faults, but there are some differences between them.

It can be seen that in the presence of SSSC at the near end of the next line and in the case of the inter phase faults, there is no overlapping between the tripping and the quadrilateral characteristics. Therefore, the probability of distance relay over-reaching does not vary due to the presence of SSSC. It can be concluded that in the case of the inter phase faults, the distance relay could perform its protective duties satisfactorily in the presence of SSSC at

the near end of the next line, i.e. no over-reaching could be observed due to SSSC presence on the next line.

References

[1] Zhang Zhizhe and C. Deshu, “An adaptive approach in digital distance protection”, *IEEE Trans. Power Delivery*, vol. 6, no. 1, pp. 135–142, Jan. 1991.

[2] Y. Q. Xia, K. K. Li, and A. K. David, “Adaptive relay setting for stand-alone digital distance protection”, *IEEE Trans. Power Delivery*, vol. 9, no. 1, pp. 480–491, Jan. 1994.

[3] S. Jamali, “A fast adaptive digital distance protection”, in *Proc. 2001 IEE 7th International Conference on Developments in Power System Protection, DPSP2001*, pp. 149-152.

[4] Khalil El-Arroudi, Geza Joos, and Donald. T. McGillis, "Operation of impedance protection relays with the STATCOM", *IEEE Trans. Power Delivery*, vol. 17, no. 2, pp. 381–387, April 2002.

[5] Wang Weiguo, Yin Xianggen, Yu Jiang, Duan Xianzhong, and Chen Deshu, "The impact of TCSC on distance protection relay", in *Proc. 1998 IEEE International Conference on Power System Technology, POWERCON'98*, vol. 1, pp. 382–388.

[6] M. Khederzadeh, "The impact of FACTS devices on digital multifunction protective relays", in *Proc. 2002 IEEE Conference and Exhibition on Transmission and Distribution, Asia Pacific IEEE/PES*, vol. 3, pp. 2043–2048.

[7] A. T. Johns, A. Ter-Gazarian, and D. F. Warne, *Flexible ac transmission systems (FACTS)*, Padstow, Cornwall: TJ International Ltd., 1999.

[8] P. J. Moore, R. K. Aggarwal, H. Jiang, and A. T. Johns, "New approach to distance protection for resistive double-phase to earth faults using adaptive techniques", *IEE Proceedings Generation, Transmission and Distribution*, vol. 141, no. 4, pp. 369–376, July 1994.

[9] P. K. dash, A. K. Pradhan, G. Panda, and A. C. Liew, “Adaptive relay setting for flexible AC transmission systems (FACTS)”, *IEEE Trans. Power Delivery*, vol. 15, no. 1, pp. 38–43, Jan. 2000.

[10] P. K. dash, A. K. Pradhan, G. Panda, and A. C. Liew, “Digital protection of power transmission lines in the presence of series connected FACTS devices”, in *Proc. 2000 IEEE Power Engineering Society Winter Meeting*, vol. 3, pp. 1967–1972.

[11] H. W. Dommel, “EMPT reference manual”, Microtran Power System Analysis Corporation, Vancouver, British Columbia, Canada, August 1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Optimal Loss Allocation of Multiple Wheeling Transactions in a Deregulated Power System

Pornthep Panyakaew and Parnjit Damrongkulkamjorn

Department of Electrical Engineering, Kasetsart University
50 Phaholyothin Road, Bangkok, Thailand
E-mail: pothpa@yahoo.com and parnjit.d@ku.ac.th

Abstract - This paper presents an approach to optimally allocate transmission losses among the utilities involving in multiple wheeling transactions. The algorithm introduces the calculation of loss ratio according to the losses the individual transaction causes the transmission system when that transaction is taken place alone. During multiple wheeling transactions, the generation output resulting from optimal power flow of the intermediate system remains unchanged. Therefore the losses associated with wheeling transactions are accommodated by supplying utilities. The losses minimizing problem for the intermediate utility is computed with an additional constraint of loss ratio when multiple wheeling transactions occur. The loss ratios are able to proportionally allocate losses for every transaction. The proposed algorithm of the optimal loss allocation for multiple wheeling transactions is tested on a modified IEEE 30 bus system. The results from the algorithm are the optimal real power generated from the supplying utilities which includes their share of losses due to the transactions. The optimal results show that the algorithm can find minimum wheeling losses and divide them among supplying utilities when loss ratio is implemented.

I. Introduction

Electric utilities all over the world are moving toward deregulation and the competition of electricity markets, however, in different paces. Even though some countries such as Thailand and Korea [1] still remain monopolistic policy, they have already introduced the independent power producers (IPPs) in their electricity sector. The IPPs aim to sell their power to the existing utilities and/or to other participants through wheeling transactions of the existing utilities [2]. The wheeling transactions may cause some impacts, including increasing of transmission losses, to the existing utilities as reported in [2]. Recently, several methods for satisfactory allocating transmission losses among all participants in competitive electricity markets have been introduced [1], [3]-[11].

Wheeling has been defined in [12] as "the use of a utility's transmission facilities to transmit power for other buyers and sellers." Generally, single wheeling transaction consists of three parties, one sending utility, one receiving utility and an intermediate utility who has connections to the former two utilities. When the sending and receiving utilities are not directly connected, the intermediate utility can allow the wheeling transaction to take place with agreeable wheeling costs. Multiple wheeling transactions (MWT) involve more than three parties, that is, there can be more than one seller or buyer sending and receiving power through an intermediate utility. For example, utility E directly connects to four different parties as shown in system topology in Fig. 1. In the multiple-connection topology, utility C wants to buy power from utility A, and utility D wants to buy power from utility B through the intermediate utility E. Note that in competitive electricity markets, utilities A and B can either be some electric utilities with their own networks or non-utility generators (or IPPs) connecting to the existing utility.

Multiple wheeling transactions, if allowed by utility E, will be accomplished by drawing a specific amount of real and reactive power out of receiving buses connecting utility E to utilities C and D, while increasing at least the same amount of power from sending buses connecting utility E to utilities A and B. Such transactions will change real and reactive power flow and increase losses of the transmission network of the intermediate utility. The intermediate utility, however, gets compensated for providing transmission services by charging wheeling costs to the participants.

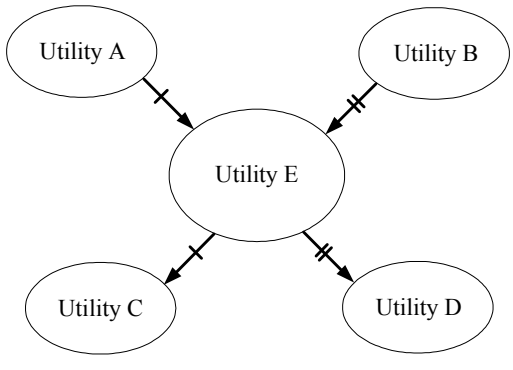


Fig. 1 Multiple wheeling transactions topology (utility E is the intermediate utility)

Normally, the intermediate utility operates its system in the way that the wheeling participants (sellers or buyers or both) must accommodate the losses occurring during the transactions. In fact, the intermediate utility would fix their generation output to the amount its generators would be generating if the wheeling transactions do not take place.

In this paper, the authors introduce an approach to optimally allocate transmission system losses among the utilities involving in multiple wheeling transactions. For simplicity of developing the approach, each transaction of MWT is assumed to be consisted of only one seller and one buyer at this time. The proposed approach uses the loss ratio according to the losses the individual transaction causes the transmission system when that transaction is taken place alone. Necessary assumptions for the approach are that the generation output of the intermediate system stays unchanged at its optimal level and the overall losses of multiple wheeling transactions are minimized. The algorithm therefore starts with solving optimal power flow (OPF) for the intermediate (or wheeling) system to obtain the optimal generation output. The wheeling system remains its optimal generation output during the transactions so that the transaction utilities must accommodate their transmission losses.

The proposed algorithm of the optimal loss allocation for multiple wheeling transactions is tested on a modified IEEE 30 bus system. The results from the algorithm are the optimal real power generated from the supplying utilities which includes their share of losses due to the transactions.

II. Problem Formulation

A. Multiple Wheeling Transactions with Optimal Loss Allocation

The additional losses in the intermediate system due to multiple wheeling transactions must be accommodated by the transaction utilities. Since there are more than one transactions occurring simultaneously, the intermediate utility may not be able to tell how much losses each and every transaction causes in order to calculate the wheeling charge. The best way to allocate losses from multiple wheeling transactions is that the intermediate utility must find out the proportion of losses that each transaction causes individually. Then the intermediate utility divides the total losses among the participated utilities by using the proportion in a loss minimization problem when all wheeling transactions occur.

Therefore, the objective function of optimal loss allocation of multiple wheeling transactions is minimizing total real power losses of all wheeling transactions as expressed in equation (1):

$$\text{Min} \quad \sum_{i=1}^{NWT}\left[\sum_{k=1}^{N} P_{SWTk,i} - \sum_{k=1}^{N} P_{RWTk,i}\right] \qquad (1)$$

where

i	is the i^{th} wheeling transaction
NWT	is the total number of wheeling transactions
k	is the bus number
N	is the total number of buses in the system
$P_{SWTk,i}$	is the real power sending from selling bus k of transaction i
$P_{RWTk,i}$	is the real power receiving at buying bus k of transaction i.

The variable to be minimized is $P_{SWTk,i}$ of every transaction, while $P_{RWTk,i}$ is fixed as known parameter from the wheeling contract.

The usual constraints of the above optimization problem are the power flow equations of the wheeling utility as shown in equations (2) and (3), for any bus k, when $k = 1,...,N$.

$$P_k - P_{Gk}^{*} + P_{Dk} - \sum_{i=1}^{NWT} P_{SWTk,i} + \sum_{i=1}^{NWT} P_{RWTk,i} = 0 \qquad (2)$$

$$Q_k - Q_{Gk} + Q_{Dk} - \sum_{i=1}^{NWT} Q_{SWTk,i} + \sum_{i=1}^{NWT} Q_{RWTk,i} = 0 \qquad (3)$$

where

P_k	is real power injected at bus k
P_{Gk}^{*}	is optimal real power generated at bus k obtained from optimal power flow of base case system
P_{Dk}	is real power load at bus k
Q_k	is reactive power injected at bus k
Q_{Gk}	is reactive power generated at bus k
Q_{Dk}	is reactive power load at bus k
$Q_{SWTk,i}$	is the reactive power sending from selling bus k of transaction i
$Q_{RWTk,i}$	is the reactive power receiving at buying bus k of transaction i

The additional constraint required in optimal loss allocation for MWT is that each transaction must accommodate its own share of losses from the wheeling. The individual loss is calculated from the total losses in equation (1) by using the ratio of real power loss of transaction i to the overall losses as expressed in equation (4). The details of the loss ratio, ρ_{Li}, will be explained in the next section. Note that each transaction of MWT consists of only one seller and one buyer.

$$P_{SWTi} - \rho_{Li} \times \sum_{i=1}^{NWT}\left[\sum_{k=1}^{N} P_{SWTk,i} - \sum_{k=1}^{N} P_{RWTk,i}\right] - P_{RWTi} = 0 \qquad (4)$$

where

ρ_{Li} is a ratio of real power loss of transaction i to overall losses of multiple wheeling transactions.

Inequality constraints for this optimization problem are transmission line flow limits, capacities of real and reactive power generation at sending buses, voltage limits at any bus in the intermediate system and limits of transformer tap ratio, as expressed in equations (5)-(9), respectively.

$$S_{kn} \le S_{kn,\max} \tag{5}$$

where

S_{kn} is apparent power flowing in transmission line connecting between buses k and n.

$$P_{RWTk,i} \le P_{SWTk,i} \le P_{SWTk,i\max} \tag{6}$$

$$Q_{SWTk,i\min} \le Q_{SWTk,i} \le Q_{SWTk,i\max} \tag{7}$$

$$V_{k,\min} \le V_k \le V_{k,\max} \tag{8}$$

$$T_{k,\min} \le T_k \le T_{k,\max} \tag{9}$$

Note that the lower limit of $P_{SWTk,i}$ in equation (6) is the amount of real power the transaction i has committed to deliver to its buyer.

B. Ratio of Real Power Losses among Multiple Wheeling Transactions

For an individual wheeling transaction, i, the ratio of real power loss of the transaction compared to total real power losses of all multiple wheeling transactions can be obtained from:

$$\rho_{Li} = \frac{\sum_{k=1}^{N} P_{SWTk}^{(i)} - \sum_{k=1}^{N} P_{RWTk}^{(i)}}{\sum_{i=1}^{NWT}\left[\sum_{k=1}^{N} P_{SWTk}^{(i)} - \sum_{k=1}^{N} P_{RWTk}^{(i)}\right]} \tag{10}$$

where

$P_{SWTk}^{(i)}$ is optimal real power sending from selling bus k when considering only wheeling transaction i

$P_{RWTk}^{(i)}$ is real power receiving at buying bus k when considering only wheeling transaction i

Equation (10) states that in order to obtain the loss ratio, real power losses of each transaction must be computed. Thus, by fixing the real and reactive power at the receiving end of the wheeling transactions, the optimal real power sending into wheeling utility of each transaction must be computed individually. The optimization problem of any single wheeling transaction can be written as:

$$\min \quad \sum_{k=1}^{N} P_{SWTk}^{(i)} - \sum_{k=1}^{N} P_{RWTk}^{(i)} \tag{11-a}$$

subject to

$$P_k - P_{Gk}^{*} + P_{Dk} - P_{SWTk}^{(i)} + P_{RWTk}^{(i)} = 0 \tag{11-b}$$

$$Q_k - Q_{Gk} + Q_{Dk} - Q_{SWTk}^{(i)} + Q_{RWTk}^{(i)} = 0 \tag{11-c}$$

$$S_{kn} \le S_{kn,\max} \tag{11-d}$$

$$P_{RWTk}^{(i)} \le P_{SWTk}^{(i)} \le P_{SWTk,\max}^{(i)} \tag{11-e}$$

$$Q_{SWTk,\min}^{(i)} \le Q_{SWTk}^{(i)} \le Q_{SWTk,\max}^{(i)} \tag{11-f}$$

$$V_{k,\min} \le V_k \le V_{k,\max} \tag{11-g}$$

$$T_{k,\min} \le T_k \le T_{k,\max} \tag{11-h}$$

where

$Q_{SWTk}^{(i)}$ is optimal reactive power sending from selling bus k when considering only wheeling transaction i

$Q_{RWTk}^{(i)}$ is reactive power receiving at buying bus k when considering only wheeling transaction i

III. Solution Methodology

By assuming that buying and selling utilities must accommodate their losses due to wheeling transactions, the intermediate system maintains generation output at the optimal value when there is no wheeling (base case system). Therefore, optimal loss allocation for multiple wheeling transactions starts with calculating optimal power flow (OPF) minimizing generation cost for base case in order to obtain optimal real power generation for intermediate system.

Wheeling transactions must be handled individually in order to obtain the proportional loss ratio expressed in equation (10). For each wheeling transaction, loss minimization problem is computed for the modified system with addition wheeling data by using the set of equation (11). As for the modified system, wheeling real power load of transaction i, $P_{RWTk}^{(i)}$, is added to the intermediate network at the receiving bus connected to buying utility, while the sending bus connected to selling utility is treated as generator bus for the intermediate network. After the optimal $P_{SWTk}^{(i)}$ has been computed for every wheeling transaction, the loss ratio, ρ_{Li}, can be calculated.

The main optimization problem which is minimizing total losses of multiple wheeling transactions can now be computed by using equations (1)-(9) with fixed generation output P_{Gk}^{*} from base case OPF and loss ratio, ρ_{Li}. The optimization yields the optimal real power sending from selling utilities including minimum real power losses due to every wheeling transaction. The flow chart of multiple wheeling transactions with optimal loss allocation is shown in Fig. 2

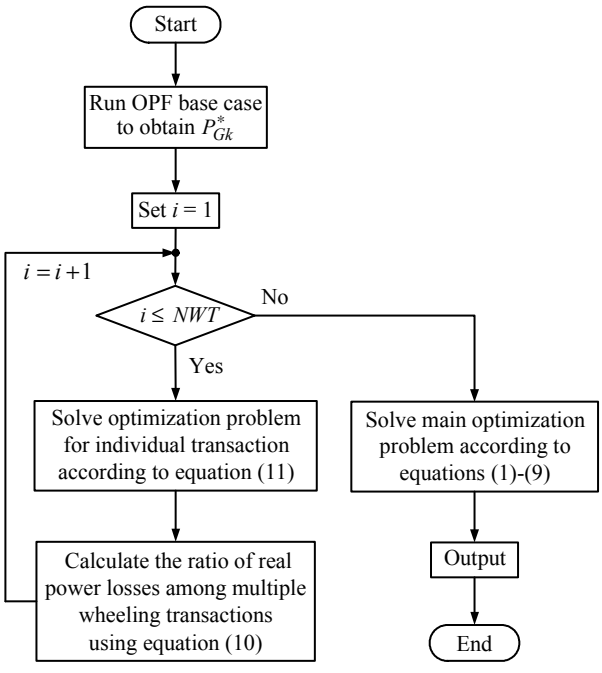


Fig. 2 Flowchart of multiple wheeling transactions with optimal loss allocation

IV. Case Studies

The proposed algorithm is tested on a modified IEEE 30 bus system as shown in Fig. 3. One of the synchronous condensers in the original system at bus 5 is modified to be a generator. Transformer tap ratios are all bounded between 0.9 and 1.1. Voltage magnitudes for all buses are bounded between 0.94 and 1.06. Transmission line limits are modified from the original IEEE 30 bus as illustrated in Table 1.

The 30-bus system is assumed to be the intermediate utility with connections to other four utilities at some specified buses as shown. Multiple wheeling transactions occurring simultaneously are:

- Transaction 1, utility A at bus 3 supplying power to utility C at bus 15,
- Transaction 2, utility B at bus 7 supplying power to utility D at bus 25.

For now, the transmission lines connected between the 30-bus system and other utilities are assumed to be large enough to carry desired power in and out of the system without any losses in them. In other words, the studies assume that power supplying from utilities A and B are injected directly into buses 3 and 7, respectively. Similarly, power drawing from the 30-bus system by utilities C and D are treated as additional load at buses 15 and 25, respectively. The maximum limit of wheeling real power is 40 MW for each transaction, while the wheeling reactive power is bounded from -40 MVar to 50 MVar.

In this paper, there are four case studies where each case has two transactions as described above. The differences among the four cases are the amount of wheeling power as shown in Table 2.

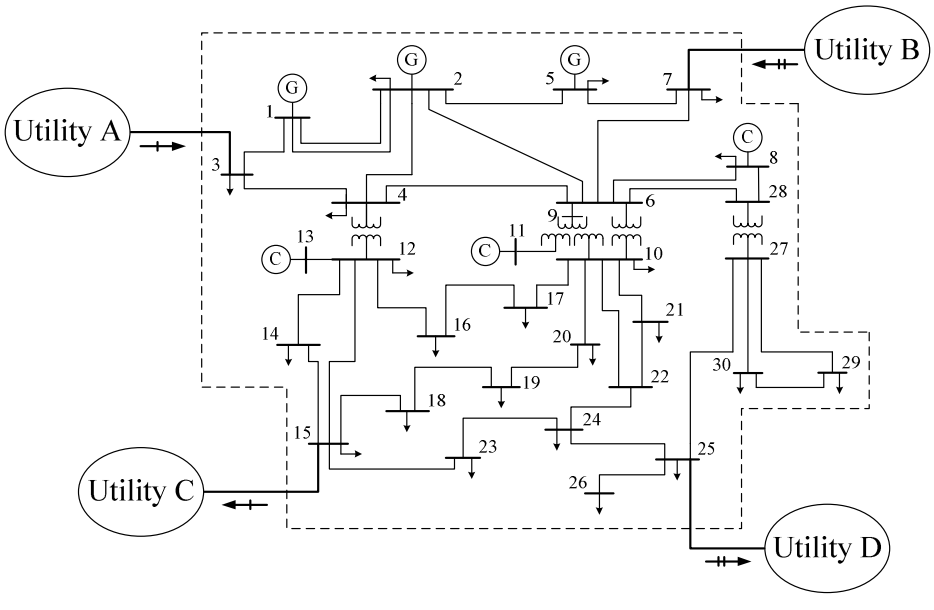


Fig. 3 Multiple wheeling transactions with optimal loss allocation of modified IEEE 30 bus test system

Table 1 Branch rating data of intermediate system

Branch No.	Bus No.	Rating MVA	Branch No.	Bus No.	Rating MVA
1	1 - 2	130	22	15 - 18	32
2	1 - 3	100	23	18 - 19	32
3	2 - 4	72	24	19 - 20	32
4	3 - 4	100	25	10 - 20	32
5	2 - 5	130	26	10 - 17	32
6	2 - 6	100	27	10 - 21	32
7	4 - 6	100	28	10 - 22	32
8	5 - 7	130	29	21 - 22	32
9	6 - 7	130	30	15 - 23	32
10	6 - 8	65	31	22 - 24	32
11	6 - 9	65	32	23 - 24	32
12	6 - 10	32	33	24 - 25	32
13	9 - 11	65	34	25 - 26	32
14	9 - 10	65	35	25 - 27	32
15	4 - 12	65	36	28 - 27	65
16	12 - 13	130	37	27 - 29	32
17	12 - 14	65	38	27 - 30	32
18	12 - 15	65	39	29 - 30	32
19	12 - 16	65	40	8 - 28	65
20	14 - 15	32	41	6 - 28	72
21	16 - 17	65			

Table 2 Data of multiple wheeling transactions

Transaction No.	1		2	
Case No.	P_{RWT} (MW)	Q_{RWT} (MVar)	P_{RWT} (MW)	Q_{RWT} (MVar)
I	20	12.395	20	12.395
II	20	12.395	30	18.592
III	30	18.592	20	12.395
IV	30	18.592	30	18.592

V. Test Results

In order for the intermediate utility to maintain its generation output, the optimal generation output of the 30-bus test system is obtained from OPF when there is no wheeling transaction. Then each wheeling transaction is added to the intermediate system in order to find losses associated with the wheeling. The minimum losses occurring during the wheeling are accommodated by supplying utilities which are included in generation power injected into the intermediate system ($P_{SWT}^{(i)}$) as shown in Table 3. The superscript (i) means that the generation output is the result from having only transaction (i) in the system. Then the proportional loss ratio of transaction i , (ρ_{Li}), is computed for each transaction and every case as shown in Table 3. For every case, the sum of loss ratio from individual transaction is equal to one. The loss ratio tells us how much losses an individual transaction causes as a proportion to the summation of losses from other transactions. Although the overall losses when every wheeling transaction takes place simultaneously may not equal the sum of individual losses, the loss ratio is considered an acceptable proportion to allocate transmission losses among multiple wheeling transactions.

Table 3 Results of loss ratio

Transaction No.	1		2	
Case No.	$P_{SWT}^{(i)}$ (MW)	ρ_{Li}	$P_{SWT}^{(i)}$ (MW)	ρ_{Li}
I	21.190	0.5192	21.102	0.4808
II	21.190	0.3332	32.381	0.6668
III	32.059	0.6514	21.102	0.3486
IV	32.059	0.4637	32.381	0.5363

Table 4 Comparison of multiple wheeling transactions with and without optimal loss allocation

MWT	with loss ratio constraint		without loss ratio constraint	
Trans. No.	1	2	1	2
Case No.	Active power at sending bus (P_{SWT}) MW			
I	21.212	21.122	20	22.309
II	21.212	32.425	20	33.615
III	32.118	21.134	30	23.200
IV	32.199	32.543	30	34.691

When all wheeling transactions are participated, the minimum losses corresponding to the transactions can be obtained while the intermediate system maintains its generation output so that the participated utilities are responsible for their losses. If the minimum losses were computed without loss ratio (equation (4) is neglected), the optimization problem would place all losses at a particular transaction with cheaper loss cost. In all cases, if transaction 1 were not forced to compensate its wheeling losses, utility A would supply only the amount of power for the transactions, while utility B would be unfairly responsible for all associated losses. However, the overall losses in this scheme are apparently the least amount of losses occurring to the intermediate network.

If the minimum losses were computed with constraint on loss ratio (equation (4) is implemented), every transaction must accommodate its own loss at a proportional ratio computed earlier. The results in Table 4 show that the overall losses are split proportionally among supplying utilities. The overall losses in this scheme are however close to those of the former scheme with the smallest difference of 0.022 MW in case II and the largest difference of 0.052 MW in case III.

VI. Conclusions

This paper presented an algorithm for optimal loss allocation of multiple wheeling transactions by introducing loss ratio constraint. The algorithm applied optimal power flow to the intermediate system to obtain optimal generation output. The intermediate system maintained its generation output during the wheeling

therefore supplying utilities must compensate losses due to wheeling transactions. The wheeling losses were minimized and allocated to every supplying utility by a proportional ratio calculated for each and every transaction. The study also compared the optimal results of loss allocation when the loss ratio was considered and when the loss ratio was not considered. The algorithm was tested on a modified IEEE 30 bus system with two wheeling transactions. Each transaction consisted of only one buyer and one seller. The results of the algorithm were the optimal real power generated from the supplying utilities which includes wheeling losses. The optimal results showed that the algorithm could find minimum wheeling losses and divide them among supplying utilities when loss ratio was implemented. However, when loss ratio was not considered, the optimization problem would place all losses at only one particular transaction.

This research could be extended by considering one buying utility purchasing power from more than one supplying utility and vice versa, to see how losses are allocated among all participating utilities. Moreover, wheeling costs associated with wheeling losses have not been considered yet in this paper and would also be left for future work.

References

[1] Y.-M. Park, J.-B. Park, J.-U. Lim, and J.-R. Won, "An analytical approach for transaction costs allocation in transmission system," IEEE Transactions on Power Systems, vol. 13, pp. 1407–1412, November 1998.

[2] T. Nakashima, T. Niimura, K. Okada, R. Yokoyama and N. Okada, "Multiple-Impact Assessment of Wheeling and Independent Power Producers In a De-Regulated Power System," in Proc. IEEE Canadian Conference on Electrical and Computer Engineering, vol. 1, pp. 89–92, 24-28 May 1998.

[3] D. Kirschen and G. Strabc, "Tracing active and reactive power between generators and loads using real and imaginary currents," IEEE Transactions on Power Systems, vol. 14, pp. 1312–1319, February 1998.

[4] J. Bialek, "Allocation of transmission supplementary charge to real and reactive loads," IEEE Transactions on Power Systems, vol. 13, pp. 749–754, August 1998.

[5] J. W. Bialek, "Topological generation and load distribution factors for supplement charge allocation in transmission open access," IEEE Transactions on Power Systems, vol. 12, pp. 1185–1193, August 1997.

[6] C. N. Macqueen and M. R. Irving, "An algorithm for the allocation of distribution system demand and energy losses," IEEE Transactions on Power Systems, vol. 11, pp. 338–343, February 1996.

[7] A. G. Exposito, J. M. Riquelme Santos, T. G. Garcia and E. A. Ruiz Velasco, "Fair Allocation of Transmission Power Losses," IEEE Transactions on Power Systems, vol. 15, no. 1, pp. 184–188, February 2000.

[8] S. Tao and G. Gross, "Transmission Loss Compensation in Multiple Transaction Networks," IEEE Transactions on Power Systems, vol. 15, no. 3, pp. 909–915, August 2000.

[9] F. D. Galiana and M. Phelan, "Allocation of Transmission Losses to Bilateral Contracts in a Competitive Environment," IEEE Transactions on Power Systems, vol. 15, no. 1, pp. 143–150, February 2000.

[10] E. D. Tuglie and F. Torelli, "Nondiscriminatory System Losses Dispatching Policy in a Bilateral Transaction-Based Market," IEEE Transactions on Power Systems, vol. 17, no. 4, pp. 992–1000, November 2002.

[11] E. A. Belati, V. A. de Sousa, A. M. de Souza and G. R. M. da Costa, "Transmission Loss Allocation by Sensitivity Approach," in Proc. IEEE Russia Power Tech, pp. 1–7, 27-30 June 2005.

[12] "U.S. Congress, Office of Technology Assessments, Electric power wheeling and dealing: Technological consideration for increasing competition," U.S. Government Printing Office, Washington, D.C., Rep. OTA-E-409, May 1989.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Neuro Fuzzy Soft Starter for Grid Integration with Pitch Regulated Wind Turbine System

L.Rajaji[+] and *Dr.C.Kumar
+Research Scholar, Sathyabama University, Chennai, India
*Principal, S.K.P. Engineering College, Tiruvannamalai, India

Abstract

Wind turbines are playing a vital role in power generation among the renewable energy sources. One of the effective ways of power regulating systems is pitch regulated wind turbine which is used to interface the grid and generator smoothly. In this paper, a novel approach of pitch regulated wind turbine using intelligent soft starter based induction generator is presented. Neuro Fuzzy approach has been used as intelligent tool in soft starter to estimate thyristors firing angle accurately so as to integrate the generator to the grid smoothly. Various wind turbine models such as 600kW and 1000kW are taken for simulation and simulation results have been presented to prove the proposed methodology. Low Voltage Ride Through (LVRT) capability of proposed methodology in 1 MW wind turbine model is also presented in this paper to prove the reliability of proposed method.

Key Words: *Induction generator, Pitch regulation, Wind turbine, Soft starter, ANFIS and LVRT*

1. INTRODUCTION

Mankind has used the wind as a source of energy for thousands of years. It was one of the most utilized sources of energy together with hydro power during the seventeenth and eighteenth centuries [1]. By the end of the nineteenth century the experiments were carried out on the use of windmills for generating electricity. Thereafter, there was a long period of a low interest in the use of wind power.

Governing support and public interest in renewable energy have caused a massive increase in wind power utilization and improvement of wind turbine technology is a natural consequence. On shore wind power potential, in India, has been assessed at 45000 MW. Exhaustive wind resource assessment has been carried out in more than 555 stations spread over 20 States in the country. Around 221 Wind Monitoring stations have indicated wind power density of more than 200 W/m^2 at 50 m above ground level.

The massive increase in installed wind power and the enormous plans for future use of wind power raise another concern from the power system side. Wind power cannot be treated as an unimportant power source any more, since it represents a higher fraction of the total power system installed capacity. The impact of wind turbines on power system stability is often mentioned and, consequently, there is great interest in modeling and predicting wind turbine response to the transient behavior of the power system.

Modeling wind turbines for predicting of their power quality impact is reported in many literatures. Engineers tend to simplify the aerodynamic and mechanic parts of the system and usually stress generator description. Moreover, they often overlook generator performance details. Some reported models seem to be over-parameterized, which obstructs their implementation because the parameters for the detailed description are not generally available.

Simplified aerodynamic modeling of wind turbines has been presented in [3], [4]. The main idea in these articles is to adjust wind speed data at one point (hub level) by the use of various filters in order to represent the interaction of turbine blades with wind speed distribution over the rotor swept area. The resulting wind data are then applied to the static power curve, $Cp(\lambda)$, in order to determine the driving torque. In contrast to this, an advanced approach to aerodynamic modeling that uses a professional software package has been presented in [5].

The impact of wind turbines on power system stability is also dealt with in the literature. The work presented in [6] describes a model of a grid-connected wind generator designed for predicting both, steady-state operation impact as well as the response to grid faults. However, verification of the simulation results against field measurements is lacking. An evaluation of the fault response of fixed-speed wind turbines and of variable-speed wind turbines equipped with doubly-fed induction generators is analyzed in [8].

2. SOFT STARTERS

The soft-starter is a power electronic converter that is not only specific for wind turbines, and is also being introduced more and more frequently in industrial plants where it is necessary to operate with induction motors controlling the start currents in a more efficient way than the traditional methods. A soft-starter device is integrated by 6 thyristors, two per phase, in a anti-parallel configuration as can be seen in Figure 1. A snubber RC network is usually included in order to limit the rate of change of the voltage, *dv/dt*, across the thyristors. Figure 2 shows the soft-starter control circuit and the way to obtain the beginning of the excitation of the gates (firing angle).

3. SOFT STARTER BASED INDUCTION GENERATOR IN PITCH REGULATED WIND TURBINES

In a wind turbine, the soft-starter has been introduced to fixed speed ones to reduce in-rush currents and voltage dropouts. Hammons and Lai [8] analyze

phenomena which affect voltage dip and inrush current due to the direct connection of induction generators running close to synchronous speed to electrical distribution in low head hydro electric schemes. This is a different case to that studied in the starting of induction motors, since now slip varies within a narrow interval around zero, but comprising motor and generator operation.

This qualitative change in the operation mode gives place to important numerical changes in the relationship between the firing angle and the voltage at the induction machine terminals. Even if there is not a shift between motor and generator operation modes, this relationship is also affected by the power factor of the induction machine which in turn depends on the voltage derivative.

Figure 3 shows a simplified soft-starter performance for a wind turbine generator. Basically, its function is to feed the induction machine with a variable voltage, whose evolution pattern is specified according to some constraint. This constraint has traditionally been to limit the start-up over current. In the present work, the soft-starter will be controlled in order to limit the voltage dropout.

In fact thyristor gates are not triggered by means of continuous pulses nor isolated impulses, but by supplying pulse trains beginning at the desired firing angle. The length of these trains can be shortened [6][7], to avoid unnecessary triggering.

4. THYRISTOR TRIGGERING

In phase A voltage, its zero crossing is the reference taken for the beginning of the pulse train that will excite the gate of the forward thyristor in phase A. It is more suitable using angles instead of time instants to refer to the point at which pulse trains begin. In this sense, pulse trains are repeated in the remaining thyristors every 60^o. Therefore the firing angle refers to the angle between the zero crossing of any voltage phase and the beginning of the train pulses exciting the corresponding forward thyristor. It determines the voltage supplied to the connected device. The triggering pulse sequence is depicted in Fig.6 Arrows indicate the beginning of the overlapped train of pulses at the corresponding thyristors. Thus, the first vertical arrow represents the point at which forward thyristor in phase A and reverse thyristor in phase B are both triggered. Firing angle or delay angle 'α' is the distance between this point and the rising zero-crossing of V_A. In order that the thyristors conduct, this first arrow must be at the left of the intersection of curves V_A and V_B. In Fig. 4 the separation angle between forward thyristor triggering pulses is shown to be 120^o. Triggering signals for forward thyristor gates and the corresponding reverse one are separated 180^o.

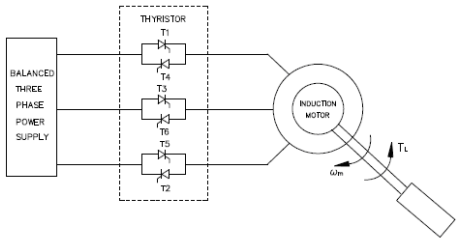


Fig. 1 Basic Soft starter Circuit

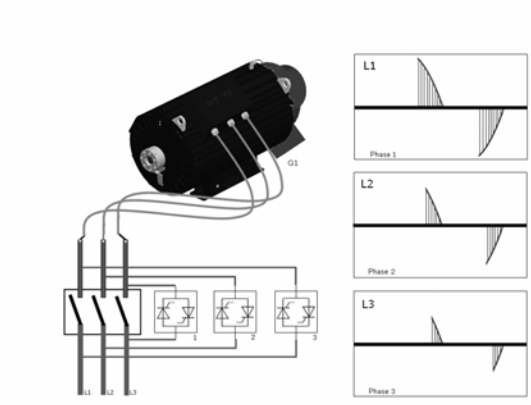


Fig. 2 Induction Generator with soft starter and associated waveforms

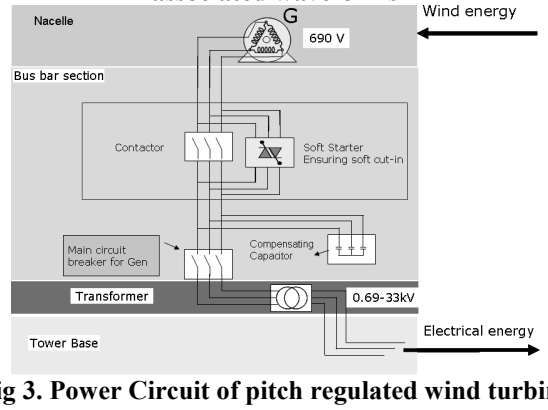


Fig 3. Power Circuit of pitch regulated wind turbine with soft starter

5. ANFIS AS INTELLIGENT METHODOLOGY

The ANFIS neuro-fuzzy system [9] [10] [11] [12] has been used to implement the proposed model. First, it uses the training data set to build the fuzzy system in which, membership functions are adjusted using the backpropagation algorithm, allowing that the system learns with the data that it is modeling. Figure 4 shows the network structure of the ANFIS that maps the inputs by the membership functions and their associated parameters, and so through the output membership functions and corresponding associated parameters. These will be the synaptic weights and bias, and are associated to the membership functions that are adjusted during the learning process. The computational work to obtain the parameters and their adjustments is helped by the gradient descendent technique.

In this paper, the ANFIS system is used at the CASPOC/MATLAB environment. Its operation can be resumed in two steps:

1) The set of membership functions has to be chosen: their number and corresponding shape.

2) The input-output training data is used by the ANFIS. It starts making a clustering study of the data to obtain a concise and significant representation of the system's behavior.

It is important to note that the system has a good modeling if the training set is enough representative, i.e., it has a good data distribution to make possible to interpolated all necessary values to the system's operation. The clustering technique used was the fuzzy c-means. After setting the number of clusters that are estimated to compose the data, the cluster's centres are searched in an iterative way based on minimizing an objective function. This represents the distance between a data value to the cluster's centre. As we do not know how much clusters exist, or the number of rules composing the neuro-fuzzy compensator, we used the technique of subtractive clustering to estimate the number of clusters.

The ANFIS model which is used for the calculation of firing angle of the appropriate thyristor as a function of the motor speed (ω_m) and torque (T_e), has two input variables (T_e and ω_m) and one output variable (α). Since the angle α is a nonlinear function of the motor speed and torque, then the tansigmoidal function is the most appropriate to model the ANFIS as given in equation (1).

$$f(n) = 2/(1+e^{-2n}) - 1. \quad (1)$$

The layers shown in Fig. 4, are defined as follows:+

- Layer 1: Every node in this layer contains membership functions.
- Layer 2: This layer chooses the minimum value of two input weights
- Layer 3: Every node of these layers calculates the weight, which is normalized.
- Layer 4: This layer includes linear functions, which are functions of the input signals.
- Layer 5: This layer sums all the incoming signals.

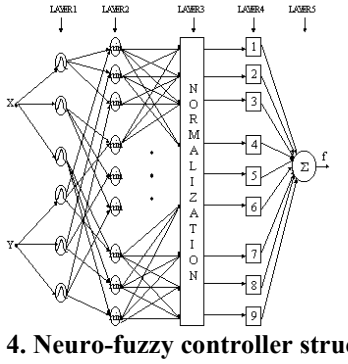


Figure 4. Neuro-fuzzy controller structure

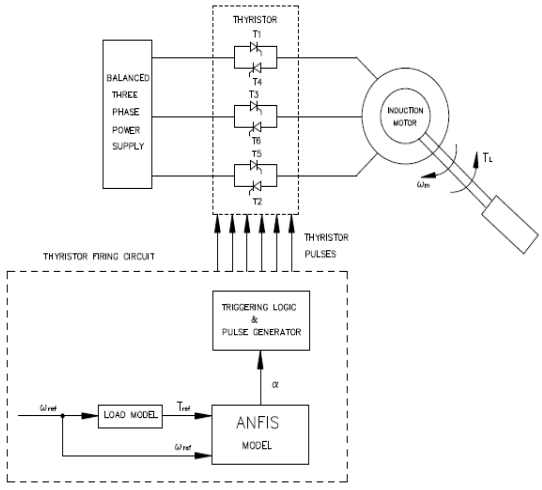


Fig 5. Schematic of Intelligent soft starter

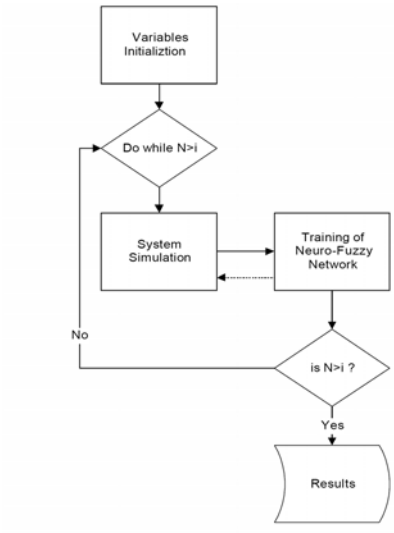


Fig 6. Flowchart of ANFIS Algorithm

6. LOW VOLTAGE RIDE THROUGH (LVRT)

In order to supply power to grid continuously and reliably, wind electric generators must confirm the standards to integrate the grid that defines requirements power imposed on power suppliers and large power consumers. In particular, LVRT requirement typically requires that a wind turbine unit remains connected and synchronized to the grid when the voltage across the terminals at the generator terminals falls below the value of prescribed value. Battery buffered pitch system has been incorporated in those turbines to achieve the offline trip. In higher rated turbines, three independent battery packs are used to turn the blades from operating position to parking position and in lower and medium rated turbines single battery pack has been used. In this system, movement of blades from operating position to park position occurs due to voltage and frequency fluctuating error. The major drawbacks of this system are that LVRT requirement could not be achieved and wind turbine system is tripped in

offline. These drawbacks can be overcome by smooth variation of firing angle of soft starter which connects the wind turbine and grid. In this paper, ANFIS is applied to achieve effective LVRT by estimating the generator voltage accurately. Simulation has been carried out to prove the LVRT using ANFIS with 1MW wind turbine model. LVRT performance curves are shown in Figures 16 and 17.

7. RESULTS AND DISCUSSIONS

Simulation of three different models of wind turbines using ANFIS soft starter strategy and low voltage ride through of 1MW wind turbine has been done using MATLAB/SIMULINK and CASPOC software packages. The detailed results and well explanation have been presented in this paper to prove the proposed methodologies. For simulation, 600kW and 1000kW wind turbines have been considered and all these three turbines are fixed speed and directly connected to the power grid. The parameters of these three turbines are given in Table I.

Table I

Rated Power	**600 kW**	**1000 kW**
Turbine rotor diameter	47 m	60 m
Type of Generator	Cage induction generator	Cage induction generator
Generator Rated Voltage / Current	690 V / 558 A	690 V / 920 A
Rated Frequency	50 Hz	50 Hz
No. of poles	4	4
Rated Speed	1531 rpm	1527 rpm

The first phase of simulation was carried out only for soft starter fed induction motor drive with and without ANFIS algorithm. By using CASPOC software package, simulation was carried out and the results were drawn in Excel sheet. For clear understanding purpose, pulses are shown in different colours for which Excel sheet has been used specially. Figure 7 shows the gate triggering pulses at thyristor gates in which GA, GB, GC are positive group thyristor gate pulses and GA', GB', GC' are negative group thyristor gate pulses. By using this triggering sequence, the line voltage across the terminal and line current waveforms are shown in Figure 8. Conventional soft starter can be accomplished by ANFIS tool to overcome the current interruptions as shown in Figures 9 and 10. In Figure 9, current interruptions appeared at the output terminals which will flow through the motor windings. These interruptions will be treated as noise and these will produce more heat to overcome this drawback motor windings are to be protected using high class insulation and it will lead cost. To avoid this, ANFIS is the best tool to avoid current interruptions. Figure 10 shows the motor voltage and motor current from ANFIS based soft starter without current interruptions. If the case in which asymmetrical gate triggering pulses are applied to the thyristors, the motor current will be discontinuous though ANFIS is used to measure the triggering pulses as shown in Figure 11. Figure 12-19 explain the performance of different wind turbine models using proposed soft starter. To achieve this, CASPOC and MATLAB/SIMULINK packages are integrated and separate m-files have been used. From Figures 12-15, it is observed that four different characteristics such as wind speed, pitch angle, generator speed and generated power have been taken to analyze the performance. As mentioned earlier, performance of the three different rated wind turbines have been analyzed using conventional and proposed soft starters. Although pitch angle can be varied in pitch regulated wind turbines between -5 degree and +88 degree, in this paper, least values of pitch angles have only been taken for the simulation. On the other hand, wind speed variation has also shown higher values. From Figures 12 & 13 and Figures 14 & 15 it is clearly shown that the power regulation is very smooth using ANFIS starter when compare to conventional starter. Figure 18 shows the power curve of 600kW wind turbine using proposed soft starter methodology and Figure 19 shows the power curve of 1000 kW wind turbine with the proposed methodology.

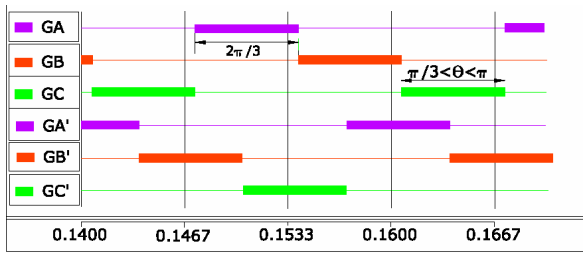


Fig 7. Pulses at Thyristor gates

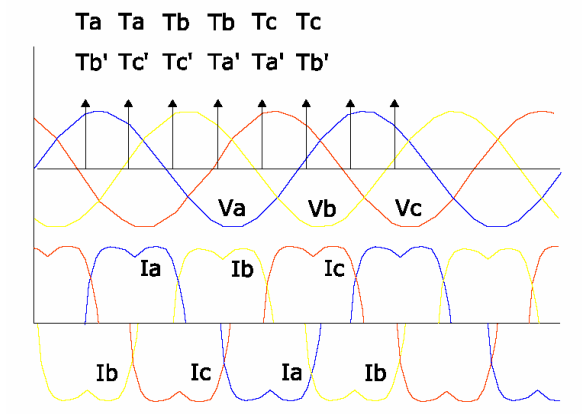


Fig 8. Gate Triggering Sequence and Line Currents

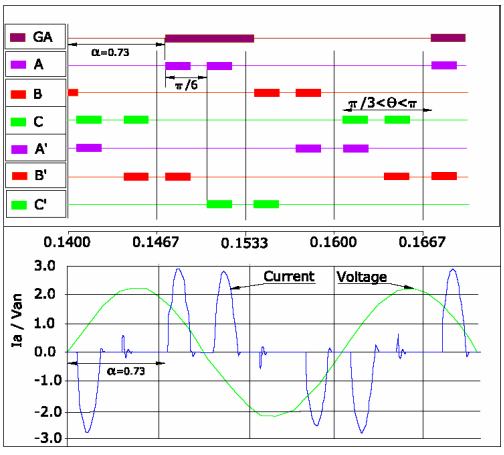


Fig 9: Operation with Current interruptions

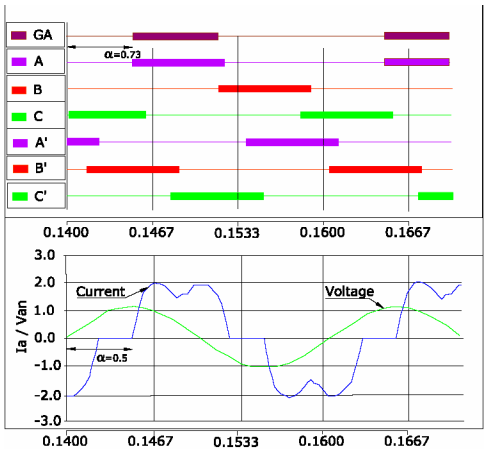


Fig 10: Operation without Current Interruptions

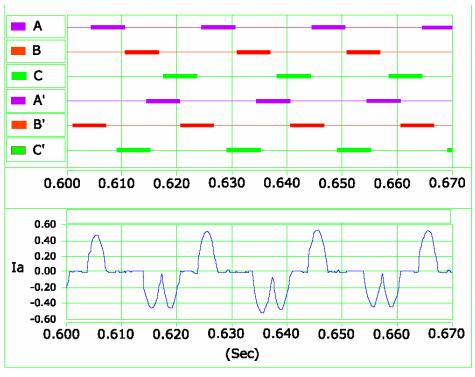


Fig 11: Asymmetrical pulse sequence at the thyristors gates and phase current

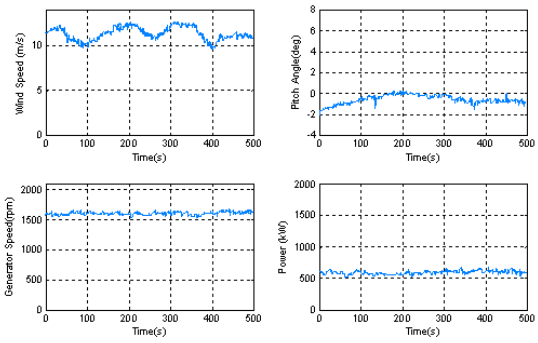


Fig 12. Simulation Results for 600kW Wind Turbine with conventional soft starter strategy

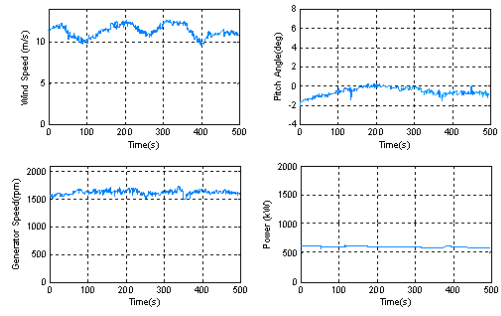


Fig 13. Simulation Results for 600kW Wind Turbine with proposed soft starter strategy

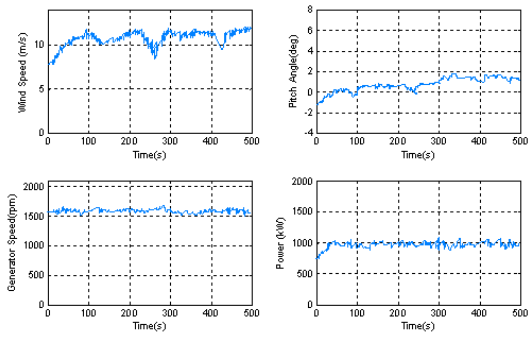


Fig 14. Simulation Results for 1000 kW Wind Turbine with conventional soft starter strategy

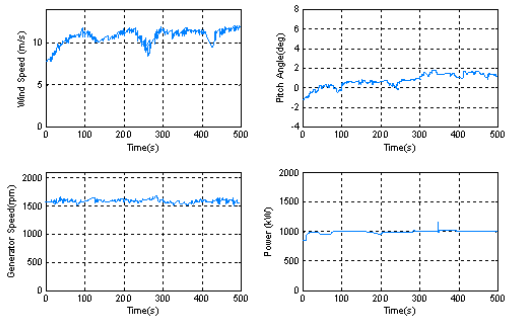


Fig 15. Simulation Results for 1000 kW Wind Turbine with proposed soft starter strategy

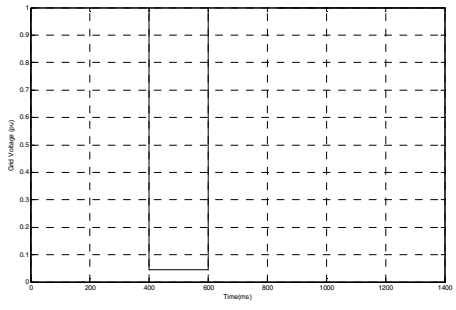


Fig 16. Low Voltage Ride Through – 5% Generator Voltage

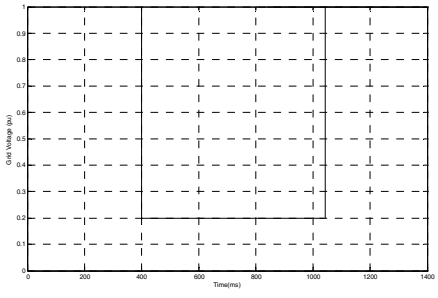


Fig 17. Low Voltage Ride Through – 20% Generator Voltage

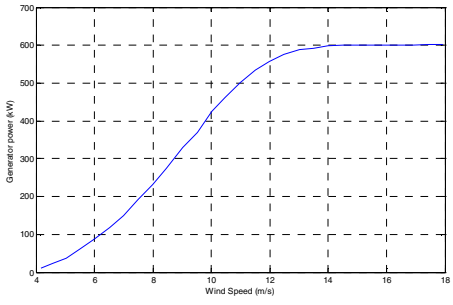


Fig 18. Power curve of 600kW pitch regulated wind turbine using proposed methodology

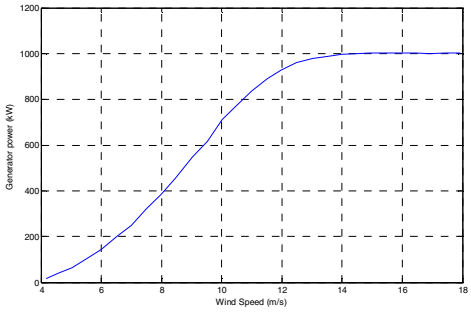


Fig 19. Power curve of 1000kW pitch regulated wind turbine using proposed methodology

8. CONCLUSION

In this paper, an efficient and intelligent soft starter based pitch regulated wind turbine systems have been presented. Adaptive neuro fuzzy inference system based soft starter ensures smooth integration of wind turbine and grid. Induction generator which is mounted in wind turbine nacelle is magnetized by grid power through this soft starter only. ANFIS is an algorithm which acts as subroutine with main wind turbine controller program. ANFIS based soft starter is also used to achieve low voltage ride through during grid a fault that is also presented in this paper. Therefore, the main contributions of this paper are summarized as (i) Simulation of novel wind turbine starter using ANFIS (ii) Modeling of 600kW, 1000 kW wind turbines with ANFIS (iii) Simulation of Low Voltage Ride Through capability of proposed soft starter in 1 MW wind turbine.

References

[1] C. Abbey and G. Joos, "Effect of Low Voltage Ride Through (LVRT) Characteristic on Voltage Stability,"

[2] Wilkie, J., Leithead, W.E., Anderson, C., Modeling of Wind turbines by Simple Models, Wind Engineering, Vol. 13, No. 4, 1990

[3] Bossanyi, E.A., Gardner, P., Craig, L., Saad-Saoud, Z., Jenkins, N., Miller, J., Design Tool for prediction of flicker, European Wind Energy Conference, Dublin Castle, Ireland, 1997

[4] Akhmatov, V., Knudsen, H., Nielsen, A.H., Advanced simulation of windmills in the electrical power supply, International Journal of Electrical Power and Energy Systems, July 2002, Vol.22, No. 6, p. 421 - 434

[5] Leithead, W.E., Rogers, M.C.M., Drive-train Characteristics of Constant Speed HAWT's: Part II - Simple Characterization of Dynamics, Wind Engineering, Vol. 20, No. 3, 1996

[6] Bao, N.Sh., Chen, Q.X., Jiang, T., Modelling and Identification of a Wind Turbine System, Wind Engineering, Vol. 20, No. 4, 1996

[7] Jenkins, N., Saad-Saoud, Z., A simpli_ed model for large wind turbines, European Union Wind Energy Conference, G oteborg, Sweden, 1996.

[8] Usaola, J., Vilar, C., Amaris, P., Ledesma, P., Rodriguez, J.L., Characterization of WECS through power spectra for power quality studies, European Wind Energy Conference (EWEC '99), Nice, France, Mars 1-5, 1999.

[9] Jang, J.R., "ANFIS: Adaptive-networks based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics," vol. 23, pp. 191-197, 1993.

[10] Purwanto, E.; Arifin, S.; Bian-Sioe So, **"**Application of adaptive neuro fuzzy inference system on the development of the observer for speed sensor less induction motor," IEEE TENCON, vol.1.,pp. 409-414, 2001.

[11] Ali Akaayol, "Application of adaptive neuro-fuzzy controller for SRM," Advances in Engineering software, vol. 35, No.3-4,pp. 129-137.

[12] L.Rajaji and Dr.C.Kumar, "Adaptive Neuro Fuzzy based soft starting of voltage-controlled induction motor drive" IEEE Southeastcon, April 2008, pp-448-453.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Educational Purpose GUI for Evolutionary Computation in Economic Dispatch Problem

A. B. M. Nasiruzzaman and M. G. Rabbani

Department of EEE, Rajshahi University of Engineering & Technology
Kazla, Rajshshi-6204, Bangladesh
E-mail: nasir_zaman_eee@yahoo.com

Abstract – In this paper we propose an evolutionary computation (genetic algorithm) based approach to economic dispatch problem. In case of thermal power plants fuel cost is an important element. Fuel price tend to rise from time to time. Therefore effort should be given on minimizing the fuel cost. In the proposed approach elitism operation in genetic algorithm is exploited very well to improve the quality of solution, and it is compared with the classical optimization technique. A Graphical User Interface (GUI) is developed for educational purposes.
Keywords: Economic dispatch, Evolutionary computation, Genetic Algorithm, GUI

I. Introduction

The term economic dispatch has been given to the problem of minimizing the cost of fuel at thermal plants, assuming that hydro generation has been previously defined and that the configuration of the network is known. It is also known which thermal units are on line. The constraints on this problem, as found in the literature, vary widely, generally trading off complexity for solution speed. Many power systems today are operated under economic dispatch with calculations made on-line every few minutes. Under normal circumstances, control signals are sent to generating stations for generating units to adjust their power output in accordance with optimization results.

The electric energy production cost becomes more and more important. In the literature dedicated to the optimization of cost objective function, a lot of researchers proposed different solutions [1]. In the beginning, one used the method that consisted of loading the most efficient power station until the maximum. Then, one developed the marginal cost method that one applied to the electric network. It is Steinberg and Smith that used it for two generators. This method has been used during several years. With the contribution of data processing one succeeded in elaborating several algorithms of calculation permitting to solve the function cost with a big speed of calculation. It is necessary to note that for the determination of economic dispatch between different electric centrals, it is necessary to consider the transmission losses. These depend on several factors among which the diagram of the network, values of impedances put in game and the distribution of loads and productions, which cheek the most important role in this domain. For this, it is necessary to express the total losses like a function of powers generated. In essence, the economic dispatch is an optimization problem. Several classical optimization techniques, such as, dynamic programming [2], linear programming, homogenous linear programming [3], nonlinear programming [4] and quadratic programming [5] were used to solve economic load dispatch problem. Evolutionary computation (genetic algorithm) optimization techniques are not rigid mathematical methods. They have the ability for nonlinear, non-convex and discontinuous problems. The genetic algorithm has been successfully applied to solve economic dispatch problem [6]. But power losses are not taken into account, which is quite useless because in a practical power system there must be some losses and these losses cannot be ignored. In this paper we include power losses in the genetic modelling to give it a realistic form. Moreover, elitism operator in genetic algorithm is used to make the computation fast. The developed GUI helps the students to grasp the idea of Genetic algorithm quickly and efficiently.

II. Mathematical Model [7]

Mathematical models are of fundamental importance in understanding any physical system. This section focuses on the modelling of economic dispatch problem and its formulation. The factors influencing power generation at minimum cost are operating efficiencies of generators, fuel cost, and transmission losses. The most efficient generator in the system does not guarantee minimum cost as it may be located in an area where fuel cost is high. Also if the plant is located far from the load centre, transmission losses may be considerably higher and hence the plant may be overly uneconomical. Hence the problem is to determine the generation of different plants such that the total operating cost is minimum. In all practical cases, the fuel cost of generator i with power output P_i can be represented as a quadratic function of real power generation as in equation (1)

$$C_i = \alpha_i + \beta_i P_i + \gamma_i P_i^2 \qquad (1)$$

where, $\alpha_i, \beta_i,$ and γ_i represent unit cost coefficients.

The economic dispatch problem is to find the real power generation for each plant such that the objective function (i.e., total production cost) as defined by the equation (2)

$$C_t = \sum_{i=1}^{n} \alpha_i + \beta_i P_i + \gamma_i P_i^2 \qquad (2)$$

is minimum subject to the inequality constraints given by equation (3)

$$P_{i(\min)} \le P_i \le P_{i(\max)} \; i = 1,2,.....,n_g \qquad (3)$$

where $P_{i(\min)}$ and $P_{i(\max)}$ are the minimum and maximum generating limits respectively for plant *i*, subject to the constraint that generation should equal total demand plus losses, i.e.,

$$\sum_{i=1}^{n_g} P_i = P_D + P_L \qquad (4)$$

where C_t is the total production cost, C_i is the production cost of the *i*-th plant, P_i is the generation of *i*-th plant, P_D is the total load demand, and n_g is the total number of displaceable generating plants and P_L is the total transmission loss as a quadratic function of the generator power outputs given by Kron's loss formula in equation (5)

$$P_L = \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} P_i B_{ij} P_j + \sum_{i=1}^{n_g} B_{0i} P_i + B_{00} \qquad (5)$$

The coefficients B_{ij} are called loss coefficients.

III. Genetic Algorithm (GA)

Intelligent systems are expected as a new methodology for solving difficult problems in power systems. Various techniques are researched and applied to power system area [8-9]. A genetic algorithm (GA) [10-13] is a biologically inspired optimization and search technique developed by Holland [14]. At first we have to code our problem into genetic domain. Then we have to use some operators. The simplest genetic algorithm uses three operators

- Selection rules select the individuals, called parents that contribute to the population at the next generation.
- Crossover rules combine two parents to form children for the next generation.
- Mutation rules apply random changes to individual parents

A. Coding

At first the power variables are coded to the binary string (called chromosomes). Although, apart from binary encoding other options are also available, it is used widely.

B. Selection

This is the first rule applied to population. Chromosomes are selected from the population to be parents to crossover and produce offspring. The basic part of the selection process is to stochastically select from one generation to create the basis of the next generation. The requirement is that the fittest individuals have a greater chance of survival than weaker ones. This replicates nature in that fitter individuals will tend to have a better probability of survival and will go forward to form the mating pool for the next generation. Weaker individuals are not without a chance. In nature such individuals may have genetic coding that may prove useful to future generations. Various selection operators are available. Such as

- Roulette-Wheel selection
- Tournament selection
- Rank selection
- Elitism

In this paper we have used Roulette-Wheel an elitism methods are used. In elitism the best chromosome (or a few best chromosomes) is copied to the population in the next generation. The rest are chosen using any selection methods. Elitism can very rapidly increase performance of GA, because it prevents losing the best found solution. In Roulette-Wheel selection Parents are selected according to their fitness (the solution which is most close to the optimal is best fit). The better the chromosomes are, the more chances to be selected they have. Consider a roulette wheel where all the chromosomes in the population are placed as in Figure 1. The size of the section in the roulette wheel is proportional to the value of the fitness function of every chromosome - the bigger the value is, the larger the section is.

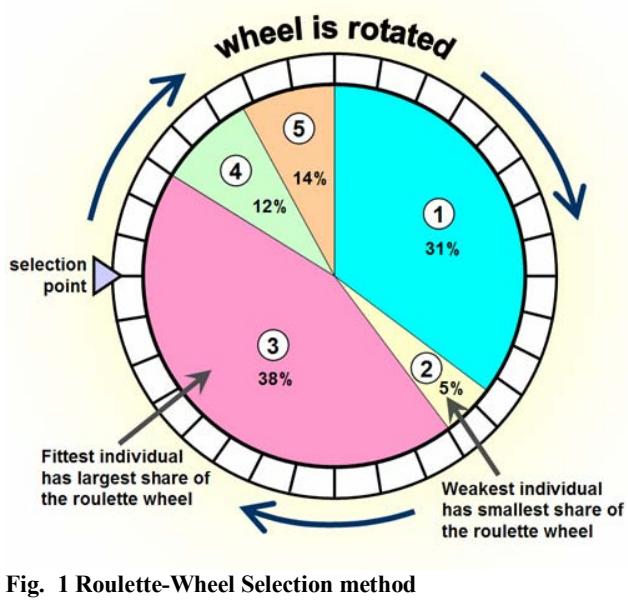


Fig. 1 Roulette-Wheel Selection method

C. Crossover

After the selection phase is over, the population is enriched with better individuals. Crossover operator is applied to the mating pool (the population ready for producing next generation) with a hope that it would create a better string. Various crossover methods are available. Such as

- Single-site crossover
- Two-point crossover
- Uniform crossover, etc.

Here in this paper we used single site crossover. If p1 and p2 are the parents

Parent p1 = [1 0 1 0 0 1 0 0]

Parent p2 = [0 0 1 0 1 0 0 1]

and the crossover point is 3, then the result of crossover is

child c1 = [1 0 1 0 1 0 0 1]

child c2 = [0 0 1 0 0 1 0 0]

Typically for a population size of 30 to 200, crossover rates are ranged from 0.5 to 1.

D. Mutation

While the crossover operation leads to a mixing of genetic material in the offspring, no new genetic material is introduced. The GA mutation operator helps to increase population diversity by introducing new genetic material. Some options are

- Bit inversion
- Order Changing, etc.

Here we used bit inversion method of mutation. If p represents chromosomes before mutation then using bit inversion method at position 4 the outcome m is as follows:

p = [1 0 1 0 0 1 0 0]

m = [1 0 1 1 0 1 0 0]

Typically for a population size of 30 to 200, mutation rates are ranged from 0.001 to 0.05.

IV. Algorithm

The algorithm for solving economic dispatch problem using GA is given below:

- Read cost function coefficients, total load in MW, generator's real power limits, loss formula coefficient matrices. Read Maximum number of generation, populating size, crossover rate, mutation rate, number of bits.
- The initial population strings are generated randomely.
- The strings are decoded to get the power generation schedule according to equation (6)

$$P_i = \frac{P_i^{\max} - P_i^{\min}}{2^{bit} - 1} \text{(Decoded Value of String)} \qquad (6)$$

- Fitness values are calculated using equation (7)

$$f = \frac{1}{1 + \frac{\varepsilon}{P_D}} \text{where}$$
$$\varepsilon = P_D - P_L - \sum_{i=1}^{n_g} P_i \qquad (7)$$

- Selection stage is performed by Roulette-Wheel method utilizing elitism. Crossover is carried out with single cross site with chosen cross over rate. Mutation stage is done with chosen mutation rate. Then new generation is produced and their fitness values are calculated again.
- The previous step is repeated until the maximum fitness in any generation reaches 1.0000 so that the objective is achieved. [15]

V. The GUI for Economic Dispatch Problem

An user friendly graphical user interface is developed for classroom uses as in figure 2. Various steps in Genetic Algorithm can be visualised so the concept of this artificial intelligence technique can be honed easily by the students who are not familiar with GA. Detail help files are provided discussing pros and cons of GA and economic dispatch problems. Some databases are available. But its capability is not limited to the built-in database. User can add new data files according to their requirements. Also the program is compiled such that it can run without the help of MATLAB saving several MB of disk space.

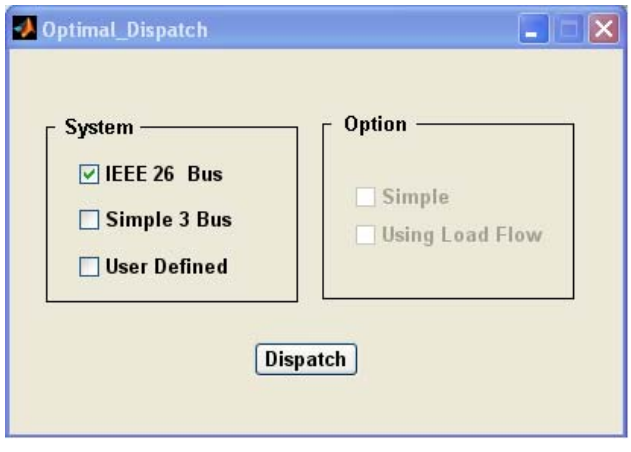


Fig. 2 Graphical User Interface for solving economic dispatch using GA

VI. Results and Comparison

Consider a three bus power system where the fuel costs are given in $/h as in equation (8)

$$C_1 = 200 + 7.0P_1 + 0.008P_1^2$$
$$C_2 = 180 + 6.3P_2 + 0.009P_2^2 \quad (8)$$
$$C_3 = 140 + 6.8P_3 + 0.007P_3^2$$

where P1, P2, and P3 are in MW. Plant outputs are subject to the following limits in equation (9)

$$10MW \le P_1 \le 85MW$$
$$10MW \le P_2 \le 80MW \quad (9)$$
$$10MW \le P_3 \le 70MW$$

The B-loss coefficients for this system are given by equation (10)

$$B = \begin{bmatrix} 0.0218 & 0.0093 & 0.0028 \\ 0.0093 & 0.0228 & 0.0017 \\ 0.0028 & 0.0017 & 0.0179 \end{bmatrix}$$
$$B_0 = \begin{bmatrix} 0.0003 & 0.0031 & 0.0015 \end{bmatrix} \quad (10)$$
$$B_{00} = 0.00030523$$

They are given in per unit on a 100MVA base. Determine the economic dispatch of generation when the total system load is 150MW.

This problem is solved using both classical and GA based methods and the result is shown in Table 1

Table 1 Comparison of Economic Dispatch Using Classical and GA based method

	Classical	GA	Difference
P1 (MW)	33.4701	33.4701	0
P2(MW)	64.0974	64.0974	0
P3(MW)	55.1011	55.1010	0.0001
Power Loss (MW)	2.66873	2.66873	0
Cost($/h)	1599.98	1599.98	0

The results are found using desktop computer in the following environment:

Processor: Intel Dual Core 1.83 GHZ
RAM: 1GB
Operating System: Windows XP (service pack 2)
Programming Language: MATLAB 2007B

The Genetic Parameters and Operators are:

Population Size = 8
Bit = 8
Crossover Rate = 0.5
Mutation Rate = 0.01
Selection Method: Roulette Wheel
Crossover Method: Single point crossover
Mutation Method: Bit Inversion
Elite Count: 1 population per generation

The GA first creates initial population of random binary strings as like Table 2

Table 2 Initial Population String in a run of GA

Pop	P1	P2	P3
1	11111010	10010100	11010011
2	11110010	00100001	01000110
3	10010110	01100100	11111101
4	11101000	10100110	11000011
5	11010000	11111001	11100011
6	11011101	01011101	01100010
7	11101011	00100000	00111001
8	01100101	00010101	00010010

This population now needs be decode to power variable using equation (6) and their finesses are calculated using equation (7) as given in Table 3

Table 3 Power Generation Schedule form initial population

Pop	P1(MW)	P2(MW)	P3(MW)	Fitness
1	24.4118	10.0000	17.7647	0.6043
2	17.3529	19.6078	18.9412	0.6134
3	14.4118	22.0784	22.2353	0.6205
4	20.8824	25.3725	12.5882	0.6206
5	34.4118	13.0196	11.6471	0.6212
6	23.5294	19.8824	19.1765	0.6304
7	12.3529	41.0196	10.2353	0.6326
8	30.2941	10.2745	24.5882	0.6372

From table it is clear that the target is not achieved. So the step is repeated until we get maximum fitness as given in the last row of Table 4.

Table 4 Power Generation Schedule form sixth generation

Pop	P1(MW)	P2(MW)	P3(MW)	Fitness
1	82.0588	24.0000	46.0000	0.9951
2	32.0588	59.6863	61.5294	0.9957
3	72.3529	45.6863	34.2353	0.9960
4	63.8235	30.0392	58.9412	0.9985

5	35.5882	68.1961	48.8235	0.9990
6	72.6471	24.0000	56.1176	0.9993
7	70.5882	42.9412	39.1765	0.9994
8	33.4701	64.0974	55.1010	1.0000

Figure 3 shows the variation of fitness function with generations.

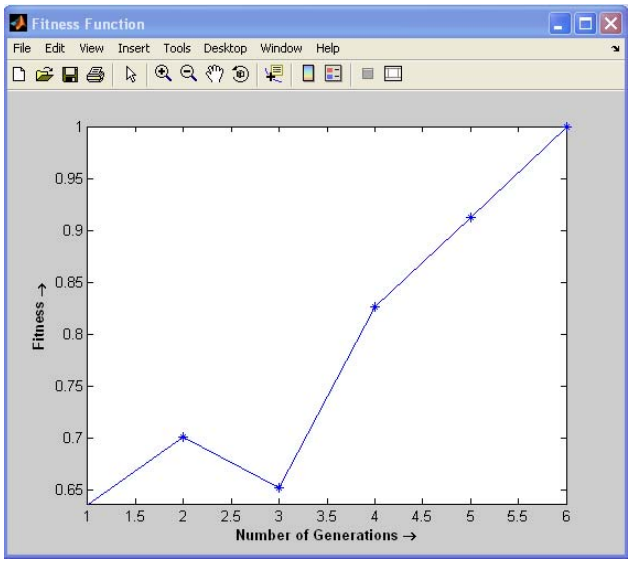


Fig. 3 Fitness function vs. Generation

The results are again compared in Table 5 for 26 bus system having six generator bus system.

Table 5 Comparison of Economic Dispatch Using Classical and GA based method of 26 bus system

	Classical	**GA**	**Difference**
P1 (MW)	447.6919	447.6919	0
P2(MW)	173.1938	173.1938	0
P3(MW)	263.4859	263.4861	0.0002
P4 (MW)	138.8142	138.8142	0
P5(MW)	165.5884	165.5884	0
P26(MW)	87.0260	87.0260	0
Loss (MW)	12.8003	12.8003	0
Cost($/h)	15447.72	15447.72	0

The result shows that both methods yield the same result. But GA based approach is more promising and deserve attention because it does not require derivative information and constraints can be handles easily than classical techniques. Moreover, if the best result is not feasible the the ultimate options are looking for second, third best solutions that are not available using classical optimization methods.

VII. Conclusion

This paper examines the possibility of genetic algorithm in thermal power plant optimization. The results are quite satisfactory. The numerical results show that GA provides acceptable and exploitable solutions. Moreover, the genetic algorithm helps the operator to manage with supple options. Also the GUI facilitates its use for educational purposes. Currently it is used in for UG and PG students in the Dept. of EEE, RUET.

References

[1] Thanathip Sum-im, "Economic Dispatch by Ant Colony Search Algorithm" in Proc. Conf. on Cybernetics and Intel. Systems, Singapore, pp. 416-421, 2004.

[2] D. W Ross and S. Kim, "Dynamic Economic Dispatch of Generation", IEEE Trans. Power Apparatus and Systems, vol. PAS-99, pp. 2060-2068, 1980.

[3] J. Rabin, A. H. Coonick and B. J. Cory, "A Homogeneous Linear Programming Algorithm for the Security Constrained Economic Dispatch Problem," IEEE Trans. Power Systems, vol. 15, pp. 930-936, 2000.

[4] J. Nanda, L. Hari and M. L. Khotari,"Economic Emission Load Dispatch with Line Flow Constraints Using a Classical Technique," IEE Proc. of Generation Transmission. and Distribution, vol. 141, pp. 1-10,1994.

[5] J. Y Fan and L. Zhang, "Real-Time Economic Dispatch with Line Flow and Emission Constraints Using Quadratic Programming," IEEE Trans. Power Systems, vol. 13, pp. 320-325,1998.

[6] C. Benachaiba, M. Haidas, and B. Tahri, "Hybrid Genetic Algorithm Method Based Approach to Optimal Load Dispatch" , Internation Journal of Applied Engineering Research, vol. 1 No. 2 , pp. 265-272, 2006.

[7] Hadi Saadat, "*Power System Analysis*," Tata McGraw HIll,2007.

[8] D. C. Walters and G. B. Sheble, "Genetic Algorithm Solution of Economic Dispatch with Valve Point Loading," IEEE Trans. Power Systems, vol. 8, pp. 1325-1332, Aug. 1993.

[9] W. M. Lin, F. S. Cheng and M. T. Tsay, "An improved Tabu Search for Economic Dispatch with Multiple Minima,"IEEE Trans. Power Systems, vol. 17, pp. 108-112, 2002.

[10] D. E. Golberg, "*Genetic Algorithm in search, Optimization, and Machine Learning*," Addison-Wesly,1989.

[11] Z. Michlewicz, "*Genetic Algorithms +Data Structures=Evolution Program*," Springer-Verlag,1994.

[12] M. Mitchell " *An Introduction to Genetic Algorithms*," Cambridge, MIT Press,1996.

[13] M. D. Vose, "*The simple Genetic Algorithm: Foundations and theory*," Cambridge, MIT Press,1999.

[14] Holland J.H., "*Adaptation in Natural and Artificial System"* Ann Arbor, the University of Michigan Press,1975.

[15] D. P. Kothari and J. S. Dhillon, "*Power System Optimization"* Prentice Hall of India,2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Implementation of Genetic Algorithm and Fuzzy Logic in Economic Dispatch Problem

A. B. M. Nasiruzzaman and M. G. Rabbani

Department of EEE, Rajshahi University of Engineering & Technology
Kazla, Rajshshi-6204, Bangladesh
E-mail: nasir_zaman_eee@yahoo.com

Abstract – Economic dispatch is one of the most challenging problems of power system. Various algorithms for solving optimal power flow problem are found in the literature. The Genetic Algorithm (GA) based solution techniques are more promising than other techniques due to its capability of global searching, robustness, and it does not require derivative information. This paper presents a genetic – fuzzy based approach for solving economic dispatch problem. Algorithms for GA economic dispatch and GA-fuzzy economic dispatch are developed and compared. The results obtained show that GA-fuzzy based approach is faster than GA based approach and less noisy.
Keywords: Economic dispatch, GA, GA-fuzzy

I. Introduction

The term economic dispatch has been given to the problem of minimizing the cost of fuel at thermal plants, assuming that hydro generation has been previously defined and that the configuration of the network is known. It is also known which thermal units are on line. The constraints on this problem, as found in the literature, vary widely, generally trading off complexity for solution speed. Many power systems today are operated under economic dispatch with calculations made on-line every few minutes. Under normal circumstances, control signals are sent to generating stations for generating units to adjust their power output in accordance with optimization results.

The electric energy production cost becomes more and more important. In the literature dedicated to the optimization of cost objective function, a lot of researchers proposed different solutions [1]. In the beginning, one used the method that consisted of loading the most efficient power station until the maximum. Then, one developed the marginal cost method that one applied to the electric network. It is Steinberg and Smith that used it for two generators. This method has been used during several years. With the contribution of data processing one succeeded in elaborating several algorithms of calculation permitting to solve the function cost with a big speed of calculation. It is necessary to note that for the determination of economic dispatch between different electric centrals, it is necessary to consider the transmission losses. These depend on several factors among which the diagram of the network, values of impedances put in game and the distribution of loads and productions, which cheek the most important role in this domain. For this, it is necessary to express the total losses like a function of powers generated. In essence, the economic dispatch is an optimization problem. Several classical optimization techniques, such as, dynamic programming [2], linear programming, homogenous linear programming [3], nonlinear programming [4] and quadratic programming [5] were used to solve economic load dispatch problem.

In the recent past methods using Genetic Algorithms (GAs) [6] have become popular to solve the optimization problems mainly because of its robustness in finding optimal solution and ability to provide near optimal solutions close to global minima. GAs are search algorithms based on the mechanics of natural selection and natural genetics. GAs are different from other optimization methods in the following ways:

- GAs search from population of several points, not a single individual point in the population.
- GAs has inherent parallel computation ability.
- GAs use pay off information (objective function) and not derivatives or auxiliary knowledge.
- GAs use probabilistic transition rules, so they can search a complicated and uncertain area to find the global optimum.

The GAs have been applied to solve economic load dispatch [7-12]. Miranda et al. [13] have provided a survey of three branches of Evolutionary Programming (EP) and the Genetic Algorithms (GAs) and their relative merits and demerits. The references [8-12] have demonstrated the superiority of GA methods in handling continuously non-differentiable objective. For better results and faster convergence, conventional GA models have been modified by including new operators such as elitism, shuffle in reproduction, multi-point or uniform crossover. Considering three added features, a refined GA was used to solve Economic Load Dispatch (ELD) in [9] and a micro GA model in [11]. A Pyramid Genetic Algorithm (PGA) has been used in [14] for voltage profile optimization. The PGA can analytically determine the bound values of mutation and crossover probabilities, which are otherwise, chosen by experience. The GA-

Fuzzy approach presented in this paper is developed to get above mentioned advantages by varying crossover and mutation probabilities throughout the generations using fuzzy-rule base.

II. Mathematical Model [15]

Mathematical models are of fundamental importance in understanding any physical system. This section focuses on the modelling of economic dispatch problem and its formulation. The factors influencing power generation at minimum cost are operating efficiencies of generators, fuel cost, and transmission losses. The most efficient generator in the system does not guarantee minimum cost as it may be located in an area where fuel cost is high. Also if the plant is located far from the load centre, transmission losses may be considerably higher and hence the plant may be overly uneconomical. Hence the problem is to determine the generation of different plants such that the total operating cost is minimum. In all practical cases, the fuel cost of generator i with power output P_i can be represented as a quadratic function of real power generation as in equation (1)

$$C_i = \alpha_i + \beta_i P_i + \gamma_i P_i^2 \tag{1}$$

where, $\alpha_i, \beta_i,$ and γ_i represent unit cost coefficients.

The economic dispatch problem is to find the real power generation for each plant such that the objective function (i.e., total production cost) as defined by the equation (2)

$$C_t = \sum_{i=1}^{n} \alpha_i + \beta_i P_i + \gamma_i P_i^2 \tag{2}$$

is minimum subject to the inequality constraints given by equation (3)

$$P_{i(\min)} \le P_i \le P_{i(\max)} \; i = 1,2,.....,n_g \tag{3}$$

where $P_{i(\min)}$ and $P_{i(\max)}$ are the minimum and maximum generating limits respectively for plant *i*, subject to the constraint that generation should equal total demand plus losses, i.e.,

$$\sum_{i=1}^{n_g} P_i = P_D + P_L \tag{4}$$

where C_t is the total production cost, C_i is the production cost of the *i*-th plant, P_i is the generation of *i*-th plant, P_D is the total load demand, and n_g is the total number of displaceable generating plants and P_L is the total transmission loss as a quadratic function of the generator power outputs given by Kron's loss formula in equation (5)

$$P_L = \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} P_i B_{ij} P_j + \sum_{i=1}^{n_g} B_{0i} P_i + B_{00} \tag{5}$$

The coefficients B_{ij} are called loss coefficients.

III. Algorithm for solution of Economic Dispatch using GA

The algorithm for solving economic dispatch problem using GA is given below:

- Read cost function coefficients, total load in MW, generator's real power limits, loss formula coefficient matrices. Read Maximum number of generation, populating size, crossover rate, mutation rate, number of bits.
- The initial population strings are generated randomely.
- The strings are decoded to get the power generation schedule according to equation (6)

$$P_i = \frac{P_i^{\max} - P_i^{\min}}{2^{bit} - 1} \textbf{(Decoded Value of String)} \tag{6}$$

- Fitness values are calculated using equation (7)

$$f = \frac{1}{1 + \dfrac{\varepsilon}{P_D}} \textbf{ where}$$
$$\varepsilon = P_D - P_L - \sum_{i=1}^{n_g} P_i \tag{7}$$

- Selection stage is performed by Roulette-Wheel method utilizing elitism. Crossover is carried out with single cross site with chosen cross over rate. Mutation stage is done with chosen mutation rate. Then new generation is produced and their fitness values are calculated again.
- The previous step is repeated until the maximum fitness in any generation reaches 1.0000 so that the objective is achieved. [16]

IV. Numerical Result using GA

Consider a 26 bus power system [15] having 46 branches, 6 generators and 7 tap changing transformers. The 26 – bus power system is shown in figure 5 and various data corresponding to this system are given in tables 3-7.The system is simulated using the previous algorithm and result is shown in Table 1. Figure 1 shows the convergence of this algorithm with generations.

Table 1 Economic Dispatch Using GA of 26 bus system

	GA
P1 (MW)	447.6919
P2(MW)	173.1938
P3(MW)	263.4861
P4 (MW)	138.8142
P5(MW)	165.5884
P26(MW)	87.0260
Loss (MW)	12.8003
Cost($/h)	15447.72

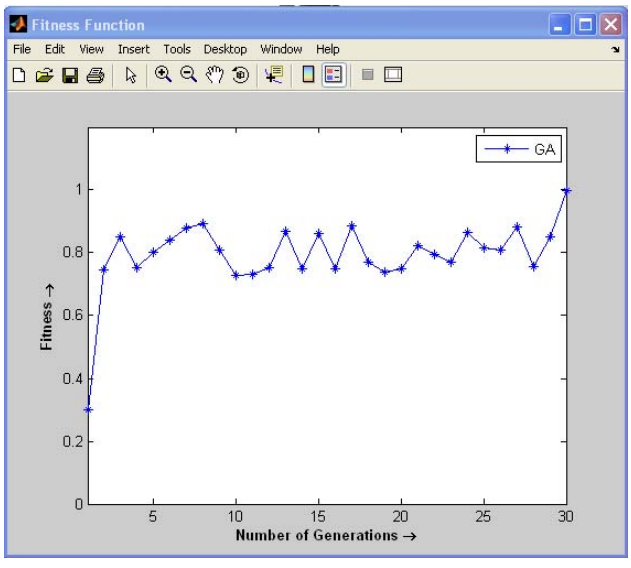


Fig. 1 Fitness function vs. Generation

The results are found using desktop computer in the following environment:

Processor: Intel Dual Core 1.83 GHZ
RAM: 1GB
Operating System: Windows XP (service pack 2)
Programming Language: MATLAB 2007B

The Genetic Parameters and Operators are:

Population Size = 8
Bit = 8
Crossover Rate = 0.5
Mutation Rate = 0.01
Selection Method: Roulette Wheel
Crossover Method: Single point crossover
Mutation Method: Bit Inversion
Elite Count: 1 population per generation

Figure 1 shows that the convergence of this algorithm is noisy. This result stems from the fact that, GA is a random natural selection technique. This oscillation before convergence can be controlled and the computational time can be effectively reduced if we use another artificial intelligence technique, fuzzy logic, to tune some genetic parameters.

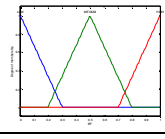
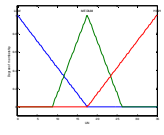
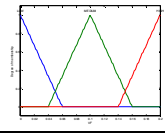
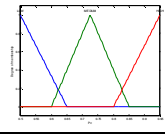
V. GA-Fuzzy Approach in Economic Dispatch

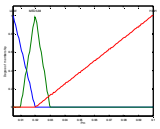
In this paper, a GA-Fuzzy approach is used for solving economic dispatch problem and two GA parameters – crossover rate / probability (Pc) and mutation rate / probability (Pm) are varied dynamically during execution of the program according to a fuzzy knowledge base which has been developed from experience to maximize the efficiency of GA. Therefore, for this purpose the ranges of above parameters have been divided into LOW, MEDIUM and HIGH membership functions and each is given some membership values as shown in Table 2. GA parameters (Pc and Pm) are varied based on the fitness function values as per following logic:

(1) The value of best fitness for each generation (BF) is expected to change over a number of generations, but if it does not change significantly over a number of generations (UN) then this information is considered to cause changes in both Pc and Pm.

(2) The diversity of a population is one of the factors, which influences the search for a true optimum. The variance of the fitness values of objective function (VF) of a population is a measure of its diversity. Hence, it is considered as another factor on which both Pc and Pm may be changed. The membership functions and membership values for these three variables (BF, UN and VF) are selected after several trials to get optimum results.

Table 2 Membership Functions

Variable	Linguistic Term	Membership Function
Best Fitness (BF)	LOW MEDIUM HIGH	
No. of generations for unchanged generation (UN)	LOW MEDIUM HIGH	
Variance of fitness (VF)	LOW MEDIUM HIGH	
Crossover rate/ probability (Pc)	LOW MEDIUM HIGH	

Mutation rate/ probability (Pc)	LOW MEDIUM HIGH	

A. Fuzzy Rule Base

The GA parameters in GA-fuzzy based approach are varied based on the following rules for economic dispatch problem

i) For controlling Pc
1. If BF is LOW then Pc is HIGH.
2. If BF is MEDIUM or HIGH and UN is LOW then Pc is HIGH.
3. If BF is MEDIUM or HIGH and UN is MEDIUM then Pc is MEDIUM.
4. If UN is HIGH and VF is LOW or MEDIUM then Pc is LOW.
5. If UN is HIGH and VF is HIGH then Pc is MEDIUM.

ii) For controlling Pm
1. If BF is LOW then Pm is LOW.
2. If BF is MEDIUM or HIGH and UN is LOW then Pm is LOW.
3. If BF is MEDIUM or HIGH and UN is MEDIUM then Pm is MEDIUM.
4. If UN is HIGH and VF is LOW then Pm is HIGH.
5. If UN is HIGH and VF is MEDIUM or HIGH then Pm is LOW.

Figure 2 and 3 shows the fuzzy rules interface and the total fuzzy inference system (FIS) in MATLAB respectively.

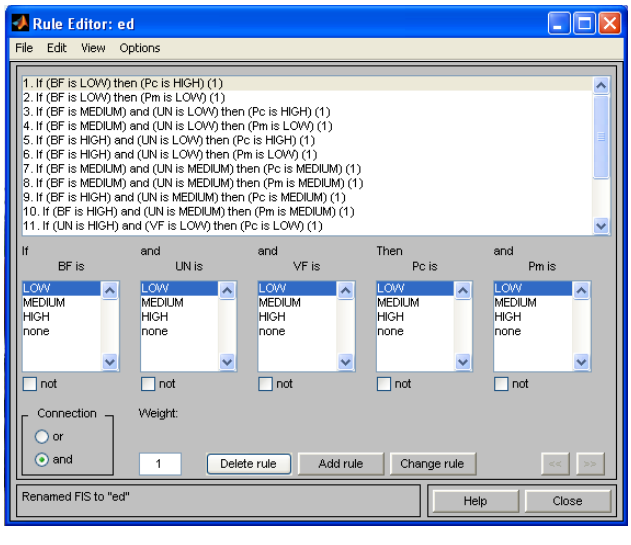


Fig. 2 Fuzzy Rule Base for solving economic dispatch problem in MATLAB

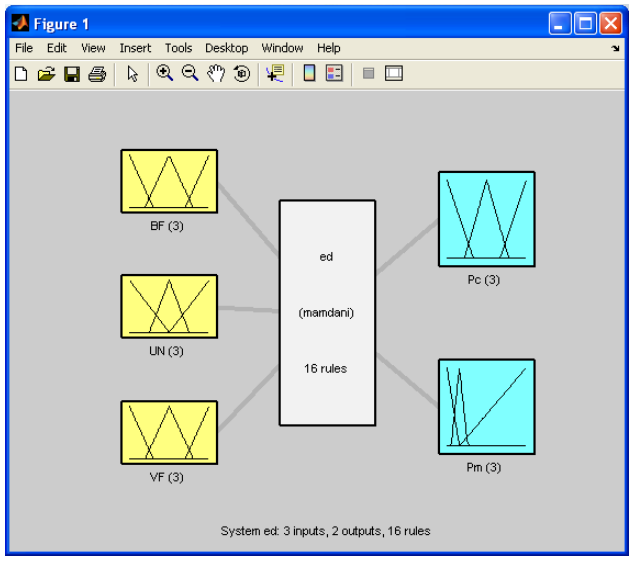


Fig. 3 Fuzzy Inference System (FIS) for solving economic dispatch in MATLAB

VI. Comparison of GA and GA-fuzzy based approach

Figure 4 shows comparison of GA and GA-fuzzy based methods for solving economic dispatch problem in case of 26 bus power system [15]. It is clear from the figure that fuzzy based approach converges more rapidly than the genetic based approach. It is also less noisy, There is no fluctuation. So, from the point of view of computational time, and quality of convergence GA-fuzzy based method exploiting natural selection method and soft computing approach is superior.

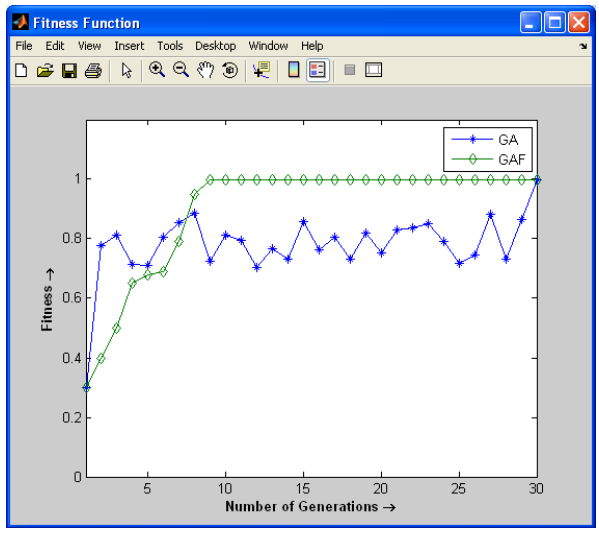


Fig. 4 Comparison of convergence in GA and GA-Fuzzy algorithms

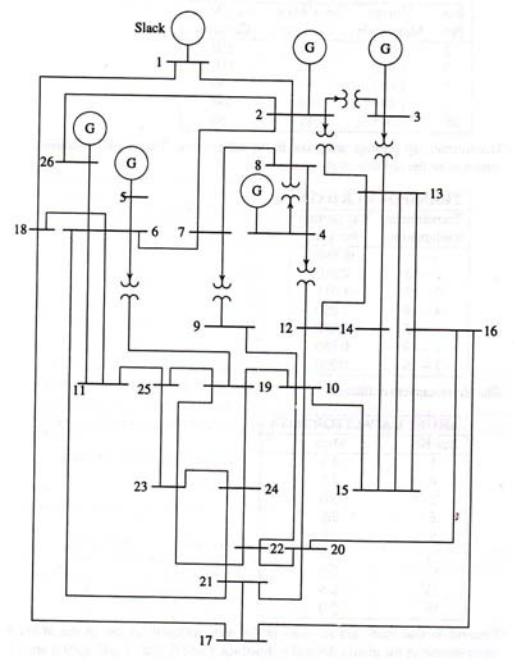


Fig. 5 Single line diagram of 26-bus power system

Table 3 Data for Voltage Controlled buses

Bus No.	Voltage Magnitude	Min MVAR Capacity	Max MVAR Capacity
2	1.020	40	250
3	1.025	40	150
4	1.050	40	80
5	1.045	40	160
26	1.015	15	50

Table 4 Transformer Tap Settings

Transformer Designation	Tap Setting Per Unit
2-3	0.960
2-13	0.960
3-13	1.017
4-8	1.050
4-12	1.050
6-19	0.950
7-9	0.950

Table 5 Shunt Capacitive Data

Bus No.	MVAR
1	4.0
4	2.0
5	5.0
6	2.0
9	3.0
11	1.5
12	2.0
15	0.5
19	5.0

Table 6 Generator Real Power Limits

Gen	Min MW	Max MW
1	100	500
2	50	200
3	80	300
4	50	150
5	50	200
26	50	120

Table 7 Generator Cost Functions

Gen	α	β	γ
1	240	7.0	0.0070
2	200	10.0	0.0095
3	220	8.5	0.0090
4	200	11.0	0.0090
5	220	10.5	0.0080
26	190	12.0	0.0075

VII. Conclusion

A GA-fuzzy based approach to solve economic dispatch problem is presented and applied to 26-bus power system. The proposed algorithm has shown better results in terms of convergence, computational time compared to simple GA based technique. These advantages are mainly due to changes in crossover and mutation probabilities governed by some set of fuzzy rules. So the soft computing method is efficiently utilized in economic dispatch problem.

References

[1] Thanathip Sum-im, "Economic Dispatch by Ant Colony Search Algorithm" in Proc. Conf. on Cybernetics and Intel. Systems, Singapore, pp. 416-421, 2004.

[2] D. W Ross and S. Kim, "Dynamic Economic Dispatch of Generation", IEEE Trans. Power Apparatus and Systems, vol. PAS-99, pp. 2060-2068, 1980.

[3] J. Rabin, A. H. Coonick and B. J. Cory, "A Homogeneous Linear Programming Algorithm for the Security Constrained Economic Dispatch Problem," IEEE Trans. Power Systems, vol. 15, pp. 930-936, 2000.

[4] J. Nanda, L. Hari and M. L. Khotari,"Economic Emission Load Dispatch with Line Flow Constraints Using a Classical Technique," IEE Proc. of Generation Transmission. and Distribution, vol. 141, pp. 1-10,1994.

[5] J. Y Fan and L. Zhang, "Real-Time Economic Dispatch with Line Flow and Emission Constraints Using Quadratic Programming," IEEE Trans. Power Systems, vol. 13, pp. 320-325,1998.

[6] D. E. Golberg, "*Genetic Algorithm in search, Optimization, and Machine Learning*," Addison-Wesly,1989.

[7] A. B. M. Nasiruzzaman, M. G. Rabbani, M. R. I. Sheikh, M. S. Anower, S. K. Shil, and M. A. Haque, 'A Genetic Algorithm Based Approach for Solving Optimal Power Flow Problem', International Conference on Electronics, Computer and Communication (ICECC 2008), pp.476-479, 2008.

[8] D.C. Walters and G.B. Sheble, 'GeneticAlgorithm Solution of Economic Dispatch with Valve Point Loading', IEEE Trans. On Power Systems, Vol. 8, No. 3, pp.1325-1332, 1993.

[9] G.. B. Sheble and K. Britigg , 'Refined Genetic Algorithm-Economic Dispatch Example', IEEE Transactions on Power Systems, Vol. 10, No. 1, pp. 117-124,Feb. 1995.

[10] Chen Po-Hung and Chang Hong-Chan, 'Large–Scale Economic Dispatch by Genetic Algorithm', IEEE Transactions on Power Systems, Vol. 10, No. 4, pp. 1919-1926, Nov. 1995.

[11] S.O. Orero and M.R. Irving, 'A Genetic Algorithm for Generator Scheduling in Power Systems', International Journal on Electric Power and Energy Systems, Vol. 18, No. 1, pp. 19-26, Jan 1996.

[12] Chira Achyuthakan, 'Genetic Algorithms Applications to Economic Load Dispatch', Master Thesis, AIT Bangkok, Thailand, August 1997.

[13] V Miranda, D. Srinivasan and L.M., Proenca 'Evolutionary Computation in Power Systems', A Survey paper, Proc. Of 12th PSCC, Dresden, Germeny, 19-23, pp.25-40, August 1996.

[14] S. K. Lee, K. M. Son and J. K. Park, 'Voltage Profile Optimization using a Pyramid Genetic Algorithm', Proc. ISAP 97, Seoul, Korea, pp.407-414. July, 1997.

[15] Hadi Saadat, "*Power System Analysis*," Tata McGraw HIll,2007.

[16] D. P. Kothari and J. S. Dhillon, "*Power System Optimization"* Prentice Hall of India,2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

EVALUATION OF OPTIMUM CAPACITY AND LOCATION OF EAST WEST INTERCONNECTROR USING TIE LINE PLANNING APPROACH

Abdul Hasib Chowdhury, Q. Ahsan
Department of EEE, Bangladesh University of Engineering and Technology
Dhaka 1000, Bangladesh
E-mail: hasib@eee.buet.ac.bd, qahsan@eee.buet.ac.bd

Abstract – Bangladesh is physically divided into two parts, Eastern and Western, by the combined flow of Jamuna-Meghna river system. Because of the availability of natural gas in Eastern part, most of the generating plants of the country are installed here. Therefore, it is important for the utility to search optimum interconnection(s), connecting Eastern and Western parts, for capacity transaction such that minimum production cost is ensured. This paper investigates for an appropriate approach to evaluate the optimum capacity and location of East-West interconnector. The tie line planning approach is applied for the evaluation. The approach considers both peak load and hour load models independently. The results are compared with the simulated results obtained using conventional transmission expansion planning technique.

Key words - Interconnection planning, optimal power flow, correlated demand, transmission expansion planning.

I. Introduction

Large power exchanges between utilities are becoming order of the day because of improved reliability and increased economical savings by exploiting global cheap generation resources, taking advantage of peak diversity and time zone difference between interconnected systems and time difference of forced outages of generating units of two systems, reducing the overall spinning reserve and minimising interruption cost. These consequently result in reduced investment requirement. Potential benefits of interconnection also include environmental protection. The incorporation of interconnection issues is thus becoming an important component of power system planning.

From an operational point of view, tie lines are not simply transmission line that carries energy from the generating plant to the load. They are interconnections that facilitate energy exchange between power systems in a pre-defined manner, which is agreed upon by the management of the interconnected systems.

Interconnection planning (IP) requires the evaluation of location and capacity of tie lines that ensures optimum operating cost and optimum rate of return on investment. The problem is combinatorial, which precludes algorithms that guarantee a global optimum.

Until now, IP is considered as a part of transmission expansion planning (TEP). [1]. The objective of TEP is to determine location and time to set up transmission line(s) to remove any existing transmission line overload at a minimum cost, given the base network configuration, the generation facilities and the forecasted demands for a target year [2]. On the other hand, the objective of IP is to connect, with the minimum investment cost, two or more power systems so that the global production cost will be the minimum and the reliability will be maximum for the planning period. TEP and IP thus have specific non-identical objectives.

Regardless of these differences, no special consideration has been given to IP in the literature. Recently, a technique is developed for IP [3,4]. The present state of the art in TEP neither has the tools available to address the problem of interconnection planning (IP), nor any research has been carried out to solve the IP problem.

The TEP problem is formulated as a linear minimization problem [5-13]. Usually, linear cost function is used to represent the transmission line expansion and the operating costs. This has the drawback of yielding only approximate results since most of the cost functions involved are not linear. The operating cost is generally expressed as cost due to loss of load [8,12,13]. Sawey et al. [9] and Kim et al. [10] included generation cost at a flat rate in the operating cost function. Levi et al. [14] included generation cost at a flat rate due to added transmission lines only. Some of the researchers do not include the generation cost [5-6,15-17].

This paper investigates to find optimum capacity and location of East-West interconnector by applying first IP approach, the results are compared with those obtained using conventional transmission expansion planning technique.

II. Tie line planning

In what follows the tie line planning technique introduced in [3,4] is briefly described.

The problem of determining the location and capacity of tie lines that would render the global investment for tie lines and production cost a minimum, even for a single peak-load instant, is a large combinatorial problem. Solving the problem with a time variant load model for firm investment decision makes it even larger and more

complex. A heuristic approach is used in the tie planning technique to solve such a complex problem.

Loads of the interconnected systems are considered correlated. Forecasted hourly load of the interconnected systems are used. A multi-level load model is developed for the analysis, which decomposes the planning problem into a number of smaller optimization problems that are solved using a full ac OPF.

Three cost components are used to evaluate the optimum tie line. These are i) cost of generation, ii) capital cost of the tie line, iii) penalty cost for tie line of under- or over-capacity.

At first, at every load level, the OPF is used to search for local minima to evaluate the optimized capacity of a tie line. In the next stage, a penalty function is applied over the whole planning horizon to calculate the contribution of each local minima to the global cost function. One aspect of the penalty function ensures the maximum utilization of a tie capacity, thus ensuring higher rate of return resulting in minimum investment cost. This amounts to imposing the capacity upper limits on a tie-set. The other aspect of the penalty function ensures optimum energy transfer for minimum global generation cost over the planning period. This imposes the lower limits on a tie-set. Finally, local minima are ranked according to global cost incurred by them.

III. Bangladesh Power System

The Bangladesh power system (BPS) is a relatively small system having 128 buses and 25 generation plants. The combined flow of Jamuna-Meghna river system has divided the country into two parts, the Eastern part and the Western part. This natural separation together with the availability of natural gas only in the Eastern part of country has influenced the development of BPS network and the installation of generating plants.

The Eastern part of the network is further divided into three zones for the ease of system operations. These are the South zone comprising of Chittagaong division, Central zone comprising greater Sylhet and Mymenshing, and Dhaka Zone, comprising Dhaka division except greater Mymenshing. The Western part of the network is divided into two zones, namely West Zone comprising Khulna division and North Zone comprising Rajshahi division.

The backbone of the system is 230 kV and 132 kV transmission systems that connect all zones. Dhaka zone is central to the system in terms of both generation and loads. All other zones are connected radially to Dhaka Zone, but they are not interconnected among themselves.

The Eastern part and the Western part were interconnected in 1982 through the 230 kV double circuit East-West Inter-connector.

In this paper, 2006 is assumed to be the base year. Accordingly the system data for 2006 has been used as the base data.

IV. Simulation

4.1 Simulation conditions

Three potential locations of interconnectors, tie lines, between the Eastern part and the Western part of BPS are considered, as listed in Table 1. These three tie lines and also their combinations are evaluated to see whether all three tie lines are necessary and whether their capacities are optimum for a planning period of fourteen years from the year 2007 to year 2020.

Table 1 Tie lines considered for BPS

Tie line	Interconnection between buses			
	From bus		To bus	
	Bus no.	Location	Bus no.	Location
T21	1101	Haripur	1401	Ishwardi
T22	1130	Ghorashal	1401	Ishwardi
T23	1201	Ashuganj	1401	Ishwardi

The forecasted hourly loads of the BPS from year 2007 through 2020 [17] are considered in this simulation. It is assumed that the loads of two parts of BPS maintain a fixed ratio. The forecasted hourly load for BPS is divided in that ratio to obtain demand for the two parts.

Table 2 summarizes the load level and their duration during the planning period. Total duration of the planning period is 1,22,640 hours.

Table 2 BPS demands over the planning period

Load level	Duration, %	Demand					
		System total		Eastern part		Western part	
		MW	MVAR	MW	MVAR	MW	MVAR
1	7.8	2744	1498	2168	1183	576	315
2	26.4	3511	1917	2774	1514	737	403
3	26.6	4486	2449	3544	1935	942	514
4	21.1	5474	2989	4324	2361	1150	628
5	11.8	6412	3501	5065	2766	1347	735
6	6.2	7776	4245	6143	3354	1633	891

4.2 Simulation results

The results are presented in Table 3. It shows the complete stage-by-stage calculation results for all the load levels and the ranks of optimum tie lines. Column 1 gives the load level index. Each row of Table 3 presents the results at each load level. Column 4 presents the optimum tie set for each load level while column 5 presents the capacity for the optimum tie set. Column 6 and 7 presents the investment cost and generation cost at a particular level for the optimum tie set. Column 8 is the summation of columns 6 and 7.

As explained earlier, Stage 1 searches for the optimum interconnection for each load level by minimizing the investment and generation cost for that level. It is seen that in stage 1 calculations, TS21 is evaluated as the optimum tie set for load level 1 and TS22 for load level 2.

Table 3 Stage by stage calculations for BPS

(1)	(2)	(3)	Stage 1					Stage 2		Stage 3	
Load level	System total demand, (MW)	Demand duration, (hr)	(4) Tie set	(5) Tie capacity, (MVA)	(6) Investment cost, ($)	(7) Gen. cost at current load level, ($)	(8) Investment cost and generation cost, ($)	(9) Gen. cost at other load levels, $	(10) Penalty, ($)	(11) Total cost, ($)	(12) Rank
1	2744	9546	TS21	577	50280000	15493158	65773158	349433842	2757046	417964046	2
2	3511	32148	TS22	740	64000000	67832280	131832280	232133720	2042051	366008051	1
3	4486	32455	TS26	306, 640	99801800	90516995	190318795	446508205	4186380	641013380	3
4	5474	25735	TS26	522, 642	110701800	89660740	200362540	436263460	7842196	644468196	4
5	6412	14441	TS26	516, 647	110651800	60089001	170740801	495392199	7742838	673875838	5
6	7776	7571	TS26	329, 267	82301800	39308632	121610432	578654568	14478660	714743660	6

For load levels 3 through 6, TS26 is evaluated as the optimum tie set.

Stage 2 calculates the generation cost at other load level for each optimum tie set. It is presented in column 9. Column 10 presents total penalty for the optimum tie obtained in stage 1. Stage 3 calculates the total cost for a particular optimum tie set by summing columns 8, 9 and 10. The value of this summation is presented in column 11. Column 12 presents the final ranking based on the value of column 11. The ranking is done in a descending manner. The tie set with rank 1 is selected for the actual interconnection.

Simulation results using IP approach using single load level is presented in Table 4. For comparison purpose, simulation has been done considering peak load level only. The considered peak loads are 6143 MW for the Eastern zone and 1633 MW for the western zone. All results that obtained tie line capacity of less then 250 MVA has been discarded as these are not feasible at the given voltage level of 230 kV. It is observed that three different tie-sets, namely TS22, TS24 and TS26, are obtained as viable through this approach.

Simulation results using transmission line planning concept is presented in Table 5. The approach developed by Graver [5] has been adapted for this purpose. In this approach, load flow is used to test for line over loads. Any line over load is removed by inserting new lines with a step-by-step increase in their capacities. Only the peak load is considered for the whole planning period.

Table 4 Optimum tie sets obtained using IP method with peak load level only

Tie set	Tie capacity, MVA	Inv. cost, $\$x10^3$	Gen. cost, $\$x10^3$	Total cost without penalty, $\$x10^3$	Total cost with penalty, $\$x10^3$	Rank
TS22	478	23900	365682	389582	391184	2
TS24	370	18500	367231	385731	389195	1
TS26	329, 267	29800	369778	399578	406685	3

Table 5 Optimum tie set obtained using TEP method

Tie set	Tie capacity, MVA	Inv. cost, $\$x10^3$	Gen. cost, $\$x10^3$	Total cost, $\$x10^3$	Rank
TS21	1113	77080	366428	443508	1
TS22	1113	82652	362996	445648	2
TS23	1242	87600	364898	452498	3
TS26	356, 753	103882	370821	474703	4

Table 6 compares the results of all three methods as applied to the BPS. It is observed from the table that all the techniques have obtained tie sets different from each other. Tie planning technique using multiple load level has obtained tie set TS22 while tie planning using peak load only has obtained tie set TS24 and transmission expansion planning obtained tie set TS21 as an optimum one. The capacity of the tie line as obtained differs considerably. While the capacity obtained from the multiple load level tie planning is 740 MVA, that for the

single load level technique is 370 MVA. Tie line capacity as obtained from transmission expansion planning technique is 1113 MVA. The total cost of TS22 is lower than that of TS24. The total cost without penalty of TS21 is the highest. Moreover, its investment cost is also the highest. Besides the cost, TEP does not ensure the optimum energy transaction between interconnected parts. Therefore, TS22 is the optimum tie line obtained using IP simulation technique with multiple load model.

Table 6 Comparison of results for three techniques

Method	Optim. tie set	Tie capacity, MVA	Inv. cost, $\$x10^3$	Gen. cost, $\$x10^3$	Total cost, (without penalty), $\$x10^3$	Total cost, (with penalty), $\$x10^3$
IP (multiple load level)	TS22	740	64000	299966	363966	366008
IP (single load level)	TS24	370	18500	367231	385731	389196
TEP	TS21	1113	77080	366428	443508	--

V. Conclusion

The methodology for planning tie lines between interconnected systems is applied to BPS to evaluate the optimum capacity and location of East-West interconnector.

The results are compared with the conventional transmission line planning approach and also with single peak load approach.

Comparison of results shows that the developed technique gives a better result as well as better understanding of the problem of tie line planning.

References

1. R. Romero, A. Monticelli, "A hierarchical decomposition approach for transmission network expansion planning", *IEEE Transactions on Power Systems*, Vol. 9, No. 1, pp. 373-380, February 1994.
2. G. C. Oliveira, A. P. C. Costa, S. Binato, "Large scale transmission network planning using optimization and heuristics techniques", *IEEE Transactions on Power Systems*, Vol. 10, No. 4, pp. 1828-1833, November 1995.
3. Abdul Hasib Chowdhury, "A new mathodology for planning tie lines between interconnected power systems", Ph.D. thesis, Dept. of EEE, Bangladesh University of Engineering and Technology, July 2008.
4. Abdul Hasib Chowdhury, Q. Ahsan, "A novel approach of tie line planning between interconnected power systems", *IEEE Transactions on Power Apparatus and Systems*, (communicating).
5. Len L. Garver, "Transmission network estimation using Linear programming", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-89, pp. 1688-1697, February 1970.
6. R. Villasana, L. L. Garver, S. J. Salon, "Transmission network planning using linear programming", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-104, No. 2, pp. 349-356, February 1985.
7. M. V. F. Pereira, L. M. V. G. Pinto, S. H. F. Cunha, G. C. Olliveira, "A decomposition approach to automated generation/transmission expansion planning", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-104, No. 11, pp. 3074-3083, November 1985.
8. R. Romero, R. A. Gallego, A. Monticelli, "Transmission system expansion planning by simulated annealing", *IEEE Transactions on Power Systems*, Vol. 11, No. 1, pp. 364-369, February 1996.
9. Kern J. Kim, Young M. Park, Kwang Y. Lee, "Optimal long term transmission expansion planning based on maximum principle", *IEEE Transactions on Power Systems*, Vol. 3, No. 4, pp. 1494-1501, November 1988.
10. Ronald M. Sawey, C. Dale Zinn, "A mathematical model for long range expansion planning of generation and transmission in electric utility systems", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-96, No. 2, pp. 657-666, March/April 1977.
11. R. Romero, R. A. Gallego, A. Monticelli, "Transmission system expansion planning by simulated annealing", *IEEE Transactions on Power Systems*, Vol. 11, No. 1, pp. 364-369, February 1996.
12. Cesar Serna, Jorge Duran, Arturo Camargo, "A model for expansion planning of transmission systems, a practical application example", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-97, No. 2, pp. 610-615, March/April 1978.
13. R. A. Gallego, A. Monticelli, R. Romero, "Comparative studies on non-convex optimization methods for transmission network expansion planning", *IEEE Transactions on Power Systems*, Vol. 13, No. 3, pp. 822-828, August 1998.
14. Viktor A. Levi, Milan S. Calovic, "A new decomposition based method for optimal expansion planning of large transmission networks", IEEE Transactions on Power Systems, Vol. 6, No. 3, pp. 937-943, August 1991.
15. A. Monticelli, A. Santos Jr., M. V. F. Pereira, S. H. Cunha, B. J. Parker, J. C. G. Praca, "Interactive transmission network planning using a least-effort criterion", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 10, pp. 3919-3925, October 1982.
16. R. Romero, C. Rocha, M. Mantovani, J. R. S. Mantovani, "Analysis of heuristic algorithms for the transportation model in statdic and multistage planning in network expansion systems", IEE Proceedings on Generation Transmission and Distribution, Vol. 150, No. 5, pp. 521-526, September 2003.
17. Moin Uddin, "Development of a Methodology for Long Term Hourly Load Forecasting and Assessment of Gas Requirement", M. Sc. Thesis, Dept. of EEE, Bangladesh University of Engineering and Technology, August 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Future Electric Energy Demand of Bangladesh

Moin Uddin[1], *Q. Ahsan*[2]

[1] Department of Electrical, Electronic & Communication Engineering
Military Institute of Science & Technology (MIST), Dhaka-1216, Bangladesh
email: moinmist @ yahoo.com
[2] Department of Electrical & Electronic Engineering
Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
email: qahsan @ eee. buet.ac.bd

Abstract- Present society demands electricity as its basic need in every moment. Future electric energy demand data is an essential requirement for the expansion analysis of all sectors of a power system. This paper presents a logical assessment of the requirement of electric energy for Bangladesh in coming 15 years with a view to provide guideline to the power system expansion planners and country's energy policy makers. The assessment is based on the forecasting of hourly electrical demands. It also forecasts the annual peak loads with a clear indication of the magnitude of their increase in future. Such an evaluation may create awareness among the power system expansion planners to meet the huge demand in future. It also warrants the attention of the energy policy makers to ensure the best utilization of the available indigenous energy resources of the country.

I. Introduction

Bangladesh is a land with a population of about 140 million people. Only 35% of its people have an access to electricity with a poor per capita electric energy consumption of about 137 kWh. Bangladesh Power System (BPS) can meet only 75% of the total demand [1]. Generation shortage forces BPS for massive load shading hindering nation's development activities. Fossil fuels of the country are still the main indigenous resources for producing electricity. The discovered fossil fuels are gas and coal whose reserve is also limited and is being depleted with the use.

All these parameters indicate insecurity in the field of electric energy sector of the country. The prevailing scenario demands a great awareness among the nation and deserves appropriate measures. A realistic electric energy and peak load forecast of the coming future will help policy makers develop an appropriate policy of the use of indigenous fuel resources and also help generation expansion planners develop an optimal generation expansion plan.

Different approaches were made in the past to forecast the energy requirement for various sectors of Bangladesh. The energy requirement for a long term (2000-2015) is forecasted for the Power System Master Plan (PSMP) of Bangladesh Power Development Board (BPDB) [2]. It uses the forecasted annual peak and constant load shape as the basic load model.

The technique to forecast hourly load for future expansion plan is broadly developed in 1977. Lot of efforts was made at various levels to forecast the hourly loads for both the short and long terms using different techniques [3-5]. The appropriateness of the projection of electric energy for a period depends mainly on the realistic forecasting of the electrical load of every minimum possible interval of that period. Q. Ahsan and M.A Jalil introduced a significant development in the field of long term hourly load forecasting [6]. The methodology establishes a relation between the forecasted hourly load and the forecasted annual peak load through some load ratios; hourly, daily, weekly and monthly. A load ratio is considered as a ratio between the peak loads of two different durations of that period. As such, the forecasted hourly loads become the function of some single valued quantities; i.e. the peaks; which cannot characterize any sorts of load variations over a period.

This paper presents the forecasted electric energy demand for a planning horizon of 15 years (2008 – 2022). It also forecasts the annual peak loads for the period. The paper utilizes the forecasting methodology presented in [7, 8]. The technique is based on the development of probabilistic models of some load ratios. It develops a relationship between load ratios and annual/seasonal average load.

II. Methodology

The methodology applied to forecast the electric energy demand is presented in [7,8]. The methodology establishes a relation between the forecasted hourly loads and forecasted annual average loads through some load ratios. Its basic steps are (i) determination of load ratios, hourly, daily and monthly from the historical data; (ii) development of the probability density function (PDF) of each load ratio, (iii) evaluation of the expected value,

mean, of each load ratio, and (iv) forecasting of annual average load.

A. Load Ratios

The first step in this process is to obtain load ratios; hourly, daily and monthly from the historical demand. The hourly load ratio (HR) of a particular hour is obtained by dividing the load of that hour by the average hourly load of the day, in which the hour belongs, that is,

$$HR_i = \frac{HL_i}{DAL_j} \ , \ i = 1, ..., n_1 .....(1)$$

Where *HLi* is the hourly load of i th hour and DAL_j is the daily average hourly load of j th day in which i th hour belongs. n_1 is the number of operating, energy supply/consume, hours in a year/season. The daily load ratio (DR) may be obtained by dividing the average hourly load of that day by the average hourly load of the month in which the day belongs. And, the monthly load ratio (MR) is the ratio of the average hourly load of the month and that of the year in which the month belongs.

B. Probability Density Function (PDF)

The next step is to obtain PDF of all load ratios. The PDF of any *i*th hour load ratio is developed by sampling the *i*th hour load ratios of known history at an equal interval of time. The interval may be a year or more.

C. Forecasted Hourly Load

Now, the forecasted hourly load of any *i*th hour in future, HL_i of *j*th day, *k*th month and *l*th year may be expressed as:

$$HL_i = m(HR_i) \times m(DR_j) \times m(MR_k) \times FYAL_l \qquad (2)$$

In equation (2), m represents the expected or mean value. $FYAL_l$ represents the forecasted annual average hourly load of *l*th year. This may be determined by using any standard forecasting technique [9, 10].

III. Forecasted Annual Energy and Peak Demand

The methodology [7] used to forecast the annual energy requires the forecasting of hourly average load of each hour of the year. The validation of the used technique to forecast hourly load and annual energy is presented. To do so, the historical data of BPDB are collected. The peak loads are also validated.

A. Validation

In what follows, the validations of average hourly loads, annual average energy and annual peak load are presented.

A.1 Hourly Load

The hourly loads of five years 1998-2002 are forecasted. The same are compared with the collected data of Central Load Dispatch Center (CLDC), BPS. The deviations of the forecasted results from the actual ones are depicted in Fig. 1.

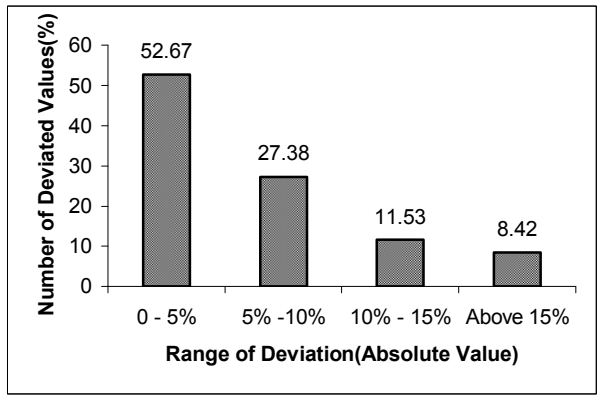


Fig. 1: Deviation of Forecasted Hourly Loads from Actual Ones

It is clearly observed from the figure that only in 20% cases the deviations are more than 10%. The deviations of most of the forecasted hourly loads from the corresponding actual ones are within 5%.

A.2 Annual Energy

The methodology presented in [7] is used to forecast the annual energies of 2000–2007. Acres International Limited, Canada forecasted the energies of these years during the preparation of Power System Master Plan (PSMP) for BPDB in 1995[2]. The forecasted values of this paper are compared with the actual ones along with the forecasted values of [2] in Table 1.

Table 1: Validation of Annual Energy

Year	Actual Energy (GWh)	Energy Forecasted in GWh by		Deviations (%)			
		This Study	PSMP	This Study		PSMP	
				%	Mean	%	Mean
2000	16723	16640	15632	0.49	2.23	6.52	4.04
2001	17865	18235	16849	2.07		5.68	
2002	19139	18957	18167	0.95		5.08	
2003	20540	20025	19596	2.51		4.60	
2004	22010	21182	21144	3.76		3.93	
2005	23558	23376	22823	0.77		3.12	
2006	25216	24595	24662	2.46		2.20	
2007	26970	25681	26651	4.78		1.18	

There is a clear observation that the forecasted values of PSMP are far from the actual ones. However, the forecasted values of this study are close to the actual. The mean deviations of this study and PSMP are, respectively, 2.23% and 4.04%.

A.3 Peak Load

The annual peak loads of BPDB system are forecasted using the technique presented in [9, 10] for the years 1998-2007. These values are compared in Table 2 with the actual annual peak loads collected from CLDC. Although the maximum deviation is 15.2%, however, the mean deviation is around 5%.

Table 2: Validation of Annual Peak Loads

Year	Peak Loads (MW)		% Deviation (Absolute Value)	Mean Deviation
	Actual	Forecasted		
1998	2897	2587	10.70	5.27%
1999	2903	2659	08.40	
2000	2925	2714	7.20	
2001	3084	2843	7.80	
2002	3208	3046	5.04	
2003	3235	3257	0.68	
2004	3618	3497	3.33	
2005	3782	3746	0.95	
2006	3480	4009	15.2	
2007	4130	4288	3.83	

B. Forecasted Values

The annual peak loads and energies of BPS are forecasted for coming 15 years. These are presented in the following subsections.

B.1 Forecasted Annual Energies

The future annual energy demand of BPS for the period of 15 years (2008-2022) is forecasted and these are presented in Table 3. To forecast the annual energy of a future year the forecasted hourly loads of each hour of that year are added. The table also presents the annual incremental energy of each year. This is determined by comparing the same with the forecasted energy of the previous year.

Table 3: Forecasted Energy (FE) in GWh

Year	FE (GWh)	% Increment	Year	FE (GWh)	% Increment
2008	28826	6.99	2016	47045	5.97
2009	30732	6.61	2017	49616	5.68
2010	32771	6.63	2018	52419	5.64
2011	34905	6.51	2019	55317	5.52
2012	37171	6.49	2020	58448	5.47
2013	39460	6.16	2021	61401	5.22
2014	41880	6.13	2022	64613	5.18
2015	44396	6.01			
Total demand for 15 years (2008 – 2022): 679000 GWh					

From the table it is clear that the annual energy demand increases over the years. But the incremental rate decreases gradually.

The cumulative energy of the forecasted period is depicted in Fig 2. It shows that the total electric energy demand over the coming 15 years is 679000 GWh or 679 kGWh. To produce this amount of energy, 8.0218 trillion cubic feet (TCF) of gas is required at the rate of 11.806 CFT/kWh, which is about 66.85% of the country's known reserve of 2008 (12.00 TCF) [7, 11]. To produce the same energy by fossil oil the amount of money required is 438714 Million US Dollar at the fuel consumption rate of 0.4224 liter/kWh [7]. This calculation considers the present oil price over the whole horizon of 15 years. It does not consider the present worth factor.

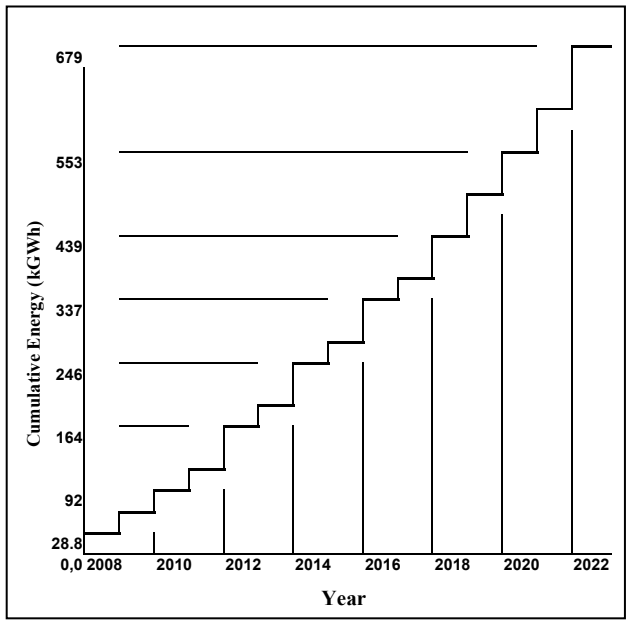


Fig. 2: Cumulative Energy for 15 Years (2008 -2022)

B.2 Forecasted Annual Peak

Table 4 presents the forecasted annual peak loads (PL) from 2008 to 2022. The table also presents the incremental value of peak load of each year. This is calculated by comparing the value of the previous year.

Table 4: Forecasted Annual Peak Loads

Year	PL (MW)	% Increment	Year	PL (MW)	% Increment
2008	4582	6.86	2016	7482	5.88
2009	4892	6.75	2017	7913	5.76
2010	5216	6.63	2018	8359	5.64
2011	5556	6.51	2019	8821	5.52
2012	5911	6.39	2020	9297	5.40
2013	6281	6.26	2021	9789	5.29
2014	6666	6.13	2022	10296	5.18
2015	7067	6.01			

The table shows that the peak load becomes more than double within fifteen years. However, the rate of increment decreases slowly with time, like the annual energy growth.

Fig 3 depicts the normalized peaks of coming 15 years. The normalized peak value of any year is obtained by dividing the forecasted peak of that year by the peak value of 2008.

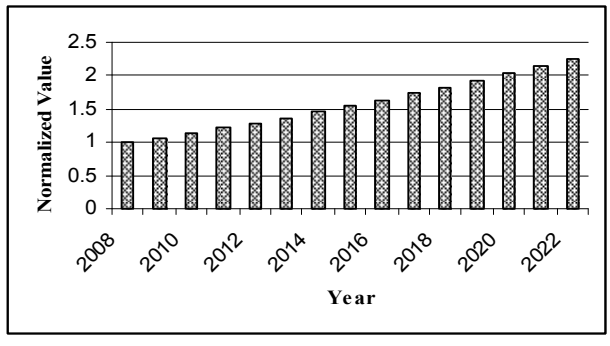


Fig 3: Normalized Value of Forecasted Annual Peak

The figure shows that the peak load grows smoothly to 225% in only 15 years. This growth of peak load indicates that the generation capacity should be proportionally increased; an installed capacity of about 10,300 MW [Table 4] or capacity transactions from neighboring countries should be made through interconnections.

IV. Conclusions

This paper forecasts the electric energy demand in coming 15 years (2008-2022) along with the corresponding annual peak loads. It also investigates the annual increment in both the cases. The energy forecast is based on the forecasting of hourly electrical demands. The validation of the proposed methodology clearly reveals that the deviation of the forecasted energy from the actual one is insignificant.

The results of the investigations warrant due awareness of the energy planners to meet the future energy demand, there by ensuring energy security to the consumers. It may also be a guideline for the decision makers to ensure the best utilization of the limited indigenous energy resources of the country.

References

[1] Statistical Yearbook of Bangladesh 2006, Twenty Seventh Edition, Bangladesh Bureau of Statistics, Chapter VI.

[2] Power System Master Plan – Bangladesh, Final Report, Volume 1-3, Prepared by Acres International Ltd, Asian Development Bank TA No. 1962-BAN, Directorate of System Planning, BPDB, August 1995.

[3] V.Pansuka, "Short-term Forecasting of Electric Power System Load from a Weather- Dependent Model", IFAC Symposium 1977 on Automatic Control and Protection of Electric Power Systems Melbourne, Australia, Feb 21 - 25, 1977.

[4] A.S. Dehdashti, J.R. Tudor, M.C. Smith, "Forecasting of Hourly Load by Pattern Recognition - A Deterministic Approach", IEEE Transactions Vol. PAS - 101, No. 9, September 1982, pp. 3290-3294.

[5] P.C. Gupta and K. Yamada, "Adaptive Short- term Forecasting of Hourly Loads using Weather Information", IEEE Transactions Vol. PAS-91, September-October 1972, pp. 2085-2094.

[6] Q**.** Ahsan and M. A. Jalil, "A New Technique for Forecasting Hourly Load for Long-Term Planning," IEEE Transactions on Power and Energy System, Vol-25 No.2, 2005.

[7] Moin Uddin "Development of a Methodology for Long Term Hourly Electrical Load Forecasting and Assessment of Gas Requirement", M. Sc. Engineering Thesis, Department of EEE, BUET, Dhaka, August 2003.

[8] Q. Ahsan and Moin Uddin "Electrical Energy Forecasting - A Novel Probabilistic Approach" IEEE Transaction for Power Systems (Communicating).

[9] K.C. Jam and L.N. Aggarwal, "Production Planning Control and Industrial Management", Khanna Publishers, Fifth Edition 1990 (Forth Reprint 1998) Part 3, Chapter 3.

[10] X. Wang and J.R. Mc Donald, "Modern Power System Planning", Mc.Graw Hill International, Editions 1994, pp.1 to 130.

[11] Moin Uddin and Q. Ahsan, "Expected Status of Natural Gas in Bangladesh in Meeting the Electricity Demand in the Next Twenty Five Years" ICECE 2004, December 28-30, 2004 Dhaka, Bangladesh.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Nonconcatenative Morphology: An HPSG Analysis

Md. Shariful Islam Bhuyan and Reaz Ahmed

Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
E-mail: sharifulislam@cse.buet.ac.bd

Abstract - In spite of being a successful syntactic theory in many respects, Head-driven Phrase Structure Grammar (HPSG) has inadequate coverage for morphological constructions, especially for nonconcatenative morphology, which is prominent in the Semitic languages such as Arabic, Hebrew etc. In this paper, we extend the HPSG framework to support rich nonconcatenative morphology. We also define the first comprehensive HPSG-construction for the morphology of the verbal system of Arabic, the best instance of nonconcatenative morphology among the living languages.

I. Introduction

Broad-coverage precision grammar [1]-[3] and computational lexicon development for deep linguistic processing is a research-intensive area with several potential applications [4]. Amidst the vast literature on formal linguistic theory [5], Head-driven Phrase Structure Grammar (HPSG) [6] has a unique position since it combines the best features of the contemporary approaches as well as establishes an integrated framework for cross-layer representation comprising phonology, morphology, syntax, semantics, pragmatics and discourse. Although, HPSG successfully describes numerous syntactic and semantic phenomena, it lacks rigorous analyses for morphological phenomena, especially for non-concatenative morphology [7]. Nonconcatenative morphology illustrates an interesting paradigm of morphological operations, which is prominent in the Semitic languages such as Arabic, Hebrew etc [10]-[11].

Among the living languages, Arabic demonstrates the best instance of nonconcatenative morphology. Arabic verb system exhibit both concatenative and nonconcatenative morphology, capable of lexically expressing diverse syntactic and semantic phenomena. Formalisms of existing morphological analyzers for Arabic are not powerful enough to capture this higher layer diversity. In this paper, we extend the HPSG framework to support rich nonconcatenative morphology. We also define the first comprehensive HPSG-construction for the morphology of Arabic verbal system using our extension.

II. Nonconcatenative Morphology

Morphology deals with the study of the patterns of word formation in a particular language, description of such patterns and the behaviour and combination of morphemes. It is difficult to define any linguistic object precisely, since they vary from language to language. However, we can identify some properties of the concept "word", which is a grammatical unit and used as a minimal possible unit in a reply. Word boundaries impose restrictions over the phonological stress. A word is the largest unit, which denies the insertion of new constituents within its boundaries. It is also the smallest constituent that can move within a sentence without making the sentence ungrammatical.

A morpheme is the smallest meaningful unit in the grammar of a language. The word "dogs" consists of two morphemes: dog, and -s, a plural marker on nouns. A bound morpheme is a grammatical unit that never occurs by itself, but always attached to some other morpheme. Here, the plural morpheme -s in dogs is a bound morpheme. A free morpheme is a grammatical unit that can occur alone. However, other morphemes such as affixes can attach to it. Here the word dog is a free morpheme.

We can form new words from existing ones by morph-syntactic operations. There are two kinds of morpho-syntactic operation, inflection and derivation. Inflectional operations create forms that can be readily embedded in the sentence with discourse compliance, whereas derivational operations create forms that cannot be necessarily embedded in the sentence and which may still require inflectional operations before they can be integrated into discourse. In the example "He speaks for people", the word "speak" inflects to "speaks". This represents the temporal aspect of the action to the time of utterance as well as third-person, singular-number actor attributes. However, in the example "They are the speaker for us", the word "speaker" cannot be readily integrated. Although, this word is derived from "speak", it needs to be inflected to "speakers". Inflection does not change the lexical category of the word and contribute syntactically constrained information, such as number, gender, or aspect. However, derivation often changes the lexical category of the word (e.g. "speak" is verb and "speaker" is noun) and contribute different lexical meaning.

In the previous examples, the bound morphemes "-er" and "-s", which are joined after the word "speak", are called affix. A root is characterized, as the part of a word, which is common to a set of derived or inflected forms, cannot be further analyzed into meaningful units when all

affixes are removed, and carries the principle portion of meaning of the words. For example, "speak" is the root of the words speaks, spoke, spoken, speaking, speaker, speakers, spokesperson and others. There are two kinds of affixes, inflectional affix and derivational affix. A derivational affix is an affix by means of which one word is derived from another. For example, "-er" of the word "speaker" is a derivational affix. An inflectional affix is an affix by means of which one word is inflected from another. For example, "-s" of the word "speakers". A stem is the root with any derivational operation, to which inflectional affixes are added. In our case, the word "speaker" is a stem.

Morphology deals with two kinds of information. First, what information is encoded by the morpheme. For example, we can take an Arabic word "kataba" – he wrote. In this paper, we use an appropriate English transliteration for Arabic alphabet. A variety of information is encoded in this word and its other inflected or derived form. Some are,

- Agreement: *kataba* - he wrote.
  Person – Third
  Number – Singular
  Gender – Masculine
  Mood – Indicative
- Agency: *kutiba* - it was written.
  Voice – Passive
- Event structure: *kataba*
  Tense – Past
  Aspect – Perfect
- Illocutionary force: *uktub* – write.
  Mode – Command
- Part-of-Speech: *kitaabun* - a book
  kataba –Verb *kitaabun* –Noun
- Definiteness: *al-kitaabu* - the book
  Determiner – Definite
- Complex Predicate: *kattaba* - he made to write.
  Semantic relation: Causation

There are many more syntactic and semantic phenomena those can be expressed using morphology.

Second issue, with which morphology deals with, is how information is encoded in the morpheme. Morpho-syntactic operations performed over the morphemes come with two flavors: concatenative and nonconcatenative.

Concatenative operations are those where morphemes are linearly concatenated. For example:

- Prefixation: clear | *un*clear
- Suffixation: walk | walk*ed*
- Circumfixation: mind | *un*mind*ful*

Nonconcatenative operations are those where morphemes are nonlinearly embedded. For example:

- Infixation: kataba | kat*t*aba
- Simulfixation: *e*at | at*e*
- Modification: man | m*e*n
- Suppletion: go | *went*

There are many other morpho-syntactic operations also. In this paper, we mainly focus on nonconcatenative operation and give a mathematical formalism to capture their rich diversity.

III. Morphology in Arabic

Arabic language exhibits an extremely rich morphology [8]-[9]. Both concatenative and nonconcatenative operations take place in the formation of an Arabic word. Inflection is made by concatenative operations whereas derivation is made by non-concatenative operations.

Arabic word formation is an excellent example of root-pattern morphology. A combination of root letters are plugged in a variety of morphological pattern with priory fixed letters and particular vowel melody that gives rise to corresponding syntactic and semantic phenomena. To feel the richness of Arabic morphological patterns, which we call "measure" in this paper, following example is given. Here, the root letters 'k', 't', 'b', bearing a concept of writing, is plugged in various measures to get a myriad of syntactic and semantic phenomena. The measures with a particular semantic paradigm are called "Form". Arabic has many forms. Among them, ten forms are used regularly. The root letters 'k', 't', 'b' can be plugged in among nine of them.

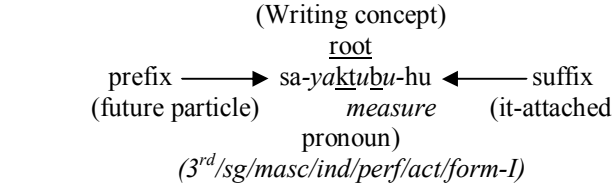
- Form I (Transitive): *kataba* – He wrote.
- Form II (Causative): *kattaba* – He caused to write.
- Form III (Ditransitive): *kaataba* – He corresponded.
- Form IV (Factitive): *aktaba* – He dictated.
- Form V (Reflexive): *takattaba* – It was written on its own.
- Form VI (Reciprocity): *takaataba* – They wrote to each other.
- Form VII (Submissive): *inkataba* – He was subscribed.
- Form VIII (Reciprocity): *iktataba* – They wrote to each other.
- Form X (Control): *istaktaba* – He asked to write

The above example illustrates the derivational paradigm of Arabic word. However, there is also an inflectional paradigm, which is governed by the agreement information. Every entry of the table 1, can take fourteen inflectional form according to there number gender and person. For imperfect form, there are three such inflectional paradigms. Table 2 and 3 show the inflectional paradigm for active perfect and passive perfect entry of form I.

Table 1 Derivational Paradigm of root "ktb"

	FORM I	FORM II	FORM III	FORM …
Active perfect	katab-a	kattab-a	kaatab-a	…
Passive perfect	kutib-a	kuttib-a	kuutib-a	…
Active imperfect	ya-ktub-u	yu-kattib-u	yu-kaatib-u	…
Passive imperfect	yu-ktab-u	yu-kattab-u	yu-kaatab-u	…
Active imperative	u-ktub	kattib	kaatib	…
Passive Imperative	litu-ktab	litu-kattab	litu-kaatab	…
Verbal noun	kitaab-atun	ta-ktiib-un	kitaab-un	…
Active participle	kaatib-un	mu-kattib-un	mu-kaatib-un	…
Passive participle	ma-ktuub-un	mu-kattab-un	mu-kaatab-un	…
Locative participle	ma-ktab-un	…	…	…
Instrumental participle	mi-ktab-un	…	…	…
…	…	…	…	…

An Arabic word can encode a complete sentence. For example, *sayaktubuhu* - He will write it. We can break the word in the following component.



From the diagram, we can conclude that an Arabic word has four components. First, a prefix part – "sa"; Second, a suffix part – "hu"; Third, the root letters – 'k', 't', 'b'; Fourth, the measure – "ya_ _u_u". It may be possible to concatenate multiple prefixes and suffixes. However, there must be a single measure and single set of root letters, where the measure packages syntactic and semantic features and root supplies the core concept. If we plug in another set of root letters, for example, 'n', 'S', 'r', which bears the concept of helping, we get *sayanSuruhu* – He will help him.

Table 2 Inflectional Paradigm of Form I-Active-Perfect

Ind/Sub/Juss	Singular	Dual	Plural
3rd/Masc.	katab-a	katab-aa	katab-uua
3rd/Fem.	katab-at	katab-ataa	katab-na
2nd/Masc.	katab-ta	katab-tuma	katab-tum
2nd/Fem.	katab-ti	katab-tuma	katab-tunna
1st	katab-tu	katab-na	katab-na

Table 3 Inflectional Paradigm of Form I-Passive-Perfect

Ind/Sub/Juss	Singular	Dual	Plural
3rd/Masc.	kutib-a	kutib-aa	kutib-uua
3rd/Fem.	kutib-at	kutib-ataa	kutib-na
2nd/Masc.	kutib-ta	kutib-tuma	kutib-tum
2nd/Fem.	kutib-ti	kutib-tuma	kutib-tunna
1st	kutib-tu	kutib-na	kutib-na

In our analysis, a measure may contain two parts – stem-measure and affix-measure. From the above tables, we note a crucial point that, for a particular inflectional paradigm, a certain portion of the word, containing all the root letters, is always constant. For table 2 and 3, these are "*katab*" and "*kutib*" respectively. We call this fixed portion, the "stem-measure" and the remaining part, containing prefix and/or suffix, the "affix-measure". Depending on this analysis, we can give the following model of an Arabic word.

A root-derived Arabic word =
Prefix + affix-measure (stem-measure (Root)) + Suffix

There are syntactic and semantic features, which governs the derivational and inflectional paradigms for Arabic roots. With a linguistic investigation, we have listed some features that will be used in this paper.

Attributes in the table 4 and 5 govern the derivational and inflectional paradigm for an Arabic root respectively.

Table 4 Attributes Governing Derivational Paradigm

Attribute	Values
POS	noun, verb, particle
FORM	I, II, III, IV, …
VOICE	active, passive
VFORM	perfect, imperfect, imperative

Table 5 Attributes Governing Inflectional Paradigm

Attribute	Values
MODALITY	emphatic, uncertainty
MOOD	indicative, subjunctive, jussive
PERSON	1st, 2nd, 3rd
NUMBER	singular, dual, plural
GENDER	masculine, feminine
CASE	nominative, accusative, genitive
DEFINITENESS	definite, indefinite
POLARITY	affirmative, negative

This is not the whole story of Arabic morphology. Another facet of Arabic morphology is the concept of root class. We call a set of roots, which share a common derivational and inflectional paradigm, a root class. Depending on the characteristics of root letters, the class is determined. The roots 'k', 't', 'b' and 'n', 'S', 'r' both are member of same root class – the "sound" root class. Some example root classes are,

- Weak: Root contains one or more weak letters
- Hamzated: Root contains a hamza
- Geminate: Root contains two similar consonant
- Sound: None of these irregularities

IV. An HPSG Framework

Head-driven Phrase Structure Grammar (HPSG) is a mathematical formalism of natural languages. To understand the motivation of HPSG we need to start from its predecessor Context Free Grammar (CFG). A Context Free Grammar *G* is a 4-tuple $G = (\sum, V, S, P)$ where,

- $\sum$ is a finite, non-empty set of terminals, the alphabet;
- *V* is a finite, non-empty set of non-terminals;
- $S \in V$ is the start symbol;
- *P* is a finite set of production rules, each of the form $A \rightarrow \alpha$, where $A \in V$ and $\alpha \in (V \bigcup \sum)^*$

For example, let we have the following CFG for very small fragment of English,

- $\sum$ = {eat, eats, …, rice, Sharif, …, I, you, he, … };
- V = {S, VP, NP, V, N, P}
- S = Start symbol
- P = { S → NP VP
  VP → V NP
  VP → V
  NP → N | P
  V → eat | eats | …
  N → rice | Sharif | …
  P → I | you | he |… }

Using the above CFG, the sentence "I eat rice" can be analyzed by the following derivation:

S → NP VP → P VP → I VP → I V NP
→ I eat NP → I eat rice

However, there are problems with CFG. The above definition also generates the sentence "I eats rice".

S → NP VP → P VP → I VP → I V NP
→ I eats NP → I eats rice

Second derivation is grammatically wrong. We do not capture the accurate agreement information with the grammar *G*. Basic problem lies with CFG is that its terminals are completely non-informative. This leads to Head-driven Phrase Structure Grammar (HPSG), a constraint-based lexicalist formalism of natural language. In HPSG our grammar fragment will look like,

$$\sum = \{\text{eats}\begin{bmatrix} HEAD & verb \\ SUBJ & \begin{bmatrix} AGR & \begin{bmatrix} NUM & sg \\ PERS & 3rd \end{bmatrix} \end{bmatrix} \end{bmatrix},$$

$$\text{I}\begin{bmatrix} HEAD & noun \\ AGR & \begin{bmatrix} NUM & sg \\ PERS & 1st \end{bmatrix} \end{bmatrix}, \ldots\};$$

V = {S, VP, NP, V, N, P}
S = Start symbol

P = {
S → NP[*AGR* [1]] VP[*SUBJ* [*AGR* [1]]]
VP → [*HEAD* *verb*]
NP → [*HEAD* *noun*]
…
}

Now, we have put the agreement information in the terminals. We do not handle the transitivity here. HPSG will not license the second derivation, since it violates the constraint in the first phrase structure rule – agreement of the noun phrase and agreement of the subject of the verb phrase must match. Here, a technique, called structure-sharing is used. There are two boxed one in the rule. It means that the value of these two agreements will share the same value. We call the information-bearing terminals as lexical sign and non-terminals as phrasal sign, $\sum$ as lexicon, phrase structure rules as constructs and the matrix associated with each sign as attribute value matrix, according to HPSG terminology.

The fundamental concept of HPSG is the notion of signs and constructs. Sign is a formal representation of a valid linguistic object, which can be a lexeme, word and phrase (including sentences). Constructs are the rule or schema that licenses a particular combination of signs. HPSG assumes that a language is an infinite set of signs.

Here, we face the problem of selecting appropriate attributes for a particular language. Attributes are selected by linguistic motivation as well as language independent requirements. A linguistic object can be captured at multiple layers. For example, a sign have attributes that can be phonological, morphological, syntactic, semantic, pragmatic and so on. Well-established representation of signs captures these different aspects of a linguistic object using an attribute value matrix.

$$\begin{bmatrix} PHON & phon-obj \\ MORPH & morph-obj \\ SYN & syn-obj \\ SEM & sem-obj \\ \vdots & \vdots \end{bmatrix}$$

Fig. 1. Representation of a linguistic sign in HPSG

In our analysis, we construct the MORPH feature that can handle non-concatenative morphology in Arabic. We also modify SYN and SEM features for Arabic.

V. Analysis in HPSG

In this section, we give our formal representation of an Arabic sign. First, we present the construction of our MORPH feature in the figure 2.

$$MORPH\begin{bmatrix} TYPE & sound \\ ROOT & \langle [1]k, [2]t, [3]b \rangle \\ FORM & I \\ STEM & [4]\langle [1], a, [2], a, [3] \rangle \\ AFFIX & [4] \oplus \langle a \rangle \end{bmatrix}$$

Fig. 2. Representation of the MORPH of *kataba* in HPSG

We have five features associated with morphology. First, the feature TYPE, which denotes the associated root

class. In this case, its value is "sound". Next, the feature ROOT, which is the list root letters. Here, its value is 'k', 't' and 'b'. Next, the feature FORM, which denotes the semantic paradigm. *kataba* – is a form-I derivative. Next, two features are for our previously defined stem-measure and affix-measure. From the table 2 we find that the fixed stem is – *katab.* Stem-measure is captured using another list which elements are structurally shared with the type ROOT. Our last feature is used to capture the affix-measure *–a*, which denotes the 3rd-person, sg-number and masc-gender. The full list of stem-measure is structurally shared and concatenated with the list of affix-measure.

$$\left[\begin{array}{ll}
MORPH & \left[\begin{array}{ll} TYPE & sound \\ ROOT & \langle [1]k,[2]t,[3]b\rangle \\ FORM & I \\ STEM & [4]\langle [1],a,[2],a,[3]\rangle \\ AFFIX & [4]\oplus\langle a\rangle \end{array}\right] \\
SYN & \left[\begin{array}{ll} CAT & \left[\begin{array}{ll} HEAD & verb \\ VFORM & perfect \\ VOICE & active \end{array}\right] \\ VAL & \left\langle\left[\begin{array}{ll} SYN & \left[\begin{array}{ll} HEAD & noun \\ CASE & acc \\ OPT & - \end{array}\right] \\ SEM & [INDEX \quad j] \end{array}\right]\right\rangle \end{array}\right] \\
SEM & \left[\begin{array}{ll} INDEX & s \\ FRAMES & \left\langle\left[\begin{array}{ll} INDEX & i \\ PNG & \left[\begin{array}{ll} PER & 3rd \\ NUM & sg \\ GEN & masc \end{array}\right] \end{array}\right], \left[\begin{array}{ll} REL & write \\ SIT & s \\ ACTOR & i \\ UNDGR & j \end{array}\right], \left[\begin{array}{ll} REL & perfect \\ ARG & s \end{array}\right]\right\rangle \end{array}\right]
\end{array}\right]$$

Fig. 3. Representation of *kataba* in HPSG

For, generic nonconcatenative templatic morphology, only the features ROOT, STEM and AFFIX can serve the purpose. If there is a root class concept, then the TYPE attribute might be needed. However, with an extensive type hierarchy, the feature TYPE and FORM can be taken out and the STEM and AFFIX values can be classified. However, these details results cannot be presented here.

In figure 3, we present our SYN and SEM matrix for the word – *kataba.* First, the CAT features identifies the syntactic category of – *kataba*. It contains the HEAD feature for syntactic head and two syntactic features of Arabic, which governs the derivational paradigm of verb lexeme. Next, the VAL feature, which captures the subcategorization of verbs. VAL is a list of signs, which are required by the syntactic head. In this case, the verb – *kataba*, requires an object. The verb – "write" is a transitive verb that takes an object. We should note a point here. In other languages, verbs subcategorize for both their subject and object. However, in Arabic, there is a concept of hidden pronoun. This means, the word – *kataba*, means – he wrote. The pronoun "he" is encoded by the inflectional morphology, when no explicit subject is used. The semantic actor is not realized syntactically. So, the verb only subcategorizes for syntactic object. We can also see the constraints imposed over the object. In this version, its syntactic head should be a noun phrase with the value of its CASE feature set to "accusative". The negative value of the OPT feature indicates that this object is not optional, rather required to be syntactically correct.

Next, we need to consider some semantic features. Here, we use a type feature version of predicate logic to capture semantics of natural language. First, we consider the INDEX feature, which is a reference to a discourse entity. Then, the PNG feature, which capture the semantics of PERSON, NUMBER and GENDER. Next, the FRAMES feature, which serves as a bag for elementary predicates to describe the situation at hand. For example, in the case of *kataba*, the event of writing is expressed. The event is completed in the past and there is a discourse referent to the actor. To capture the core event, *write*-predicate is introduced. To capture the temporal constraint, we use the *perfect*-predicate. Finally, to express the actor of the event, the hidden pronoun, we introduce a discourse referent with corresponding PNG feature. Predicates have their respective arguments. *write*-predicate has a situation hook, expressed by the feature SIT. There are two semantic role associated with this predicate. First, we consider the role of *writer*, who plays a doer role, expressed by the feature ACTOR. Second, we consider the role of *written*, who plays an undergoer role, expressed by the feature UNDGR. The *perfect*-predicate takes a situation hook as an argument, which is expressed as the feature ARG. We use the important technique of co-indexing for sharing of semantic objects. The discourse referent predicate is actually the actor of the *write*-predicate. To denote this constraint, the INDEX value of hidden pronoun and the ACTOR value of the *write*-predicate are co-indexed, both are given the value *i.* This is an example of reference co-indexing. We also use event co-indexing. The event hook SIT of *write*-predicate, situation hook of the entire scenario and argument ARG of the *perfect*-predicate, all are co-indexed and expressed using the value *s.* Another important of this HPSG representation is the syntax-semantics interface. In this example, this is done by co-indexing the INDEX value of the syntactic object and the UNDGR value of the *write*-predicate with a value *j.* This indicates that the syntactic object is our semantic undergoer whereas from our previous discussion we can note that the semantic actor is not syntactically realized.

In figure 4, we show the HPSG representation of the passive – *kutiba.* We identify the associated changes for this conversion. The stem-measure is changed to capture the derivational morphological operation whereas we can see that the affix-measure is not changed. This phenomenon illustrate the motivation of isolating the measure into two parts – stem-measure and affix-measure. Next change can be found obviously in the

feature VOICE, changing its value to *passive*. Unlike English, which can have a prepositional complement in passives, Arabic passives do not subcategorize for a subject or any other argument. For this reason, the VAL list is empty. Moreover, the discourse referent in the feature FRAMES is now co-indexed with the UNDGR feature of the *write*-predicate, expressed by the value *j*. Semantic actor now completely unknown by not having any syntactic or semantic reference, which is a distinctive property of Arabic passive.

$$\left[\begin{array}{ll}
\textit{MORPH} & \left[\begin{array}{ll} \textit{TYPE} & \textit{sound} \\ \textit{ROOT} & \langle [1]k,[2]t,[3]b\rangle \\ \textit{FORM} & I \\ \textit{STEM} & [4]\langle [1],u,[2],i,[3]\rangle \\ \textit{AFFIX} & [4]\oplus\langle a\rangle \end{array}\right] \\
\textit{SYN} & \left[\begin{array}{ll} \textit{CAT} & \left[\begin{array}{ll} \textit{HEAD} & \textit{verb} \\ \textit{VFORM} & \textit{perfect} \\ \textit{VOICE} & \textit{passive} \end{array}\right] \\ \textit{VAL} & \langle\ \rangle \end{array}\right] \\
\textit{SEM} & \left[\begin{array}{ll} \textit{INDEX} & s \\ \textit{FRAMES} & \left\langle \left[\begin{array}{ll} \textit{INDEX} & j \\ \textit{PNG} & \left[\begin{array}{ll} \textit{PER} & 3rd \\ \textit{NUM} & sg \\ \textit{GEN} & masc \end{array}\right] \end{array}\right], \left[\begin{array}{ll} \textit{REL} & \textit{write} \\ \textit{SIT} & s \\ \textit{ACTOR} & i \\ \textit{UNDGR} & j \end{array}\right], \left[\begin{array}{ll} \textit{REL} & \textit{perfect} \\ \textit{ARG} & s \end{array}\right] \right\rangle \end{array}\right]
\end{array}\right]$$

Fig. 4. Representation of *kutiba* in HPSG

In this way, we can construct other matrix also for every entry of table 1, not all of which can be covered in this paper. However, we sketch the fundamental strategy of capturing the nonconcatenative morphology in the core of an HPSG sign.

VI. Conclusion

In this paper, we give the proposal how to capture nonconcatenative morphology, especially Arabic verb morphology within the framework of HPSG. There are lot of works to do in the future. To construct matrix from table 1 we need to cope with a wide range of diversity that an Arabic verb can take. Results will be immensely helpful for the construction of resource grammar for languages with rich nonconcatenative morphology.

References

[1] A. Copestake, and D. Flickinger, "An open-source grammar development environment and broad-coverage English grammar using HPSG," *Second conference on Language Resources and Evaluation*, 2000.

[2] M. Marimon, N. Bel, S. Espeja, and N. Seghezzi, "The Spanish Resource Grammar: pre-processing strategy and lexical acquisition," *ACL Workshop on Deep Linguistic Processing,* 2007.

[3] B. Comrie, R. Fabri, B. Hume, M. Mifsud, T. Stolz, and M. Vanhove, (Eds), "Towards an HPSG Analysis of Maltese," *1st International Conference on Maltese Linguistics,* 2007.

[4] F. Bond, S. Oepen, M. Siegel, A. Copestake, and D. Flickinger, "Open source machine translation with DELPH-IN," *Open-Source Machine Translation Workshop at the 10th Machine Translation Summit*, pp. 15-22, 2005.

[5] P. Sells, *Lectures on Contemporary Syntactic Theories*, Stanford: CSLI Publications, 1985.

[6] I. A. Sag, and T. Wasow, *Syntactic Theory: A Formal Introduction*, Stanford: CSLI Publications, 1999.

[7] S. Bird, and E. Klein, "Phonological Analysis in Typed Feature Systems", *Computational Linguistics*, vol. 20, pp. 55-90, 1994.

[8] K. R. Beesley, "Finite-State Morphological Analysis and Generation of Arabic at Xerox Research: Status and Plans in 2001", *ACL Workshop on Arabic Language Processing: Status and Prospects*, pp. 1–8, 2001.

[9] O. Smrž, *Functional Arabic Morphology. Formal System and Implementation*, PhD Dissertation, Charles University in Prague, 2007.

[10] S. Z. Riehemann, *A Constructional Approach to Idioms and Word Formation*, PhD Dissertation, Stanford University, 2001.

[11] S. Z. Riehemann, "Type-Based Derivational Morphology," *Journal of Comparative Germanic Linguistics*, vol. 2, pp. 49-77, 1998.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Parameter Estimation Method for Linear Amplitude Modulated Chirp Signals Based on Discrete Fractional Fourier Transform

Saurav Zaman Khan Sajib[1], and Ahmed Mostayed[2]

[1]ALCATEL-LUCENT, Dhaka, Bangladesh
[2] Department of Electrical Engineering, Kongju National University, Kongju, South Korea
E-mail: sajib127@hotmail.com[1], shaibal125@yahoo.com[2]

Abstract – A new parameter estimation method for linear amplitude modulated chirp signal is proposed. This method utilizes the Discrete Fractional Fourier Transform (DFRFT) along with Radon Wigner Transform (RWT) to estimate the amplitude and phase parameters of multi-component chirp signals. Expressions of DFRFT for estimating amplitude and initial phase parameters are derived. The chirp rate and initial frequency is estimated using RWT. Performance of the proposed method in noise is analyzed by Monte-Carlo simulation for different Signal to Noise Ratio (SNR). Better performance was achieved at SNR as low as -10 dB.

I. Introduction

Polynomial Phase Signals (PPS), especially linear chirp signals are frequently used in Mobile communications, Sonar and Synthetic Aperture Radar (SAR) imaging of moving targets. It is well known that the radar echo of a moving target with constant acceleration is a chirp signal. By estimating the chirp rate and initial frequency of the received signal, one can achieve valuable information about the velocity and acceleration of the target. In such applications the amplitude of the chirp signal is considered as a nuisance parameter which does not contain any significant information. A more recent application of linear chirp signals have been the modelling of speech signals with multi-component linear amplitude modulated chirp signals [1]. In that case, the amplitude parameters are as important as the chirp rates and initial frequencies of component signals. Hence, techniques that estimate the phase parameters (Initial frequency, chirp rate and initial hase) as well as amplitude parameters (Amplitude slope and constant amplitude) accurately under noisy environment are to be established.

This paper addresses the parameter estimation problem of linear amplitude modulated chirp signals. Before proceeding it is worthy to take a quick look at the history of related researches. In the early 90s the time-frequency analysis based on Wigner –Ville (WVD) distribution was applied for detection and imaging of moving objects with SAR [2]. But the cross-term associated with Wigner distribution hampers the estimation performance in extreme noisy conditions. However, the time-frequency analysis is restricted to applications where chirp rate is the only parameter of interest. To overcome that, the radon transform of WVD (RWT) was proposed by Wood [3]. This method is based on the line integral of the time-frequency plane along all possible lines and the outcome of the transform is localized maxima on the initial phase-chirp rate plane. The computational complexity of RWT was reduced by Wang [4] employing the Radon transform of the ambiguity function (2D FFT of WVD), well known as RAT. RAT limits the line integral to lines passing through the origin hence losing information about the initial frequency parameter. A sequential estimation procedure is proposed by Zhao [5] which employs RAT to estimate chirp rate and Fractional Fourier transform (FRFT) [6, 7] to estimate amplitude and initial frequency. But this method is restricted to constant amplitude chirps only. In fact conventional FRFT based method is useful for mono-component constant amplitude chirp signals. But with multi-component signals and high amplitude difference between components, make it extremely difficult to estimate parameters accurately. Moreover, with those methods the signal has to be investigated over the full range of $[0, 2\pi]$. Methods based on wavelet [8, 9] and Discrete Polynomial Phase Transform (DPT) are also reported [10]. Some recent works are Discrete chirp Fourier transform for chirp rate estimation [11], and slope calculation of ambiguity function for parameter estimation [1] etc.

In this paper a new sequential estimation method for linear amplitude modulated chirp signals are proposed. The method is based on RWT and Discrete Fractional Fourier transforms (DFRFT). Although the RWT has some extra computational cost but considering the total numbers of parameters to be estimated (the parameters are: chirp rate, initial frequency, constant amplitude, amplitude slope or modulation parameter and initial phase for each component signal) compared to other single parameter (most cases) or multi parameter (at most three parameters) estimation methods and advances in digital computers make it viable. The Radon Transform's Ability to detect lines even in extreme noise condition ensures highly accurate estimation of chirp rate and initial frequency. Thus the inherent chirp localization property of DFRFT can be effectively used to estimate amplitude

parameters even if the peaks are buried by noise or other amplitude dominant components. Moreover, the DFRFT can be applied efficiently using Fast Fourier Transform (FFT) Algorithm.

The paper is organized as follows. In section II the signal model along with preliminaries of Wigner Distribution, Radon Transform and Discrete Fractional Fourier Transform is presented. In section III the proposed estimation method is discussed. In section IV simulation results are given. Performance analysis and discussions are presented in section V. Finally the paper is concluded in section VI with direction for future research.

II. Preliminaries

A. Signal Model

In this paper the signals are modeled as following.

$$x(t) = \sum_{i=1}^{M} s_i(t) + \mu(t) \quad (1)$$

where $s_i(t)$ is an amplitude modulated chirp signal defined as

$$s_i(t) = (a_i t + b_i) e^{j(\frac{1}{2}c_i t^2 + \gamma_i t + \phi_i)}$$

Here, $\mu(t)$ is zero mean Gaussian noise whose variance is known to be σ_μ. a_i and b_i are the amplitude parameters and c_i, γ_i, ϕ_i are chirp rate, initial frequency and initial phase respectively.

B. Wigner-Ville Distribution

The Wigner-Ville distribution of a signal $x(t)$ is defines as

$$W[x(t)](t,\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t+\frac{\tau}{2}) x^*(t-\frac{\tau}{2}) e^{-j\omega\tau} \, d\tau \quad (2)$$

where * denotes complex conjugate operation. It represents the time-frequency (t,ω) distribution of a signal. The time-frequency analysis of a linear chirp is a straight line with slope equal to chirp rate.

The Wigner distribution follows the following properties

$$|x(t)|^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} W[x(t)](t,\omega) \, d\omega$$

$$|X(\omega)|^2 = \int_{-\infty}^{\infty} W[x(t)](t,\omega) \, dt$$

$$\|x(t)\|^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W[x(t)](t,\omega) \, d\omega \, dt \quad (3)$$

where $|X(\omega)|$ is the frequency spectrum of $x(t)$ and $\|x(t)\|^2$ denotes energy of the signal.

C. Radon Transform

The Radon transform [Wood 14] of a 2D signal $f(x,y)$ is defined as

$$R_{\rho,\theta}[f(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\delta(\rho - x\cos\theta - y\sin\theta) dx dy \quad (4)$$

where $\rho \in [-\infty, \infty]$ and $\theta \in [0, 2\pi]$. The argument inside the delta function indicates that the integral is taken along straight lines with parameters θ and ρ.

According to equation (4) the Radon transform of Wigner Distribution takes the form

$$R_{\rho,\theta}[W(t,\omega)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(t,\omega)\delta(\rho - t\cos\theta - \omega\sin\theta) dt d\omega$$

$$= \int_{-\infty}^{\infty} W(t, \rho\cos ec\theta - t\cot\theta) dt \quad (5)$$

In equation (5) $\cot\theta$ represents the slope of the lines, which is also the sweep rate of a chirp on time-frequency plane and $\rho \cos ec\theta$ represents the initial frequency of the chirp.

D. Fractional Fourier Transform

The Fractional Fourier transform of $x(t)$ is defined as

$$F^{\alpha}[x(t)] = X_{\alpha}(u) = \int_{-\infty}^{\infty} x(t) k_{\alpha}(t,u) dt \quad (6)$$

where

$$k_\alpha(t,u) = \begin{cases} C_\alpha e^{jat^2 + jau^2 - jbut} & \alpha \neq 0, n\pi \\ \delta(t-u) & \alpha = 0, 4n\pi \\ \delta(t+u) & \alpha = \pi, (4n+2)\pi \end{cases}$$

with $C_\alpha = \sqrt{\frac{1 - j\cot\alpha}{2\pi}}$, $a = \frac{1}{2}\cot\alpha$ and $b = \frac{1}{2}\csc\alpha$, n is positive or negative integer.

An interesting property of the radon transform is that equation (5) replacing $\theta = \alpha$ is same as the squared magnitude spectra of $x(t)$ in the Fractional domain, that means,

$$|X_\alpha(u)|^2 = \int_{-\infty}^{\infty} W(t, \rho\sec\alpha - t\tan\alpha) \quad (7)$$

Equation (7) is the direct consequence of Wigner rotation property of Fractional Fourier Transform and the main essence of the chirp rate and initial frequency estimation procedure.

However, the Discrete Fractional Fourier transform of a discrete time signal $x(nT)$ is defined as (for $\alpha \neq 0, n\pi$) [15]

$$X(kU) = e^{j\frac{\cot\alpha}{2}k^2U^2} \sum_{n=-N}^{N} x(nT) e^{j\frac{\cot\alpha}{2}n^2T^2} e^{-j\frac{2\pi}{2N+1}kn}, \quad -N \le k \le N \quad (8)$$

where α is the transform order, T and U are the time and fractional domain sampling rate respectively and n and k denotes time and fractional domain sample numbers respectively. T and U must satisfy the following relation

$$TU = \frac{2\pi \sin\alpha}{2N+1} \quad (9)$$

DFRFT of a chirp signal with chirp rate c_i is a delta if the transform order is $\alpha = -\cot^{-1} c_i$. In the following section it will be shown how this delta peak is used to estimate amplitude parameters.

III. Proposed Estimation Method

A. Chirp Rate and Initial Frequency Estimation Using Radon Wigner Transform (RWT)

A component signal s_i can be written as

$$s_i(t) = s_i^m(t)\, s_i^c(t) \quad (10)$$

where, $s_i^m(t)$ is the modulated term and $s_i^c(t)$ is the chirping term. The auto-term of Wigner Distribution $W_{ii}(t,\omega)$ is the interest for parameter estimation. The cross term between chirps $S_i(t)$ and $S_k(t)$ ($i \neq k$) is not useful for parameter estimation and hence not shown. However, the WD of the modulated-term will be,

$$W_{ii}^m(t,\omega) = \int_{-\infty}^{+\infty} s_i^m(t+\frac{\tau}{2}) s_i^{m*}(t-\frac{\tau}{2}) e^{-j\omega\tau}\, d\tau$$

$$= (a_i t + b_i)^2 \delta(\omega) + \frac{a_i^2}{4}\delta''(\omega)$$

WD of the chirping term will be,

$$W_{ii}^c(t,\omega) = \int_{-\infty}^{+\infty} s_i^c(t+\frac{\tau}{2}) s_i^{c*}(t-\frac{\tau}{2}) e^{-j\omega\tau}\, d\tau$$

$$= \delta(\omega - c_i t - \gamma_i)$$

Since the two signals are convolved along ω direction then their WD also is convolved along the ω direction;

$$W_{ii}(t,\omega) = \int_{-\infty}^{+\infty} W_{ii}^m(t,\omega') W_{ii}^c(t,\omega-\omega')d\omega'$$

$$= (a_i t + b_i)^2 \int_{-\infty}^{+\infty} \delta(\omega')\delta(\omega' - \omega + c_i t + \gamma_i)d\omega'$$

$$+ \frac{a_i^2}{4}\int_{-\infty}^{+\infty} \delta''(\omega')\delta(\omega' - \omega + c_i t + \gamma_i)d\omega'$$

$$= (a_i t + b_i)^2 \delta(\omega - c_i t - \gamma_i) + \frac{a_i^2}{4}\delta''(\omega - c_i t - \gamma_i) \quad (11)$$

The cross-terms for two components with different parameters are not important for estimation and therefore skipped to save space.

Now we have to take the Radon transform of the auto-term

$$\Re_{ii}(\rho,\theta) = \int_{-\infty-\infty}^{+\infty+\infty} |W_{ii}(t,\omega)| \delta(\rho - t\cos\theta - \omega\sin\theta) d\omega dt$$

$$= \frac{1}{|\sin\theta|}\int_{-\infty}^{+\infty} |W_{ii}(t, -t\cot\theta + \rho\csc\theta)| dt$$

$$= \frac{1}{|\sin\theta|}\int_{-\infty}^{+\infty} \left| (a_i t + b_i)^2 \delta(\omega - c_i t - \gamma_i) + \frac{a_i^2}{4}\delta''(\omega - c_i t - \gamma_i) \right|_{\omega = -t\cot\theta + \rho\csc\theta}$$

$$= \frac{\csc\theta}{|\cos\theta + c_i \sin\theta|}\{a_i\left(\frac{\rho - \gamma_i \sin\theta}{\cot\theta + c_i}\right) + b_i\} \quad (12)$$

Equation (12) indicates that RWT term will have localized peak at $\theta = -\cot^{-1} c_i$, $\rho = \gamma \cos ec\theta$. The integral terms along other lines will be much smaller than this and therefore not shown here.

B. Amplitude and Phase Parameters Estimation Using Discrete Fractional Fourier Transform (DFRFT)

The component signal in equation (1) can be sampled with a sampling rate $T = \frac{2\pi}{2N+1}$. Thus,

$$S_i(nT) = (a_i nT + b_i) e^{j(\frac{1}{2}c_i n^2 T^2 + \gamma_i nT + \phi_i)} \quad (13)$$

Replacing (13) in (8) with $\alpha = -\cot^{-1} c_i$ (as we already know this optimal value from RWT)

$$S_i^{\alpha}(kU)=e^{j\frac{c_i}{2}k^2U^2}\sum_{n=-N}^{N}(a_inT+b_i)\,e^{j\phi_i}e^{-j\frac{2\pi}{2N+1}(k-\gamma)n},\quad -N\le k\le N$$

$$=e^{j\frac{c_i}{2}k^2U^2}\,e^{j\phi_i}[a_i\sum_{n=-N}^{N}ne^{-j\frac{2\pi}{2N+1}(k-\gamma)n}+b_i\sum_{n=-N}^{N}e^{-j\frac{2\pi}{2N+1}(k-\gamma)n}] \tag{14}$$

Now at $k=\gamma$

$$S_i^{\alpha}(kU)|_{k=\gamma}=e^{j\frac{c_i}{2}k^2U^2}\,e^{j\phi_i}(2N+1)b_i$$

And at $k=\gamma\pm1$

$$S_i^{\alpha}(kU)|_{k=\gamma\pm1}=\frac{4\pi j}{2N+1}e^{j\phi_i}e^{j\frac{c_i}{2}(\gamma\pm1)^2U^2}\,a_i\sum_{n=1}^{N}n\sin(\frac{2\pi}{2N+1}n)$$

So from the magnitude spectrum amplitude parameters can be estimated as,

$$b_i=\frac{\left|S_i^{\alpha}(kU)\right|_{k=\gamma}}{(2N+1)}\qquad a_i=\frac{(2N+1)\left|S_i^{\alpha}(kU)\right|_{k=\gamma\pm1}}{4\pi\sum_{n=1}^{N}n\sin(\frac{2\pi}{2N+1}n)} \tag{15}$$

Now as the other four parameters are already known, the fifth parameter initial phase ϕ_i can be estimated by injecting a chirp with zero initial phase with other estimated parameters and using the relation below

$$\cos\frac{\phi_i}{2}=\frac{\left|\hat{S}_i^{\alpha}(kU)\right|_{k=\gamma}}{\left|S_i^{\alpha}(kU)\right|_{k=\gamma}} \tag{16}$$

where $\left|\hat{S}_i^{\alpha}(kU)\right|$ is the magnitude spectra of the chirp injected signal.

Equation (15) and (16) can be used to approximate chirp parameters with a very small bias even when multiple components are present

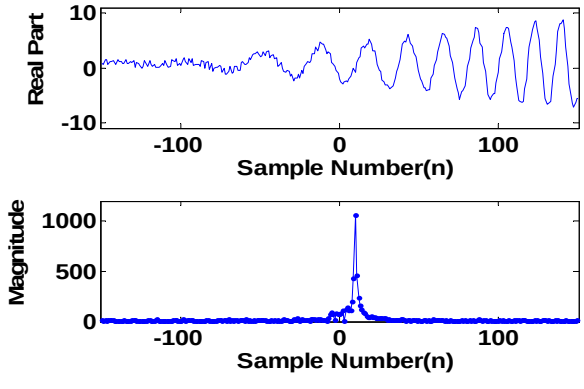


Fig. 1 (Top) Real plot of a noisy chirp signal and (Bottom) DFRFT of the chirp at the appropriate angle. At this optimal angle the magnitude spectra shows delta like peak.

IV. Simulation Results

The proposed estimation method is investigated with digital simulation. The test signal consists of 4 amplitude modulated chirp signals. The parameters are:

$$c_1=0.5,\gamma_1=50,a_1=0.7,b_1=1.89,\phi_1=0.91$$
$$c_2=-1.6,\gamma_2=-60,a_2=1.23,b_2=3.1,\phi_2=1.8$$
$$c_3=0.4,\gamma_3=30,a_3=0.5,b_3=7.43,\phi_3=1.03$$
$$c_4=0.9,\gamma_4=10,a_4=3.5,b_4=3.5,\phi_4=2.37\,.$$

Number of samples is $2N+1=301$. Seven different observation noise levels were considered. Fig. 2 shows the 2D RWT result (γ vs. θ) for estimation of c and γ at SNR=-5 dB. The peaks are clearly visible at (-63.5°,50), (31.5°,-60), (-68°,30) and (-48°,10).

Estimation results of other three parameters for different noise levels are shown in Table 1. Fig. 3 presents the actual and estimated composite signal on same plot for SNR=-5 dB.

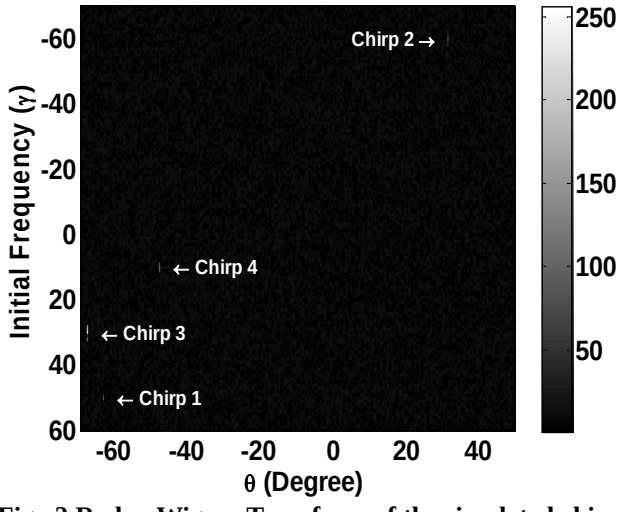


Fig. 2 Radon Wigner Transform of the simulated chirp signal. The maximum peaks are clearly visible in the figure.

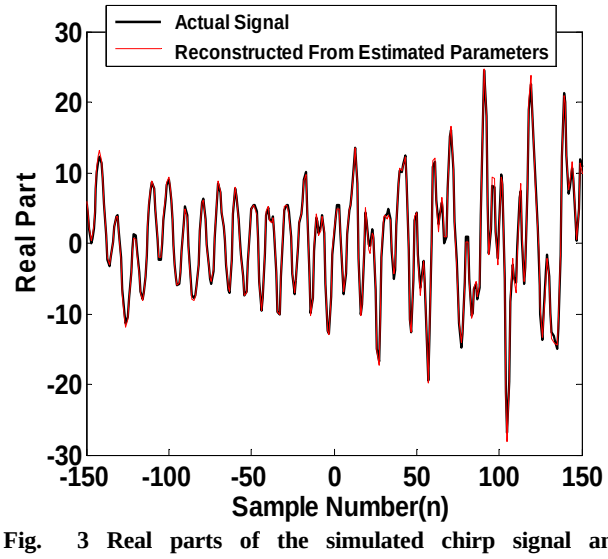


Fig. 3 Real parts of the simulated chirp signal and reconstructed chirp signal from the estimated parameters on the same plot.

Table 1 Values of estimated parameters at different noise levels

Para-meters	Actual Values	Estimated Values in Different Noise Levels (dB)						
		-5	-3	-1	0	1	3	5
a_1	0.7	0.854	0.822	0.795	0.784	0.774	0.758	0.745
b_1	1.89	2.126	2.047	2.033	2.016	2.000	1.975	1.954
ϕ_1	0.91	0.823	0.830	0.835	0.838	0.840	0.843	0.846
a_2	1.23	1.293	1.280	1.268	1.268	1.260	1.253	1.249
b_2	3.1	3.272	3.248	3.228	3.220	3.214	3.202	3.193
ϕ_2	1.8	1.732	1.743	1.753	1.756	1.760	1.766	1.771
a_3	0.5	0.412	0.428	0.443	0.450	0.456	0.466	0.475
b_3	7.43	7.238	7.250	7.260	7.265	7.269	7.275	7.280
ϕ_3	1.03	0.998	1.003	1.007	1.009	1.010	1.013	1.015
a_4	3.5	3.680	3.642	3.612	3.599	3.589	3.570	3.555
b_4	3.5	3.478	3.477	3.477	3.476	3.476	3.476	3.476
ϕ_4	2.37	2.359	2.371	2.379	2.380	2.380	2.381	2.381

V. Performance Analysis and Discussion

To evaluate the performance of the system under noise, Monte-Carlo simulations are performed. A mono component chirp signal with parameters $c_1 = 2.3, \gamma_1 = -17, a_1 = 1.13, b_1 = 3.5, \phi_1 = \pi$ is generated. The estimation values are computed 100 times for each of a_i, b_i, ϕ_i, for each case of SNR=-10 dB,-8 dB, -6dB,-4dB,-2 dB, 0dB, 2 dB, 4dB, 6 dB, 8 dB and 10 dB. The accuracy of estimation is measured as Mean Squared Error (MSE) as,

$$MSE = 10 \log_{10} [\frac{1}{M} \sum_{i=1}^{M} (\hat{p}_i - p)^2] \quad dB \qquad (17)$$

where $\hat{p}_i$ denote the estimated value of parameter p at the i_{th} computation and M is the number of Monte-Carlo computation. Fig. 4 shows the MSE of estimated parameters for various SNR.

As shown in Fig. 4, the MSE ranges between -30 to -50 dB for amplitude parameters. For the initial phase parameter the MSE is around -25 dB. The performance of the proposed method is similar to other constant amplitude chirp parameter estimation like [4, 5] but it estimates more parameters. Moreover, this method has better accuracy than the method proposed in [1].

VI. Conclusions

A sequential estimation method combining Radon Wigner transform (RWT) and Discrete Fractional Fourier transform (DFRFT) is proposed. The RWT is capable of estimating chirp rate and initial frequency in high noise condition. The DFRFT enables to compute the amplitude parameters of amplitude modulated signals, which are important for speech modeling, along with the initial phase. Most importantly the proposed method is well suited for applications involving multiple-parameter estimation of chirp signals. However, effectiveness of this proposed method for real world application like speech modeling and synthesis is yet to be performed and therefore attributed to future research.

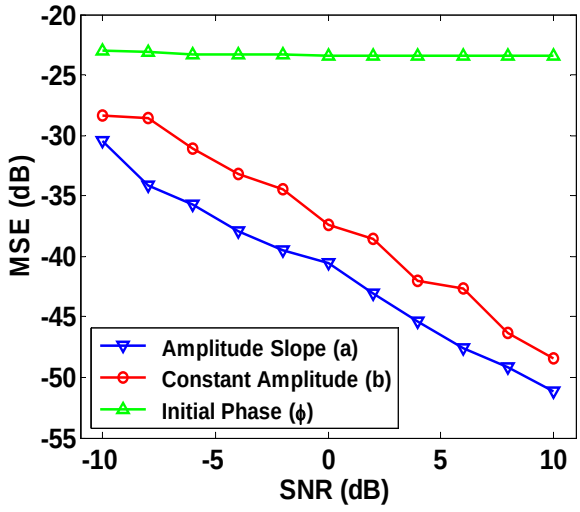


Fig. 4 MSE vs. SNR plot for parameter estimation of the simulated mono-component chirp signal.

References

[1] Y. Huang, and R.D. Dony, "Speech modelling by non-stationary partials with time varying amplitude and frequency," Electrical and Computer Engineering, 2004. Canadian Conference on, vol.3, pp. 1273-1276, 2-5 May 2004.

[2]. S. Barbarossa, and A. Farina, "Detection and imaging of moving objects with synthetic aperture radar. 2. Joint time-frequency analysis by Wigner-Ville distribution," Radar and Signal Processing, IEE Proceedings, vol.139, no.1, pp.89-97, Feb 1992.

[3] J.C. Wood, D.T. Barry, "Radon transformation of time-frequency distributions for analysis of multicomponent signals," Signal Processing, IEEE Transactions on, vol.42, no.11, pp.3166-3177, Nov 1994.

[4] M. Wang; A.K. Chan, and C.K. Chui, "Linear frequency-modulated signal detection using Radon-ambiguity transform," Signal Processing, IEEE Transactions on, vol.46, no.3, pp.571-586, Mar 1998.

[5] Z. Xinghao, T. Ran; Z. Siyong, "A novel sequential estimation algorithm for chirp signal parameters," Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on, vol.1, pp. 628-631, 14-17 Dec. 2003.

[6] L.B. Almeida,"The fractional Fourier transform and time-frequency representations ," Signal Processing, IEEE Transactions on, vol.42, no.11, pp.3084-3091, Nov 1994.

[7] H.M. Ozaktas, O. Arikan, M.A. Kutay, and G. Bozdagt, "Digital computation of the fractional Fourier transform," Signal Processing, IEEE Transactions on, vol.44, no.9, pp.2141-2150, Sep 1996.

[8] M.M. Wang, A.K. Chan, C.K. Chui, "Linear frequency modulated signal detection using wavelet packet, ambiguity function and Radon transform," Antennas and Propagation Society International Symposium, 1995. AP-S. Digest , vol.1, pp.308-311, 18-23 Jun 1995.

[9] Z. Huafeng; Z. Jianping; H.J. Guo, "Instantaneous parameter estimation based on continuous wavelet transform and some improvements," Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, vol.4, pp.2169-2172, 12-15 May 1998.

[10]S. Peleg, and B. Friedlander, "The discrete polynomial-phase transform," Signal Processing, IEEE Transactions on, vol.43, no.8, pp.1901-1914, Aug 1995.

[11]Xiang-Gen Xia, "Discrete chirp-Fourier transform and its application to chirp rate estimation," Signal Processing, IEEE Transactions on, vol.48, no.11, pp.3122-3133, Nov 2000.

[12]J.C. Wood, D.T. Barry, "Tomographic time-frequency analysis and its application toward time-varying filtering and adaptive kernel design for multicomponent linear-FM signals," Signal Processing, IEEE Transactions on, vol.42, no.8, pp.2094-2104, Aug 1994.

[13]T. Erseghe, P. Kraniauskas, and G. Carioraro, "Unified fractional Fourier transform and sampling theorem," Signal Processing, IEEE Transactions on, vol.47, no.12, pp.3419-3423, Dec 1999.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A new perceptual post filter for single channel speech enhancement

[1]Md. Jahangir Alam, [2]Douglas O'Shaughnessy, and [3]Sid-Ahmed Selouani

[1,2]INRS-EMT, University of Quebec, Montreal QC H5A 1K6 Canada
[3]University of Moncton, Shippagan Campus, NB, Canada

E-mail: {alam, dougo}@emt.inrs.ca, selouani@umcs.ca

Abstract – The major drawback of the most speech enhancement methods in speech applications is the generation of an annoying residual noise with musical character. A potential solution to this artifact is the incorporation of a psychoacoustic model in the suppression filter design. In this paper a frequency domain optimal linear estimator with perceptual post filtering is proposed, which incorporates the masking properties of the human hearing system to make the residual noise distortion inaudible. The performance of the proposed enhancement algorithm is evaluated by the Segmental SNR, and Perceptual Evaluation of Speech Quality (PESQ) measures under various noisy environments and yields better results compared to the Wiener denoising technique.

Index Terms: speech enhancement, perceptual post filter, musical critical band, MMSE, modified masking threshold

I. Introduction

The performance of speech communication systems in applications such as hands-free telephony, degrade considerably in adverse acoustic environments. The presence of noise can cause loss of intelligibility as well as the listener's discomfort and fatigue. Speech enhancement methods seek to improve the performance of these systems and to make the corrupted speech more pleasant to the listener. These methods are also useful in other applications such as automatic speech recognition.

The removal of additive noise from speech has been an active area of research for several decades. Numerous methods have been proposed by the signal processing community. Among the most successful signal enhancement techniques have been spectral subtraction [1, 2], Wiener filtering [3, 4] and signal subspace methods [16]. Although these techniques improve speech quality, they suffer from the annoying residual noise known as musical noise. Tones at random frequencies, resulting from poor estimation of the signal and noise statistics, are at the origin of this artifact. The quality and the intelligibility of the enhanced speech signal could be improved by reducing or in better cases eliminating this kind of musical residual noise.

Many variations have been developed to cope with the musical residual noise phenomena including spectral subtraction techniques based on masking properties of the human hearing system. A number of methods have been developed to improve intelligibility by modeling several aspects of the enhancement function present in the auditory system [5]–[7]. These attractive methods use a noise masking threshold (NMT) as a crucial parameter to empirically adjust either thresholds or gain factors. This auditory system is based on the fact that the human ear cannot perceive additive noise when the noise level falls below the NMT.

In this paper, we have only considered simultaneous masking, i.e., a weak signal is made inaudible by a strong signal occurring simultaneously. This phenomenon is modeled via a noise-masking threshold, below which all spectral components are inaudible. The masking-based speech enhancement approach basically incorporates the noise masking properties into a speech enhancement algorithm. Since the human ear builds critical bands around each frequency and behaves like a band-pass filter [8], in this paper, we have developed a post processing method with a modified masking threshold to reduce the musical residual noise in each critical band generated by classical speech enhancement methods. The tonality coefficient in each critical band is utilized to characterize the residual musical noise. Experimental results show that the proposed method outperforms the Wiener denoising method based on decision-directed a priori SNR and two step a priori SNR estimators.

This paper is organized as follows: section 2 provides a description of the baseline speech enhancement system. In section 3, descriptions of the a priori SNR estimation, noise estimation and proposed method are given. A discussion on the experimental results and a conclusion are drawn in section 4 and section 5, respectively.

II. Baseline speech enhancement method

Let the distorted signal be expressed as

$$y(n) = x(n) + d(n), \quad (1)$$

where $x(n)$ is the clean signal and $d(n)$ is the additive random noise signal, uncorrelated with the original signal. Taking FFT to the observed signal gives

$$Y(m,k) = X(m,k) + D(m,k). \quad (2)$$

Where $m = 1,2,...,M$ is the frame index, $k = 1,2,...,K$ is the frequency bin index, M is the total number of frame

and K is the frame length, $Y(m,k)$, $X(m,k)$ and $D(m,k)$ represent the short-time spectral component of $y(n)$, $x(n)$ and $d(n)$, respectively. With the assumption that different spectral components on index k are statistically independent, an estimate $\widehat{X}(m,k)$ of the clean speech component is given by

$$\widehat{X}(m,k) = H(m,k)Y(m,k), \quad (3)$$

where $H(m,k)$ is the noise suppression gain (denoising filter). In general, it can be expressed as a function of the a posteriori SNR and a priori SNR given by

$$\gamma(m,k) = \frac{|Y(m,k)|^2}{\Gamma_d(m,k)}, \quad (4)$$

$$\xi(m,k) = \frac{\Gamma_x(m,k)}{\Gamma_d(m,k)}, \quad (5)$$

where $\Gamma_d(m,k) = E\{|D(m,k)|^2\}$, by definition, is the noise power spectrum, an estimate of which can be made easily during speech pauses and $\Gamma_x(m,k) = E\{|X(m,k)|^2\}$.

The instantaneous SNR can be defined as

$$\vartheta(m,k) = \gamma(m,k) - 1. \quad (6)$$

An estimate $\hat{\xi}(m,k)$ of $\xi(m,k)$ is given by the well known decision-directed approach [9] and is expressed as

$$\hat{\xi}(m,k) = \max\left(\alpha\frac{|H(m-1,k)Y(m-1,k)|^2}{\Gamma_d(m,k)} + (1-\alpha)P'[\vartheta(m,k)], \xi_{\min}\right), \quad (7)$$

where $P'[x] = x$ if $x \geq 0$ and $P'[x] = 0$ otherwise. In this paper we have chosen $\alpha = 0.98$ and $\xi_{\min} = 0.0032$ (i.e., -25 dB) by the simulations and informal listening tests.

Several variants of the noise suppression gain $H(m,k)$ have been reported in the literature, here, without loss of generality, the gain function is chosen as the Wiener filter similar to [13]

$$H(m,k) = \frac{\xi(m,k)}{1+\xi(m,k)}. \quad (8)$$

The temporal-domain denoised speech is obtained with the following relation

$$\hat{x}(n) = IFFT(|\widehat{X}(m,k)|.e^{j\arg(Y(m,k))}). \quad (9)$$

Since the human ear is fundamentally phase deaf and perceives speech primarily based on the magnitude spectrum, we have used noisy signal phase to obtain temporal-domain denoised speech.

III. Overview of the proposed method

A. Estimation of *a priori* SNR

An important parameter of numerous speech enhancement techniques is the *a priori* SNR. In the well-known decision-directed approach, the *a priori* SNR depends on the speech spectrum estimation in the previous frame [10], which results in degradation of the speech enhancement performance. In order to alleviate this problem while keeping its benefits we have used the MMSE based two-step *a priori* SNR estimation approach proposed in [17] and which is expressed as.

$$\hat{\xi}_{MMSE} = \frac{\hat{\xi}}{1+\hat{\xi}}\left(1+\frac{\hat{\xi}}{1+\hat{\xi}}\gamma\right), \quad (10)$$

where $\hat{\xi}$ and γ are given by the equations (7), and (4), respectively.

B. Noise Estimation

Noise estimation is also an important factor in speech enhancement systems. In this paper the noise power spectrum is estimated during speech pauses using the following recursive relation [12]:

$$\Gamma_d(m,k) = \begin{cases} \lambda_D\Gamma_d(m-1,k) + (1-\lambda_D)W(m,k)|Y(m,k)|^2 & \text{if W(m,k)>0} \\ \Gamma_d(m-1,k) & \text{if W(m,k)=0} \end{cases}, \quad (11)$$

where λ_D is a smoothing factor satisfying $0 < \lambda_D < 1$ and W(m,k) is the weighting factor on the noisy power spectrum. The weighting factor is designed so that it is almost inversely proportional to the estimated SNR (dB) given by

$$\tilde{\gamma}(m,k) = 10\log_{10}\left(\frac{|Y(m,k)|^2}{\Gamma_d(m-1,k)}\right) \quad (12)$$

and the weighting factor is given by the following relation

$$W(m,k) = \begin{cases} 1 & \text{if } \tilde{\gamma}(m,k) \leq 0 \\ -\frac{1}{\tau}\tilde{\gamma}(m,k) + 1 & \text{if } 0 < \tilde{\gamma}(m,k) \leq \varepsilon \\ 0 & \text{if } \tilde{\gamma}(m,k) > \varepsilon \end{cases}, \quad (13)$$

where τ is a slope deciding constant of the graph of (13) and ε is a threshold to eliminate an unreliable $\tilde{\gamma}(m,k)$. In this paper we have used $\lambda_D = 0.9$, $\tau = 12$, and $\varepsilon = 6$ on the basis of simulations.

C. Spectral Gain Calculation

The spectral gain $H(m,k)$ for the Wiener denoising technique is given by (8). The spectral gain $H_r(m,k)$ for the reference signal is computed from $H(m,k)$ as

$$H_r(m,k) = \begin{cases} H(m,k) + \frac{\eta}{q} & \text{if } \left(H(m,k) + \frac{\eta}{q}\right) \leq 1 \\ 1 & \text{otherwise} \end{cases}, \quad (14)$$

where η and q are adjustment constants chosen experimentally. It is the shifting up of the Wiener denoising filter by $\frac{\eta}{q}$. In this paper we have used $\eta = 0.35$ and $3 \leq q \leq 10$. It is assumed that $H_r(m,k)$ introduces minimum distortion and results in a reference signal which does not contain residual musical noise. The reference signal obtained using $H_r(m,k)$ is used instead of the noisy speech signal to improve the accuracy of the musical noise detector [13].

D. Calculation of Tonality Coefficients

In the denoised signal obtained by subtractive methods, annoying musical tones appear in the power spectrum, which leads to an increase of tonality coefficient. Thus it is possible to detect the presence or absence of musical tones by means of a tonality coefficient. The steps for calculating the tonality coefficient are taken from [14] and described below:

I. Frequency analysis of both signals (reference and denoised) along the critical band (CB): The power spectrum of the denoised signal and that of the reference signal are partitioned in critical bands. We have considered CBs between 0 kHz and 4 kHz as we chose the Aurora database, having the sampling frequency of 8 kHz, for this experiment. In this paper we tried to detect CB musical noise for the CBs between 9 and 18 as the annoying musical noise is situated only in the frequency range between 1 kHz and 4 kHz, for frequencies under 1 kHz it is masked by the presence of real tones of clean speech [15].

II. Calculation of the tonality coefficients: The tonality coefficient is measured using the ratio of the geometric mean (GM) and the arithmetic mean (AM) of the signal power spectrum, known as the spectral flatness measure (SFM). SFM is used to determine whether the signal is tone-like or noise-like. The coefficient of tonality is expressed as

$$tc(i) = \min\left(\frac{SFM_{dB}(i)}{-60}, 1\right), \qquad (15)$$

where i is the CB index and $SFM_{dB}(i)$ is given as

$$SFM_{dB}(i) = 10\log_{10}\left(\frac{GM(i)}{AM(i)}\right), \qquad (16)$$

where

$GM(i) = \prod_{j=1}^{I} P(j)^{\frac{1}{M(i)}}$ and $AM(i) = \frac{\sum_{j=1}^{I} P(j)}{M(i)}$, I is the number of critical bands, and $M(i)$ & $P(j)$ denote the number of frequency bins and the power spectral density, respectively in each critical band i. A tonality coefficient of unity indicates that the signal is tone-like and a tonality coefficient close to zero indicates that the signal is noise-like.

Using (15) and (16) the tonality coefficients for the denoised signal $tc_d(i)$ and that of the reference signal $tc_r(i)$ are computed for each CB. Then, the tonality coefficient difference for each CB is given by

$$\Delta tc(i) = tc_d(i) - tc_r(i). \qquad (17)$$

Musical residual noise appears in the ith CB if $tc_d(i) > tc_r(i)$ and it becomes audible if $\Delta tc(i) > \mathrm{T}'(i)$, where $\mathrm{T}'(i)$ is the threshold for the ith CB, which depends on the order of CB and masking properties of the human ear. For the calculation of the threshold, we used a narrow band noise and a sinusoidal signal and computed tonality coefficients for each the signals. The differences between the tonality coefficients of the both signals have been taken as the thresholds $\mathrm{T}'(i)$. The threshold, $\mathrm{T}'(i)$ for all CB is found to be approximately constant and is $\mathrm{T}'(i) = 0.06$ [13].

E. Modified Relative threshold offset and noise-masking threshold computation

The noise-masking threshold is obtained through modeling the frequency selectivity of the human auditory system and its masking properties [14]. The steps for calculating the modified relative threshold offset and the modified masking threshold are taken from [14] and are as follows:

I. Partition of the signal power spectrum into CBs and the energies, $E(i)$ in each CB are added.

II. Calculation of the spread CB spectrum, $C(i)$ by convolving the spread function $SF(i)$ and the bark spectrum $E(i)$ in order to take into account the masking effect between different CBs.

III. In the Johnston model [11], an offset $O(i)$ is determined according to the tonality coefficient and the CB order as

$$O(i) = tc(i)\left(14.5 + i\right) + (1 - tc(i))5.5 \quad \text{dB}. \qquad (18)$$

The total tonality coefficient used by Johnston gives a general idea about the nature of the power spectrum. In our context, we seek to detect musical noise in the selected CBs, i.e., between 9 and 18. The CB tonality coefficient of the denoised signal was taken into account when calculating the threshold offset $O(i)$ in the proposed method.

In order to calculate the modified relative threshold offset, a Boolean flag $M(i)$ is constructed first based on the tonality coefficient difference $\Delta tc(i)$, and the threshold $\mathrm{T}'(i)$, over which an additive tone becomes audible in the presence of narrow-band noise. $\Delta tc(i)$ and $\mathrm{T}'(i)$ were determined in section 3.4.1. The Boolean flag $M(i)$ indicates the presence or absence of CB musical noise and is given as

$$M(i) = \begin{cases} 1 & \text{if } \Delta tc(i) \geq \mathrm{T}'(i) \\ 0 & \text{otherwise} \end{cases}. \qquad (19)$$

In case of critical-band musical noise ($M(i) = 1$), the used tonality coefficient to calculate $O(i)$ is close to one. However, it shouldn't be for better estimation of masking threshold. For better estimation we proposed to correct the tonality coefficient of the ith musical critical band by replacing it with the tonality coefficient of the ith critical band tonality coefficient of the reference signal. Thus the corrected offset threshold for the modified Johnston model becomes

$$O_{M'}(i) = tc_m(i)\left(14.5 + i\right) + (1 - tc_c(i))5.5 \quad \text{dB}, \qquad (20)$$

where

$$tc_m(i) = \begin{cases} tc_d(i), & \text{for } M(i) = 1 \\ tc_r(i), & \text{for } M(i) = 0 \end{cases}. \qquad (21)$$

Since we only have the noisy speech signal, the preliminary estimate of the speech signal is usually not very accurate. Consequently, the estimation of the relative threshold offsets obtained using (18) and/or (20) are also

affected. This is because the residual noise of the preliminary estimated speech may severely change the original tonality of the speech signal. As a result of the inaccurate relative threshold offset, the estimation error of the noise masking threshold increases. In [7], the fixed relative threshold offset is exploited and compensated with a slight modification by taking into account the tone-like nature of the musical noise for the CB $i > 15$. In this paper, we propose a modified relative threshold offset by merging both the fixed relative threshold offset and the modified relative threshold offset (i.e., (20)) to achieve a more effective application of the masking properties [18]. The final modified relative threshold offset, $O_M(i)$, is expressed as

$$O_M(i) = \beta O_{M'}(i) + (1-\beta) O_F(i), \qquad (22)$$

where β is a weighting constant, $O_F(i)$ is the fixed relative threshold offset in each CB as shown in figure 5. In the paper we have used $\beta = .95$, on the basis of simulations.

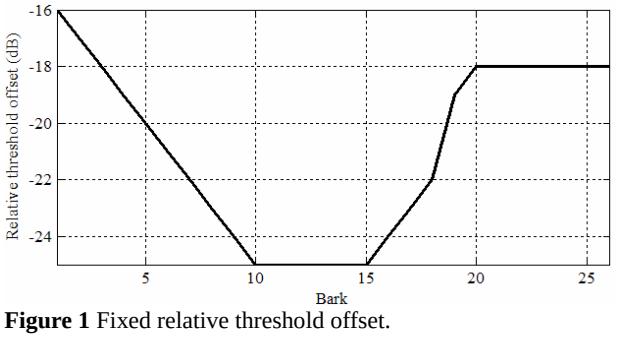


Figure 1 Fixed relative threshold offset.

The modified relative threshold offset is then subtracted from the spread CB spectrum to yield the spread threshold estimate $T(i)$

$$T(i) = 10^{\left[\log_{10}(C(i)) - (O_M(i)/10)\right]}. \qquad (23)$$

IV. In the final step, the modified noise-masking threshold (NMT) is estimated as

$$NMT(i) = \max\left(T_q(i), T(i)\right), \qquad (24)$$

where $T_q(i)$ is the absolute threshold of hearing [14].

IV. Experimental Results and Discussion

In order to evaluate the performance of the proposed perceptual Wiener denoising technique, we conducted extensive objective quality tests under various noisy environments. The frame sizes were chosen to be 256 samples (32 msec) long with 40% overlap; a sampling frequency of 8 kHz and a Hamming window were applied. To evaluate and compare the performance of the proposed perceptual Wiener denoising technique, we carried out simulations with the *TEST A* database of the Aurora [21]. Speech signals were degraded with three types of noise at global SNR levels of -5 dB, 0 dB, 5 dB, 10 dB and 15 dB. The noises were N1 (Subway noise), N3 (Car Noise), and WGN (White Gaussian Noise). Figure 2, Figure 3, and Figure 4 show the average segmental SNR[1], and Perceptual Evaluation of Speech Quality ([2]PESQ) scores [19, 20] for the enhanced speech signals of the Wiener denoising technique and proposed method when speech signals are degraded with the car noise, subway noise and WGN, respectively. It is observed that the proposed approach yields better segmental SNR than that of the Wiener denoising technique under all tested noisy environments. In the case of the PESQ measure, the proposed perceptual Wiener denoising technique gives better PESQ scores than the Wiener denoising technique.

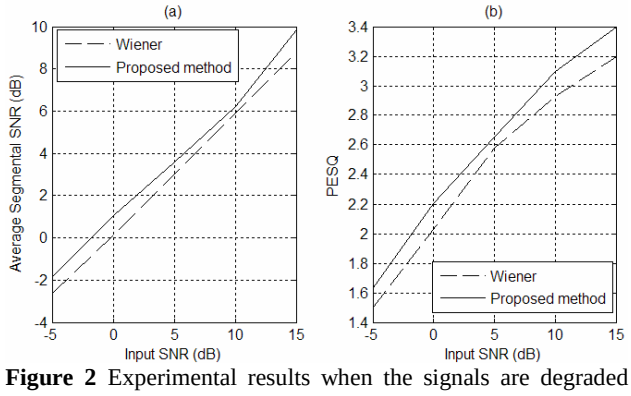


Figure 2 Experimental results when the signals are degraded with car noise.

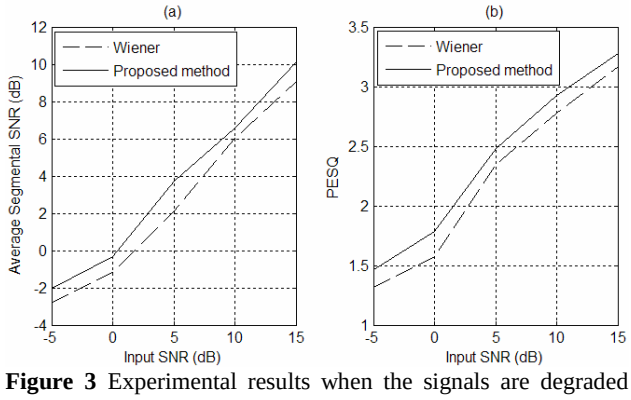


Figure 3 Experimental results when the signals are degraded with subway noise.

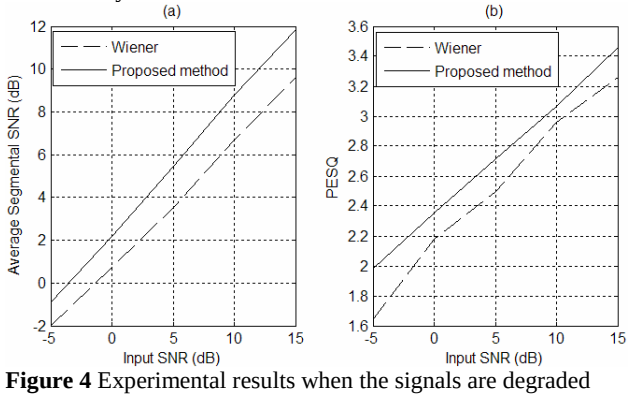


Figure 4 Experimental results when the signals are degraded with white gaussian noise.

Figure 5 represents the spectrograms of the clean speech signal, noisy signal and enhanced speech signals obtained using the Wiener denoising technique and the proposed technique. The speech spectrograms provide more accurate information about the residual noise and speech

[1] The higher value of the segmental SNR indicates the weaker speech distortions.

[2] The higher PESQ score indicates better perceived quality.

distortion than the corresponding time domain waveforms. We compared the spectrograms for each of the methods and confirmed a reduction of the residual noise and speech distortion. Speech spectrograms presented in Figure 5 use a Hamming window of 256 samples with 50% overlap and the noisy signals include N3 (Car Noise) with SNR = 0 dB. It is seen that the musical noise is almost removed for the most part in figure 3(d).

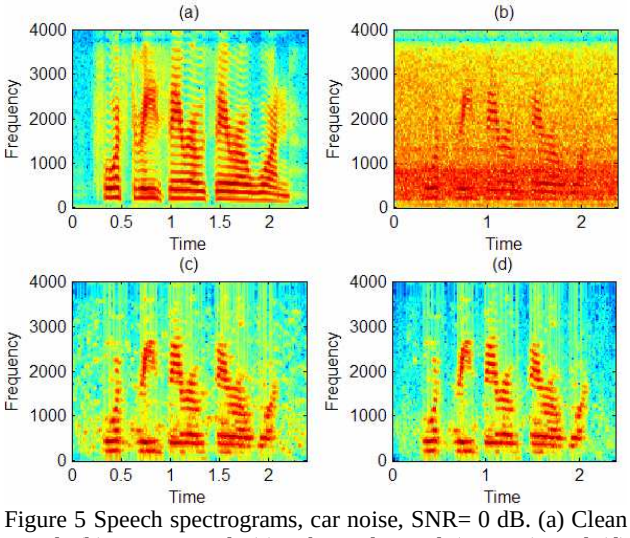


Figure 5 Speech spectrograms, car noise, SNR= 0 dB. (a) Clean signal, (b) noisy signal, (c) enhanced signal (Wiener), and (d) enhanced signal (proposed method).

V. Conclusion

The aim of our study is to develop a perceptual post filter for subtractive type speech enhancement algorithm that would maximize noise reduction while minimizing speech distortion. In order to improve speech enhancement performance by detecting and eliminating musical phenomena in the denoised signal using Wiener filtering, a new perceptual Wiener denoising approach is proposed. This method has the advantage to be applied as a post-processing for any subtractive denoising technique. The proposed approach is based on the detection of musical critical bands using the tonality coefficient and a modified Johnston masking threshold. Experimental results, plotted spectrograms and informal listening tests show that proposed technique performs better in all tested objective quality measures; it does not introduce additional speech distortion, and results in significant reduction of the musical phenomenon.

References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoustics, Speech, Signal Processing, vol. 27, pp. 113–120, Apr. 1979.

[2] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing, vol. 1, (Washington, DC), pp. 208–211, Apr. 1979.

[3] H. L. V. Trees, Detection, Estimation, and Modulation: Part I - Detection, Estimation and Linear Modulation Theory. John Wiley and Sons, Inc., 1st ed., 1968.

[4] T.F. Quatieri and R.B. Dunn, "Speech enhancement based on auditory spectral change," IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol. 1, pp. 257-260, Orlando, FL, USA, 2002.

[5] Y.M. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," IEEE Trans. Signal Processing, vol.39, no.9, pp.1943–1954, 1991.

[6] D. Tsoukalas, M. Paraskevas, and J. Mourjopoulos, "Speech enhancement using psychoacoustic criteria," IEEE ICASSP, pp.359–362, Minneapolis, MN, 1993.

[7] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech Audio Processing, vol.7, no.2, pp.126–137, 1999.

[8] E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models. Springer-Verlag, 2nd ed., 1999.

[9] Y. Ephraim and D. Mallah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

[10] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," IEEE Trans. Speech and Audio Processing, vol. 2, no. 1, pp. 345–349, April 1994.

[11] A. J. Accardi and R. V. Cox, "A modular approach to speech enhancement with an application to speech coding," in Proc. IEEE Int. Conf. Acoustic., Speech, Signal Processing, 1999.

[12] M. Kato, A. Sugiyama and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA," IEICE Trans. Fundamentals, vol. E85-A, no. 7, pp. 1710-1718, July 2002.

[13] Ben Aicha, S. Ben Jebara, "Perceptual musical noise reduction using critical bands tonality coefficients and masking thresholds," *INTERSPEECH conference*, Anvers, Belgium, August 2007.

[14] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. on Selected Areas in Comm., vol. 6, pp. 314–323, Feb. 1988.

[15] Ben Aicha, S. Ben Jebara and D. Pastor "Speech denoising improvement by musical tones shape modification," International Symposium on Communication, Control and Signal Processing ISCCSP, Morocco 2006.

[16] Kris Hermus, Patrick Wambacq, and Hugo Van hamme, "A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition," EURASIP Journal on Advances in Signal Processing, Vol. 2007, Article ID 45821, pp. 1-15, April 2006.

[17] Md. Jahangir Alam, Douglas O'Shaughnessy and Sid-Ahmed Selouani, "Speech enhancement based on novel two-step *a priori* SNR estimators," to appear in *INTERSPEECH Conference*, Brisbane, Australia, September 2008.

[18] Chang Huai You, Soo Ngee Koh, and Susanto Rahardja, "Masking-based β-order MMSE speech enhancement," speech communication, vol. 48, pp. 57-70, 2006.

[19] Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," IEEE Trans. on Audio, Speech and Language Processing, vol. 16, No. 1, pp. 229-238, January 2008.

[20] Quackenbush S., T. Barnwell and M. Clements, *Objective Measures of Speech Quality*, Englewood Cliffs, NJ, USA, Prentice Hall, 1988.

[21] H. Hirsch and D. Pearce, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy environments," ISCA ITRW ASR, September 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Iterative Noise Power Subtraction Technique for Improved Speech Quality

M. Ryyan Khan[1], Taufiq Hasan[2], M. Rezwan Khan[2]

[1]Department of EEE, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.
[2]United International University, Dhaka, Bangladesh.
E-mail: ryyan.khan.eee@gmail.com

Abstract - An iterative spectral subtraction technique for speech enhancement that gives improved performance has been presented. In all the speech enhancement methods, a predetermined average power is subtracted in the spectral domain or a gain function less than unity is generated for the individual harmonics to reduce noise power. What ever the process is, it is a common observation that some of the power spectrums become negative after the enhancement process and is avoided by forcing the coefficients to zero. Under such a condition the actual amount of noise power subtracted is less than the amount intended. If the power subtraction is done in an iterative way such that the predetermined power is correctly subtracted from the noisy signal, the segmental SNR improves appreciably. The results show that an improvement in average segmental SNR by more than 0.5 dB can be achieved by adopting this technique.

I. Introduction

There are different speech enhancement techniques for noise corrupted speech signals whose performance depends on the type and level of noise. The popular ones include spectral subtraction [1], the Wiener filters [2, 3], the minimum mean square error (MMSE) estimators [4] and Kalman [5] filters. Recently wavelet based filtering methods [6] are also showing promise in speech de-noising. The basic idea behind all the speech enhancement techniques in the transformed domain (Fourier or wavelet transformation) includes an estimation of noise and a direct or indirect subtraction of noise from the coefficients in the transformed domain. In Wiener filter, a gain factor less than unity is calculated based on the noise power and is multiplied with the noisy coefficients to reduce the average noise level. In case of spectral subtraction method, an estimated noise is subtracted directly from the noisy coefficients. In either of the cases there arise situations when the gain factor becomes negative or the subtracted coefficient becomes negative. In those situations, negative coefficients are forced to assume zero values, as negative coefficient in the power spectrum is not a meaningful quantity. This, in other words, means that the actual amount of power that is effectively subtracted is less than intended. In case of any signal enhancement technique, it is a well accepted fact that we can never estimate the amount of noise power in a particular harmonic component. But subtraction of a pre-calculated power from the coefficients statistically reduces the noise content. So, when we force the negative coefficients to zero in the de-noising process, the unaccounted amount of power remains hidden in other harmonic components, effectively leaving the speech segment with some of the pre-calculated noise power still remaining.

In this paper, we propose an Iterative Power Subtraction (IPS) method where a pre-calculated amount of power is subtracted from the noisy signal. This method ensures that the exact amount of power is subtracted to achieve higher signal to noise ratio. Experimental results presented in this paper show that there is an improvement in output average segmental SNR by about 0.5 dB compared to Wiener filter using the optimal estimate of *a priori* SNR [7] when the iterative de-noising method is adopted.

II. Proposed Method

In the proposed method we follow the usual power subtraction technique in an iterative form. Let, a signal $x(n)$ is corrupted by an additive random white noise $v(n)$ and the resultant noisy signal is represented by $y(n)$ where,

$$y(n) = x(n) + v(n). \tag{1}$$

Say, the corresponding Fourier Domain relationship is,

$$Y(k) = X(k) + V(k) \tag{2}$$

where $Y(k)$, $X(k)$ and $V(k)$ are the Fourier transformation coefficients of $y(n)$, $x(n)$ and $v(n)$ respectively, and k is the frequency index. Let us assume that the average noise power can be quite accurately estimated from the noisy signal. There are a number of standard techniques like identifying the pause period noise power or the higher frequency components of the noisy signal. What ever the process we adopt, let a noise power of $P_n(0)$ is to be subtracted from the noisy signal. Here,

$$P_n(0) = \gamma P_{nn} \tag{3}$$

where, P_{nn} is the estimated average noise power and $\gamma>1$. This is a common practice, as subtraction of exact noise power leave many of the noise peaks left out in the signal. The usual power subtraction would estimate de-noised Fourier coefficients by using equation (4),

$$\tilde{Y}_1(k) = \begin{cases} \dfrac{Y(k)}{|Y(k)|}\left[|Y(k)|^2 - P_n(0)\right]^{1/2}, \\ \qquad for\ |Y(k)|^2 > P_n(0) \\ 0, \quad for\ |Y(k)|^2 \le P_n(0) \end{cases} \tag{4}$$

In the above equation, it is clear that power in some of the coefficients $Y(k)$ is less than the average noise power $P_n(0)$ and it is not possible to subtract $P_n(0)$ from them. In our proposed method, we consider $\tilde{Y}_1(k)$ as the first iteration in power subtraction. After first iteration, the amount of noise power still left in the signal can be estimated using equation (6),

$$\delta P(1) = \frac{1}{N}\sum_{k=1}^{N}\left[|Y(k)|^2 - \left|\tilde{Y}_1(k)\right|^2\right] \tag{5}$$

$$P_n(1) = P_n(0) - \delta P(1) \tag{6}$$

Where, $\delta P(1)$ is the amount of power that has been subtracted and $P_n(1)$ is the amount of noise power still left in the signal after the 1st iteration. N is the total number of data points.

In the 2nd iteration, the residual noise in the estimated signal $\tilde{Y}_1(k)$ is $P_n(1)$. So, the new signal estimate after 2nd iteration is found from equation (7),

$$\tilde{Y}_2(k) = \begin{cases} \dfrac{Y(k)}{|Y(k)|}\left[\left|\tilde{Y}_1(k)\right|^2 - P_n(1)\right]^{1/2}, \\ \qquad for\ \left|\tilde{Y}_1(k)\right|^2 > P_n(1) \\ 0, \quad for\ \left|\tilde{Y}_1(k)\right|^2 \le P_n(1) \end{cases} \tag{7}$$

Proceeding this way, the estimate for the m[th] iteration is,

$$\tilde{Y}_m(k) = \begin{cases} \dfrac{Y(k)}{|Y(k)|}\left[\left|\tilde{Y}_{m-1}(k)\right|^2 - P_n(m-1)\right]^{1/2}, \\ \qquad for\ \left|\tilde{Y}_{m-1}(k)\right|^2 > P_n(m-1) \\ 0, \quad for\ \left|\tilde{Y}_{m-1}(k)\right|^2 \le P_n(m-1) \end{cases} \tag{8}$$

where,

$$\delta P(m-1) = \frac{1}{N}\sum_{k=1}^{N}\left[\left|\tilde{Y}_{m-2}(k)\right|^2 - \left|\tilde{Y}_{m-1}(k)\right|^2\right] \tag{9}$$

$$P_n(m-1) = P_n(m-2) - \delta P(m-1) \tag{10}$$

Let us set a threshold value α for the residual noise such that we stop iterating the process when $\delta P(m) = \alpha P_n(0)$, α being a small number less than unity. Hence, after the completion of the de-noising process, $\tilde{Y}_m(k)$ is the enhanced version of the noisy signal $Y(k)$. The inverse Fourier Transform of $\tilde{Y}_m(k)$ would produce the enhanced signal in the sampled domain.

III. Results

The experimental investigation for the proposed de-noising technique was performed on a number of utterances, both male and female, chosen from TIMIT database. For male utterances, speech no. 'sx229m', 'SI493', 'SI563', 'SI587' and 'sa2' were used. Speech no. 'SA1', 'SA2', 'SI649', 'SI1279' and 'SI1909' were chosen for female utterances. The sampling rate was 8kHz. A computer generated Gaussian white noise was added to the speech to generate the noisy signal. The de-noising process was implemented on such noisy speeches.

It is a well known phenomenon that human speech shows spectral stability over a short period of 32-40ms. To exploit this advantage, the speech data was divided in to a number of frames having N_F data points (N_F varied from 200 to 300 data points). In all cases, the signal was reconstructed using standard overlap-add method with a 75% overlap. The de-noising was done in the DCT domain, as it gives the advantage of better spectral resolution and does not involve complex calculations. The DCT domain frame was subdivided into a number of sub-frames having 20 data points and the de-noising technique as described in Section II of this paper was implemented. So, the residual power calculation was done on each sub-frame.

Initially, the average noise power P_{nn} was calculated from the pause period of the speech segments. The value of $P_n(0)$ was set to be $1.55 \times P_{nn}$ (i.e., $\gamma = 1.55$). This is a necessary condition to achieve high SNR and the value of the multiplier (1.55) was evaluated empirically. Although the average noise is P_{nn}, there are a good number of noise components whose values are significantly higher than the average value P_{nn}. So, a higher magnitude of noise subtraction was necessary to suppress the high noise peaks. The iterative process was stopped when $\delta P(m) = P_n(0) \times 10^{-6}$, i.e., value of α was set to 10^{-6}.

The sub-plots in Fig.1 show typical clean signal, the noisy signal and the signal after de-noising by the proposed technique. Speech no. 'SI1909' was used to generate the figure. This figure shows some of the features those are responsible for back ground musical noise. If watched closely, it can be seen that some finer details of the original signal, particularly in the narrow speech zones, are lost. This is a common feature with all the de-noising

techniques. Fig.1 shows that such distortions are quite low with the proposed enhancement technique.

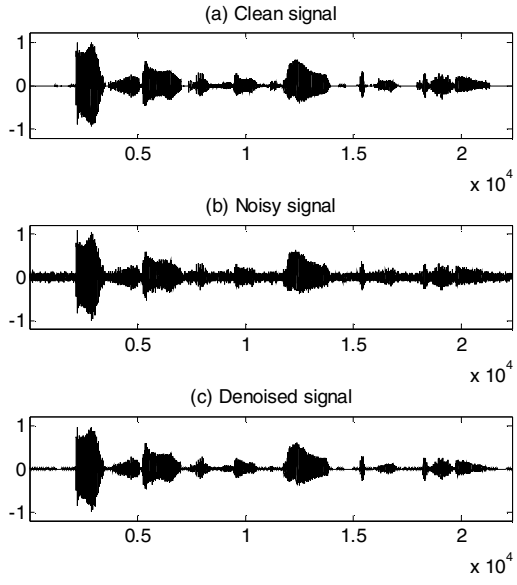


Fig. 1 Clean signal, noisy signal, de-noised signal for speech no. 'SI1909'

All the selected utterances mentioned earlier were tested under different input SNR. The average results of the utterances are presented in Fig.2. Average outputs of 10 runs of each of the utterances were used. The curves show

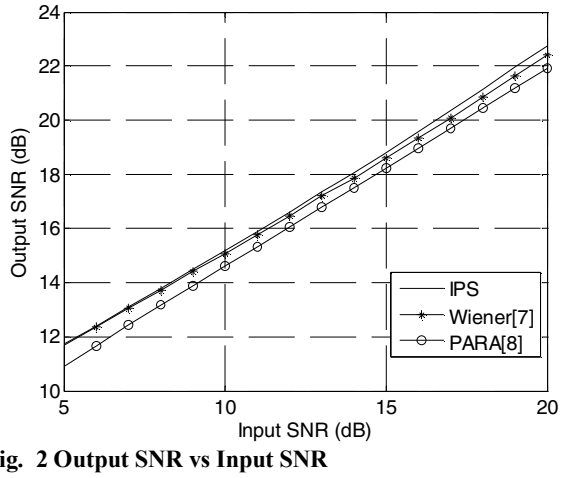


Fig. 2 Output SNR vs Input SNR

that the proposed IPS method has similar output SNR characteristics compared to Wiener filter using the optimal estimate of *a priori* SNR [7]. But IPS shows slightly better output SNR for higher input SNRs. On the other hand, compared to Parametric formulation of the generalized spectral subtraction method (PARA) [8], the proposed method shows much better output. Results for the average segmental SNR is shown in Fig.3. The proposed method shows higher average segmental SNR all through when compared to other de-noising techniques such as Wiener [7] and PARA [8]. For example, when compared to Wiener [7], IPS has more than 0.5 dB improvement in average segmental SNR for any input SNR ranging from 5 dB to 20 dB.

It is a general belief that, process generated noise is a by-product when high output SNR is to be yielded in a de-noising technique. However, the proposed technique enhances the sound quality while holding on to the high output SNR of Wiener [7] method.

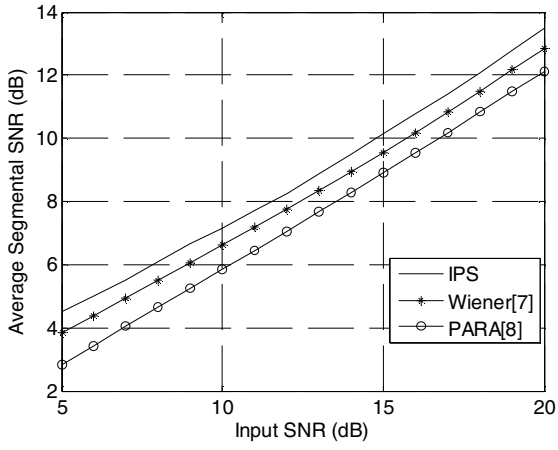


Fig. 3 Average Segmental SNR vs Input SNR

IV. Conclusions

The iterative power subtraction technique, presented in this paper, shows promise in the field of speech enhancement. The proposed technique subtracts noise power more accurately, eliminating the noise coefficients more effectively and leaves the average segmental SNR better than other existing techniques. Although improvement in overall SNR is not much, improvement in speech quality is quite significant as apparent from the comparison of segmental SNR (Fig.3). Although a formal listening test was not performed, the utterances enhanced by IPS were judged to give a better quality sound by the informal listeners. In the proposed method a fixed value of γ (equation (3)) was used all through. It is expected that a variable γ, which depends upon the SNR of a particular segment, can produce better results and is being investigated.

References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, pp. 113-120, 1979.

[2] L. Akter and M. K. Hasan, "Crosscorrelation Compensated Wiener Filter for Speech Enhancement," in *Proc. Int. Conf. Acoust., Speech and Signal Processing (ICASSP), 2006*, Toulousu, France, 2006.

[3] P. Scalart and J. Vieira-Filho, "Speech enhancement based on a priori signal to noise estimation," *Proc. ICASSP*, pp. 629-632, 1996.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Speech Audio Processing*, vol. ASSP-32, pp. 1109-1121, 1984.

[5] S. Gannot, D. Burshtein, E. Weintstein, "Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms." *IEEE Trans. Speech and Audio Processing*, Vol. 6, no. 4, pp 373-385, July 1998.

[6] M. Bahoura, J. Rouat, "Wavelet speech enhancement based on the teager energy operator," *IEEE Signal Processing Letters*, vol. 8, no. 1, pp. 10-12, January 2001.

[7] M. K. Hasan, S. Salahuddin and M. R. Khan, "A modified a priori SNR for speech enhancement using spectral subtraction rules," *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 450-453, April 2004.

[8] B. L. Sim, Y. C. Tong, J. S. Chang and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 328-337, 1998. elax Harlow, England: Addison-Wesley, 1999.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Perceptually weighted multi-band spectral subtraction speech enhancement technique

Md. Faqrul Alam Chowdhury[1], Md. Jahangir Alam[2], Md. Fasiul Alam and Douglas O'Shaughnessy[3]

[1]Dept. of EEE, BUET, Dhaka 1000, Bangladesh
[2]Dept. of EEE, KUET, Khulna 9203, Bangladesh
[3]INRS-EMT, University of Quebec, Montreal, Canada

E-mail: faqrul.buet@yahoo.com , jahangir@eee.kuet.ac.bd, fosi_bd@yahoo.com, dougo@emt.inrs.ca

Abstract –It is gaining extensive research concentration to estimate a suitable factor that will subtract just the necessary amount of the noise spectrum from each frequency bin (ideally) to prevent destructive subtraction of the speech while removing most of the residual noise. Considering the advantage of multi-band spectral subtraction method which takes into account the fact that colored noise affects the speech spectrum differently at various frequencies and also of perceptual frequency weighting technique that allows for an automatic adaptation in the time and frequency of the enhancement system and finds a suitable noise estimate according to the frequency of the corrupted speech, our proposed method offers a better spectral approximation of the clean speech signal than these two method.

Index Terms: speech enhancement, perceptual weighting filter, spectral subtraction, MBSS

I. Introduction

The removal of additive noise from speech has been an active area of research for several decades. Numerous methods have been proposed by signal processing community. Among the most successful signal enhancement techniques have been spectral subtraction and Wiener filtering. Conventional power spectral subtraction method substantially reduces the noise levels in the noisy speech. However it also introduces an annoying distortion in the speech signal called musical noise. Also occasional negative estimates of the enhanced power spectrum can occur [1]. Multi-band spectral subtraction (MBSS) method provides a definite improvement over the conventional power spectral subtraction method. The improvement of MBSS is due to the fact that the multi-band speech takes into account the non-linear effect of colored noise on the spectrum of speech [3][4] i.e. some frequencies are affected more adversely than others.

On the other hand, introduction of speech enhancement based on perceptual frequency weighting offers a better spectral approximation of the clean speech signal than spectral subtraction does. This method outperforms spectral magnitude subtraction for all types of noise and can efficiently remove additive noise related to various types of noise corruption while keeping the formant regions of speech [8]. This technique takes into consideration that the noise elements are masked by the speech power in the formant regions and conversely unmasked in the valleys between the formants.

In this paper we have taken these two facts into consideration and our proposed method, perceptually weighted multi-band spectral subtraction speech enhancement technique shows a noticeable improvement in better spectral approximation of the clean signal.

The remaining paper is organised as follows: section II provides an overview of the MBSS speech enhancement system. The perceptual frequency weighting technique is presented in section III, and a description of the proposed method is given in section IV. A discussion on the experimental results and a conclusion are drawn in section V, and section VI, respectively.

II. Multi-band spectral subtraction method

Since Boll's original work of conventional spectral subtraction [1], different variations of spectral subtraction have been proposed [2, 3, 5, 6, 7]. Most of the implementations of the spectral subtraction approach are variants of the approach proposed by Berouti et. al [2]. In the implementation proposed by [2] the estimate of clean speech spectrum is obtained as

$$\left|\hat{S}(k)\right|^2 = \left|Y(k)\right|^2 - \alpha \left|\hat{D}(k)\right|^2 \quad (1)$$

where α is an over-subtraction factor $Y(k)$, $S(k)$ and $D(k)$ are the magnitude spectra of $y(n)$, $s(n)$ and $d(n)$ which are the corrupted speech signal, clean speech signal and the noise respectively. $\hat{S}(k)$ and $\hat{D}(k)$ are the estimates of clean speech signal and noise. As noise spectrum cannot be directly obtained and it changes with the environment, the estimate $\hat{D}(k)$ is calculated during periods of silence.

If we take the assumption that the additive noise is stationary and uncorrelated with clean speech signal, the resulting corrupted speech signal can be expressed as

$$y(n) = s(n) + d(n). \tag{2}$$

Taking into account the fact that the colored noise affects the speech spectrum differently at various frequencies, in [4] a multi-band approach to spectral subtraction is proposed. The speech spectrum is divided into *N* non-overlapping bands and spectral subtraction is performed independently in each band. So the estimate of the speech spectrum in the i^{th} band is obtained by

$$\left|\hat{S}_i(k)\right|^2 = \left|Y_i(k)\right|^2 - \alpha_i \delta_i \left|\hat{D}_i(k)\right|^2, \qquad b_i < k < e_i \tag{3}$$

where b_i and e_i are the beginning and ending frequency bins of the i^{th} frequency band, α_i is the over-subtraction factor of the i^{th} band and δ_i is a "tweaking factor" that can be individually set for each frequency band to customize the noise removal properties.

The value of δ_i in (3) were empirically determined and set to [4]

$$\delta_i = \begin{cases} 1 & f_i \le 1 \text{ kHz} \\ 2.75 & 1 \text{ kHz} < f_i \le \left(\frac{F_s}{2} - 2\right) \text{ kHz}, \\ 1.75 & f_i > \left(\frac{F_s}{2} - 2\right) \text{ kHz} \end{cases} \tag{4}$$

where f_i is the upper frequency of the *i*th band, and F_s is the sampling frequency. The motivation for using smaller δ_i values for the low frequency bands is to minimize speech distortion, since most of the speech energy is present in the lower frequencies. Relaxed subtraction was also used for the high frequency bands.

α_i is the function of the segmental SNR_i of the i^{th} frequency band, calculated as

$$SNR_i(dB) = 10 \log_{10}\left(\frac{\sum_{k=b_i}^{e_i}\left|Y_i(k)\right|^2}{\sum_{k=b_i}^{e_i}\left|\hat{D}_i(k)\right|^2}\right). \tag{5}$$

Using the value calculated above, α_i can be determined as

$$\alpha_i = \begin{cases} 5 & SNR_i < -5 \\ 4 - \frac{3}{20}(SNR_i) & -5 \le SNR_i \le 20 \\ 1 & SNR_i > 20 \end{cases} \tag{6}$$

While the use of the over-subtraction factor α_i provides a degree of control over the noise subtraction level in each band, the use of multiple frequency bands and the value of δ_i weights provide an additional degree of control within each band.

The negative values in the enhanced spectrum were floored to the noisy spectrum as

$$\left|\hat{S}(k)\right|^2 = \begin{cases} \left|\hat{S}_i(k)\right|^2 & \left|\hat{S}_i(k)\right|^2 > 0 \\ \beta\left|Y_i(k)\right|^2 & \text{else} \end{cases} \tag{7}$$

where the spectral floor parameter was set to β =0.002.

III. Perceptual frequency weighting technique

In spectral subtraction method deciding the optimal value of α, the over-subtraction factor is very difficult due to a reduction in the residual noise while the speech quality is maintained [8]. So in [8] an adaptation of the parameters is performed based on the concept of perceptual weighting. It is due to the fact that the SNR is much higher around spectral peaks than it is near spectral valleys. Noise auditory impressions are generally provided by the parts of the spectrum with a low SNR, such as the spectral valleys. On the other hand, spectral peaks carry the most important information. Therefore the attenuation of the spectral valleys is thought to be very effective due to the reduction of speech distortion to a human listener. This encourages us to think about treating spectral peaks and spectral valleys differently with the help of a perceptual weighting filter shown in figure 2.

A. Perceptual weighting filter

It is an IIR filter which shapes the overall spectrum of noisy speech to exploit the masking properties of human ear [14, 15]. The PWF used in this study is 6th order IIR filter defined by the transfer function

$$H(z) = \frac{A\left(z/r_1\right)}{A\left(z/r_2\right)} = \frac{\sum_{i=0}^{p} a_i r_1^{\,i} z^{-i}}{\sum_{i=0}^{p} a_i r_2^{\,i} z^{-i}} \tag{8}$$

where *A(z)* is the LPC synthesis filter, { a_i , where i= 1, 2, 3, ...p, a_0 =1} are the LPC co-efficient, r_1 and r_2 are weighting factors between 0 and 1, and *p* is the order of the LPC filter. r_2 compensates for the tilt of the speech spectrum being flat and r_1 weights the spectral valleys. The value of r_1 and r_2 modify the frequency response of the filter. In this paper we have chosen $r_1 = 0.9$ and $r_2 = 0.8$ by experiment. The reason for using the 6th order filter is that this filter is sufficient for calculating the spectral envelop which consists of the 1st-3rd formants.

IV. Proposed method

Our proposed method, Perceptually weighted multi-band spectral subtraction speech enhancement technique can be thought of as an improvement of MBSS predicting the value of α_i (which provides a degree of control over the noise subtraction level in each band) effectively by perceptual frequency weighting technique with the benefit of using multiple frequency bands and also with the use of weights to provide an additional degree of control within

each band. Figure 1 represents the block diagram of our proposed method.

According to the MBSS method equation (6) shows that α_i is constant over for a particular band and several bands can have the same value of α_i. By using the perceptual frequency weighting technique, the noise elements are considered to be masked by the speech power in formant regions and conversely unmasked in the valleys between the formants. Therefore the gain factor which decides the amount of estimated noise subtracted from the noisy input signal is controlled to be lower in formants and higher in valleys and therefore can have different values in the same band.

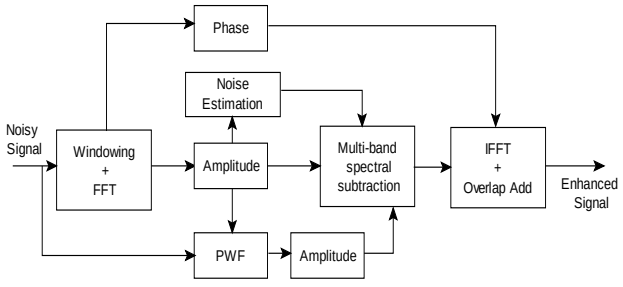


Figure 1 Block diagram of the proposed perceptually weighted multi-band spectral subtraction speech enhancement technique.

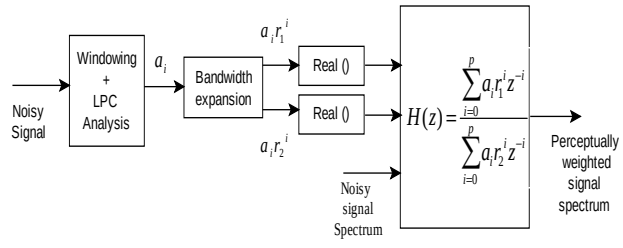


Figure 2 Block diagram of the perceptual weighting filter (PWF)

The frequency weighting function is derived by the noise shaping process for speech coding system proposed by Atal and Schroder [9]. The weighting function is calculated by the following equation

$$h_w(k) = S\left[\left|H_w(z)\right|\right] \tag{9}$$

where S[.] denotes a shaping process. The magnitude characteristic $\left|H_w(z)\right|$ is found from the equation (8) and normalized value of shaping function is taken. Thus in the proposed method, the spectral magnitude of enhanced speech is given by the following equations.

$$\left|\hat{S}_i(k)\right|^2 = \left|Y_i(k)\right|^2 - h_w(k)\alpha_i\delta_i\left|\hat{D}_i(k)\right|^2, \qquad b_i < k < e_i \tag{10}$$

where b_i and e_i are the beginning and ending frequency bins of the i^{th} frequency band, α_i is the over-subtraction factor of the i^{th} band measured by the equation (5) and (6), δ_i is a "tweaking factor" as described above in the multi-band spectral subtraction method by equation (4). The negative values in the enhanced spectrum were floored to the noisy spectrum as in equation (7).

The enhanced spectrum within each band is then combined. Because the phase of the noisy input signal is not modified, the enhanced speech signal is determined by the inverse-FFT. Finally the standard overlap-add method is used to obtain the enhanced signal.

V. Experimental results

In order to evaluate the performance of the proposed perceptually weighted MBSS technique, we conducted two objective quality tests, namely the Segmental SNR (SegSNR) and the Cepstrum Distance (CD), under various noisy environments. A subjective test is not carried out in this paper but is necessary in the future.
It is well known that segmental SNR is more accurate in indicating the speech distortion than the overall SNR. The higher value of the segmental SNR indicates the weaker speech distortions. This CD measure is considered to be a human auditory measure. The higher CD reflects the stronger speech distortions.

Table 1 depicts the experimental conditions used in our implementation. To evaluate and compare the performance of the proposed perceptually weighted MBSS method, we carried out simulations with the *NOIZEUS* database [13]. Speech signals were degraded with four types of noise at global SNR levels of 0 dB, 5 dB, 10 dB and 15 dB. The noises were Car noise, and Babble noise.

Figure 3, (a) and (b) show the Average SegSNR (Avg. SegSNR) when the signal is degraded with the Car noise and Babble noise, respectively whereas Figure 3, (c) and (d) show the CD when the signal is degraded with the Car noise and Babble noise, respectively, for the different input global SNR levels mentioned before. The speech enhancement algorithms for performance evaluation comparison include the proposed perceptually weighted MBSS technique and the conventional MBSS technique. It is observed that the proposed method yields better average segmental SNR than that of the MBSS method. In the cases of the CD measure the proposed method exhibits lower values of CD for all noisy environments compared to those obtained by the conventional MBSS method. These results demonstrate that the proposed method offers a better spectral approximation of the clean signal than does the MBSS.

Table 1 Experimental conditions

Sampling rate	8 kHz
Quantization bit rate	16 bit
Frame length	256
Frame period	128
Window function	Hamming
Speech Corpus	Noizeus
Noise	Car and Babble
Objective evaluation	SegSNR, CD

Figure 4 represents the spectrograms of the clean speech signal, noisy signal and enhanced speech signals obtained using the proposed perceptually weighted MBSS and the conventional MBSS method. The speech spectrograms provide more accurate information about the residual noise and speech distortion than the corresponding time domain waveforms. We compared the spectrograms for each of the methods and confirmed a reduction of the residual noise and speech distortion. Speech spectrograms presented in Figure 4 use a Hamming window of 256 samples with 50% overlap and the noisy signal includes Babble noise with SNR = 5 dB.

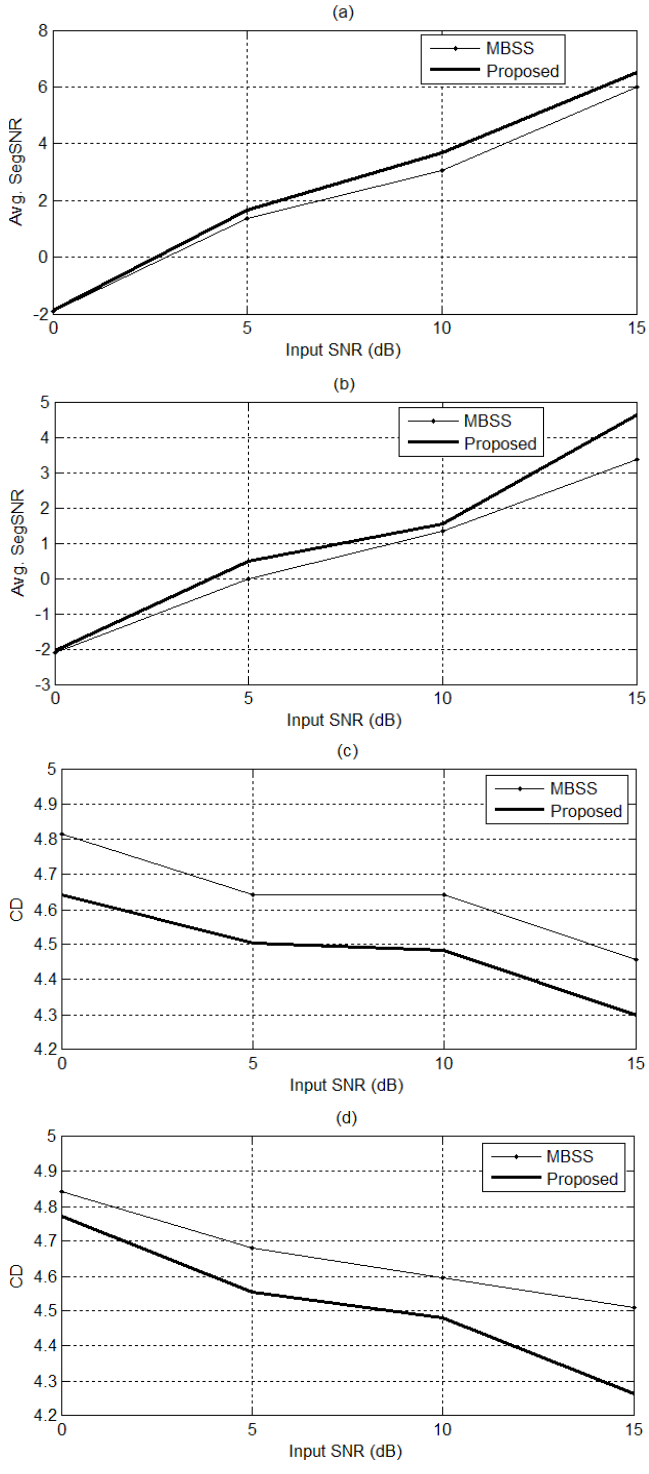


Figure 3 Experimental results. (a) Average segmental SNR (Avg. SegSNR), (c) Cepstrum Distance (CD), when the signal is degraded with Car Noise and (b) Average segmental SNR (Avg. SegSNR), (d) Cepstrum Distance (CD), when the signal is degraded with Babble Noise.

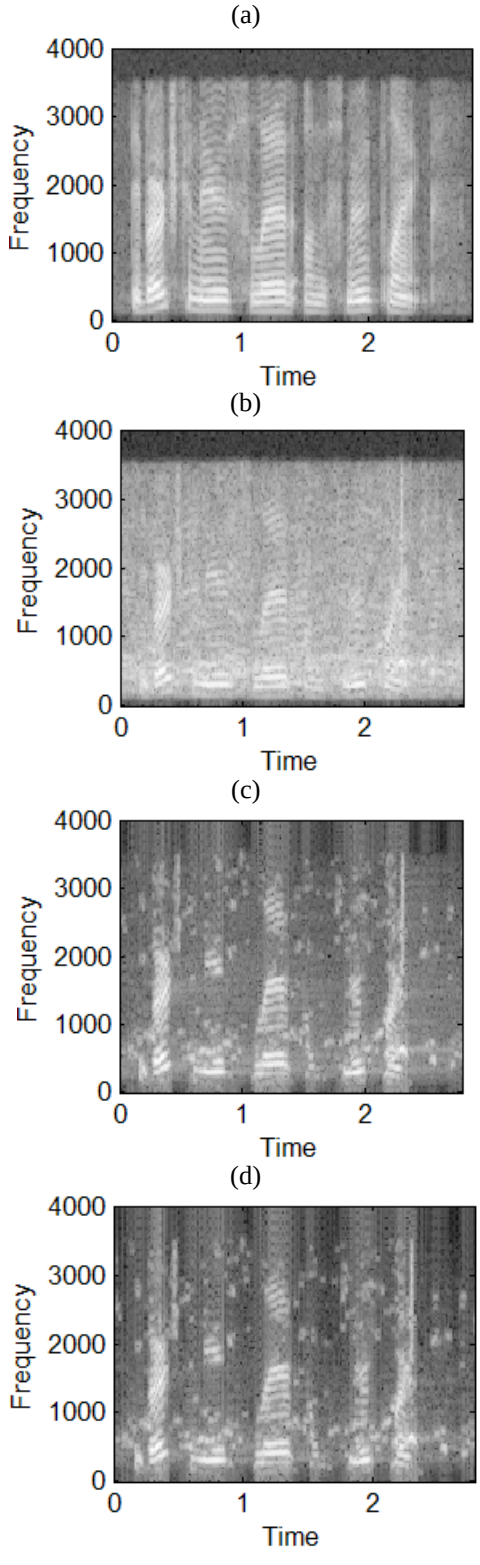


Figure 4 Speech spectrograms, Babble noise, SNR= 5 dB. (a) Clean signal, (b) Noisy signal, (c) Enhanced signal obtained using the MBSS method, and (d) Enhanced signal obtained using the proposed method.

VI. Conclusion

We have proposed a MBSS speech enhancement method based on the introduction of the perceptual frequency weighting function using a perceptual weighting filter (PWF). Experiments results and plotted spectrogram show that the proposed perceptually weighted MBSS speech enhancement method can efficiently maximize additive noise reduction while minimizing the speech distortion.

References

[1] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction." *IEEE Trans. Acoust. Speech Signal Process...*vol.ASSp-27, no.2, pp. 113-120, 1979.

[2] M.Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp.208-211, Apr. 1979.

[3] C. He, G. Zweig, "Adaptive two-band spectral subtraction with multi-window spectral estimation," *ICASSP*, vol.2, pp.793-796, 1999.

[4] Sunil D. Kamath, Philipos C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, July 2002.

[5] P. Lockwood, J. Boundy, "Experiments with a non-linear spectral subtractor (NSS), hidden markov models and the projection, for robust speech recognition in cars," *Speech Communication*, vol. 11, Nos. 2-3, pp. 215-228, 1992.

[6] I. Soon, S. Koh, C. Yeo, "Selective magnitude subtraction for speech enhancement," *Proceedings. The Fourth International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region*, vol.2, pp. 692-695, 2000.

[7] N. Virag, "Single channel enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio Processing, pp. 126-137, vol. 7, March 1999.

[8] Seiji HAYASHI, Mashahiro SUgUIMOTO, "Single channel speech enhancement based on perceptual frequency weighting," IEICE Trans. INF & SYST., vol. E90-D, no. 6, June 2007.

[9] B. S. Atal, M. R. Schroder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust. Speech Signal Processing*, vol.ASSP-27, no. 3, pp. 247-254, 1999.

[10] L. Arslan, A. McCree, V. Viswanathan, "New methods for adaptive noise suppression," *ICASSP*, vol. 1, pp. 812-815, may 1995.

[11] Y. Cheng, D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," *ICASSP*, vol. 2, pp. 961-964, Apr. 1991.

[12] J. Deller, Jr., J. Hansen, J. Proakis, "Discrete-Time Processing of Speech Signals," *NY: IEEE Press, 2000.*

[13] Hu, Y. and Loizou, P., "Subjective comparison of speech enhancement algorithms," *Proceedings of ICASSP-2006*, I, pp. 153-156, Toulouse, France, May 2006.

[14] H. Tolba, Z. Li, and D. O'Shaughnessy, "Robust automatic speech recognition using a perceptually-based optimal spectral amplitude estimator speech enhancement algorithm in various low-SNR environments," *INTERSPEECH - Eurospeech*, pp. 937–940, September 2005.

[15] Z. Li, H. Tolba, and D. O'Shaughnessy, "Robust automatic speech recognition using an optimal spectral amplitude estimator algorithm in low-SNR car environments," INTERSPEECH - ICSLP, pp. 2041–2044, October 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Two Pass Method to Impulse Noise Reduction from Digital Images Based on Neural Networks

Alireza Rezvanian [a], Karim Faez [a,b], and Fariborz Mahmoudi [a]

[a] Department of Electrical and Computer Engineering, Azad University of Qazvin, Iran
[b] Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
rezvan@qazviniau.ac.ir , faez@aut.ac.ir , f.ahmoudi@qazviniau.ac.ir

Abstract - The image enhancement and removal impulse noise from digital images is one of the most research challenges in image processing scope. There are many various methods for removal impulse noise such as Median based filters or nonlinear filters. But these methods more or less cause images to blur and to remove important details from images, as in high noise ratio will destroy vital information such as edges. Also some ways proposed to impulse noise removal using soft computing that have better performance. This paper presents an efficient hybrid method in two pass for removal impulse noise. At the first pass Impulse Noise Detection using ANFIS and the second pass Impulse Noise Estimation, that corrupted noise pixel replaced with new value based on ANN. Our method experimented on some popular test images and compared with other methods using subjective and objective measures. Results are shown that our proposed method is efficient in impulse noise removal and better than the other compared methods.

I. Introduction

The impulse noises in digital images are usually happened during image acquisition and/or transmission because of undesirable sensors or/and transmission channel [1]. Impulse noise removal from digital images is important for researcher as it was one of the most challenges in digital image processing [2]. The aim of impulse noise removal is suppress the noise while preserving the important fine details and edges [3]. There are a variety large number of techniques for impulse noise removal in the literatures [20]. The standard median filter, replaced median value of neighbours in center pixel [5] but this method changed uncorrupted noise pixel and in high noise ratio destroyed fine details and image appears very blurred. Other types of median filter [6] also proposed such as Center-Weighted Medium Filter or Center-Adaptive Weighted Medium Filter [7] as these methods better than standard median filter [8]. Although many methods was derivative from standard median filter with widely usage [20] but it most cause to damage natural features of image [28]. Other types of impulse noise removal methods are combination, that impulse noise detection and impulse noise cancellation [9] [12] [13] [14] [15] [16]. Except of replacement value certainly efficiency of these methods depends on impulse noise detection. Some methods used impulse noise detection are iterative procedure [10], Classifier-augmented median filters [12], threshold boolean filters [11], universal noise removal [29], pixel counting methods [14], Jarque-Berra test [17], global-local noise detector [13]. In most of nonlinear methods happened undesirable effects [4], so need the trade-off between impulse noise removal and edge preservation [3]. New methods proposed based on soft computing as good as nonlinear filter and act better than nonlinear filters and include combination features [2] [4] [21]. But these methods are more complicated from traditional methods. In this paper, a two-pass method is presented, which exhibits better impulse noise removal than many other more complicated methods. Proposed method with details presented in section II, experimental results in section III shown the proposed method, performs better than the other compared state-of-the-art methods.

II. Our Proposed Method

The our proposed method is based on two pass. In the first pass attempts to impulse noise detection using ANFIS (ANFIS: Adaptive-Network-Based Fuzzy Inference) with enough precision. Details of impulse noise detector are presented in the II.A section. In the second, pass estimate new value using neighbor pixels for replacement noise pixel corrupted by another artificial neural network. Details of impulse noise estimator are presented in the II.B section. General schema of our proposed is shown in Fig. 1.

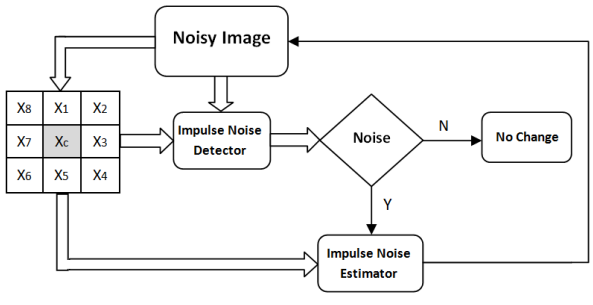


Fig. 1 General schema of proposed method

A. Impulse Noise Detector

Because of uncertainty in impulse noise and preserving the important fine details, we used a type of Fuzzy Neural Network named Adaptive-Network-Based Fuzzy Inference (ANFIS). ANFIS, is a fuzzy inference system implemented in the framework of adaptive networks [22] [30] [31].ANFIS serves as a basis for constructing a set of

fuzzy if-then-else rules with appropriate membership functions to generate the stipulated input-output pairs. Details of ANFIS are explained in [22] and a number of various applications are demonstrated in [31]. In this paper the computational structure of ANFIS has not been mentioned because many studies have been achieved on ANFIS in the literatures [22] [30] [31]. Structure of ANFIS is shown in Fig 2.

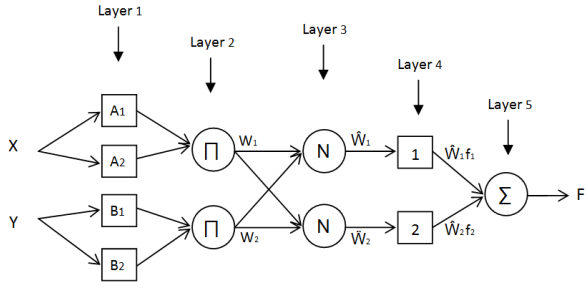
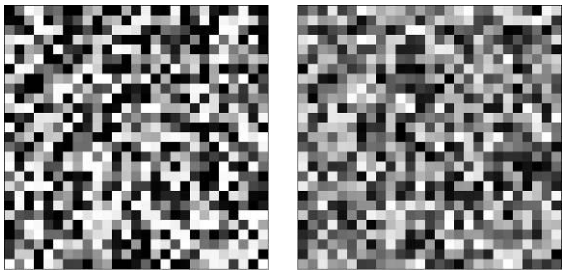


Fig. 2 ANFIS structure [22]

For training Impulse noise detector was used artificial training image with random pixels and the size of training image is 81-by-81. Noisy training image corrupted by an impulse noise 40% noise ratio. Training images is shown in Fig 3.



(A) (B)

Fig. 3 The training images. (A) Original (B) Noisy

In table 1 we show results of accuracy in noise detection after training on some popular gray-scale test images [32] like *Lena*, *Baboon* and *Peppers* in size of 512-by-512. Of course with different properties of images get different results.

Table 1 Comparison of accuracy results in impulse noise detection for images Lena, Baboon and Peppers.

Noise ratio	Lena	Baboon	Peppers
10 %	99.9866	99.8825	97.5482
20 %	99.9882	99.8913	97.8286
30 %	99.9889	99.9045	98.0583
40 %	99.9924	99.9122	98.3576
50 %	99.9947	99.9245	98.6443
60 %	99.9947	99.9442	98.9025
70 %	99.9951	99.9457	99.1924
80 %	99.9973	99.9683	99.4636

B. Impulse Noise Estimator

In second pass replace detected noise pixel with new value that estimated by feed-forward back propagation network, we named Impulse Noise Estimator. In this method unlike traditional methods [5] [6] [7] uncorrupted pixels left no changes. Estimator tries to estimate precise value using uncorrupted noise neighbors pixels in some iteration. Threshold of converge difference between iteration makes to stop iterations.

X8	X1	X2
X7	XC	X3
X6	X5	X4

Fig. 4 Center pixel and the neighbors with boundary of one

In [4] [18] [27] for training network were used previous training images likely in Fig.3, but the authors believe that based on Correlation Theorem [1], those methods were not reasonable and good results achieved by really training image. Fig. 5 shows accuracy of our idea.

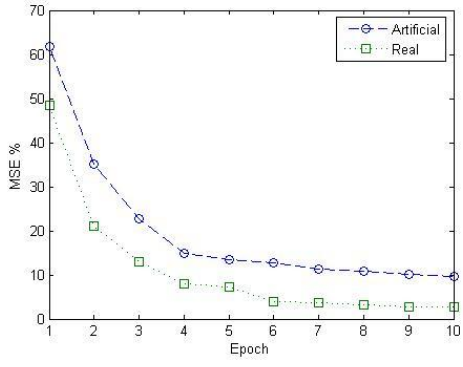


Fig. 5 Comparison of Mean Squared Error for Artificial image training and real image training

The neural network designed in three layers. No discussion about input and output layer. Best number of nodes in hidden layer for perfect value output obtained by some experiments. Table 2 shows results of accuracy in the train and the test samples.

Table 2 Comparison of train accuracy and test accuracy to obtained hidden layer nodes.

Structure	Training	Testing
8x9x1	97.844	82.398
8x10x1	97.250	81.452
8x11x1	97.169	80.868
8x12x1	96.981	81.206
8x13x1	97.498	80.502
8x14x1	97.136	82.190
8x15x1	97.501	81.925
8x16x1	98.008	82.950
8x17x1	97.167	80.355
8x18x1	95.925	77.898
8x19x1	97.623	82.310
8x20x1	97.595	82.299

The results of Table 2 presented in Fig.6 for better demonstration.

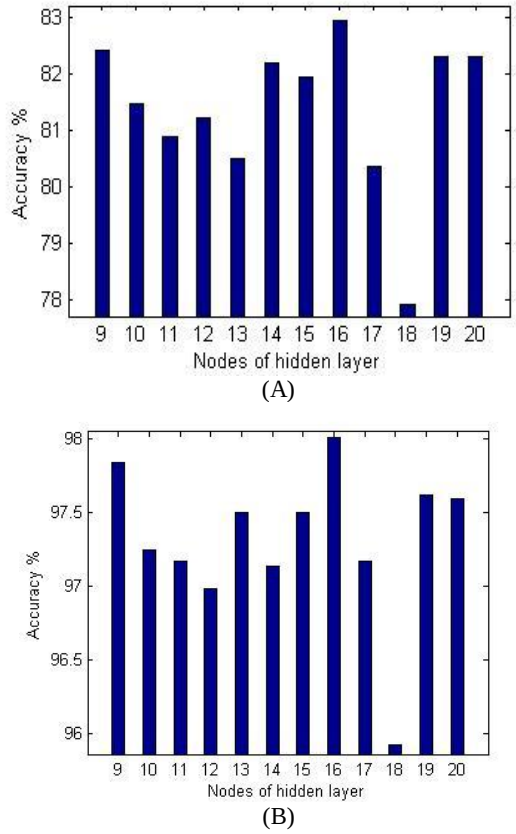


Fig. 6 Comparison of accuracy for hidden layer nodes. (A) Train accuracy (B) Test accuracy

Based on results in Table 2 and Fig. 6 number of nodes obtained *16*. It means that in our design, perfect estimate value for noise pixel achieved by 8x16x1 structure. Proposed structure is demonstrated in Fig. 7.

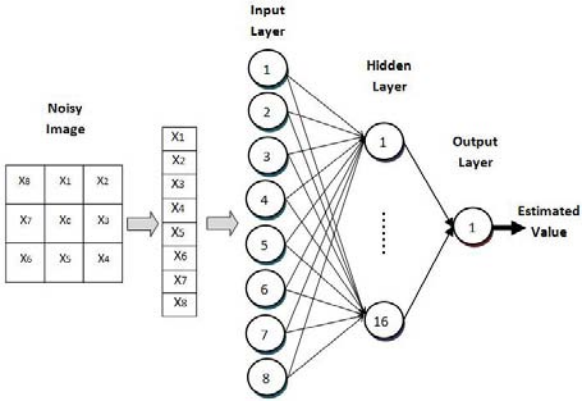
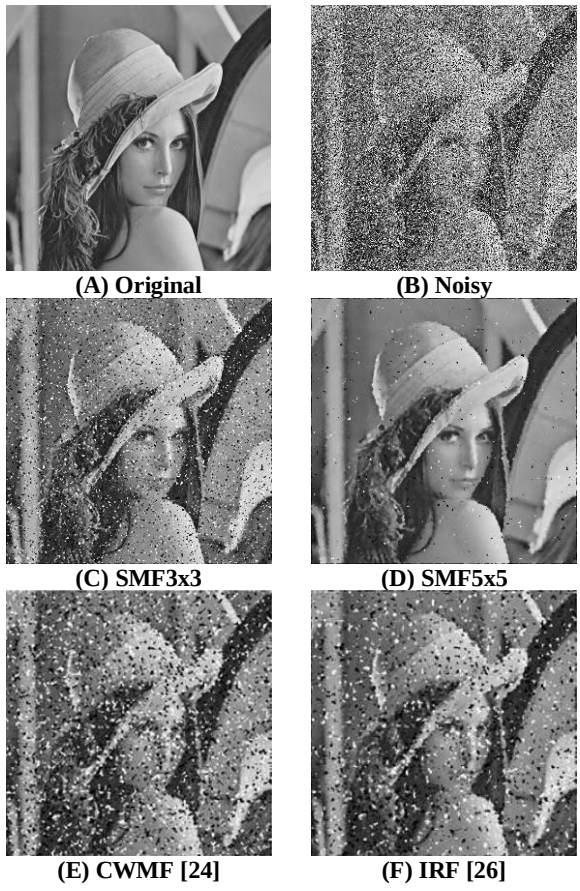


Fig. 7 Proposed artificial neural network structure for Impulse Noise Estimator

III. Experimental results

The Proposed approach in two stages as Impulse Noise Detector and Impulse Noise Estimator explained in previous section. In this section we compare the performance of our proposed approach with some methods for impulse noise reduction. The experiments were carried out on the most popular test image from the literature including *Lena* and *Baboon* [32]. For comparison, the corrupted experimental images are also restored by using several traditional and state-of-the-art impulse noise reduction methods including, *Standard Median Filter* with 3x3 mask as SMF3x3 and 5x5 mask as SMF5x5 [5] [6], *Center Weighted Median Filter* as CWMF [24], *Impulse Rejecting Filter* as IRF [26], *Switching Median Filter using Edge Detection Kernels* as SMFEDK [16], *Progressive Switching Median Filter* as PSMF [23], *Recursive Signal Dependent Rank Order Mean Filter* as SDROMRF [25], *Fuzzy Filter* as FF [19], *recursive adaptive-center weighted median filter* as ACWMRF [7], *Iterative Median Filter* as IMF [23], *Simple Neuro-Fuzzy Filter* as SNFF [27]. At the first the subjective quantitative measure as visual presented in Fig. 8 for *Lena* and Fig. 9 for *Baboon*.



(A) Original **(B) Noisy**

(C) SMF3x3 **(D) SMF5x5**

(E) CWMF [24] **(F) IRF [26]**

(G) SMFEDK [16]

(H) PSMF [23]

(I) SDROMRF [25]

(J) FF [19]

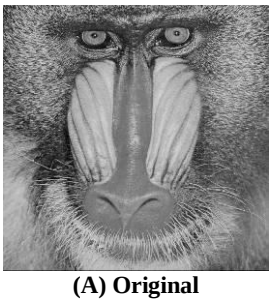
(K) ACWMR [7]

(L) IMF [23]

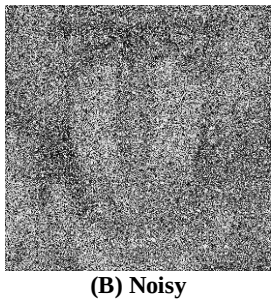
(M) SNFF [27]

(N) Proposed

Fig 8. Comparison of the restored images of the methods for the Lena by impulse noise with 50 %. (A) Original (B) Noisy (C) SMF3x3 (D) SMF5x5 (E) CWMF (F) IRF (G) SMFEDK (H) PSMF (I) SDROMRF (J) FF (K) ACWMR (L) IMF (M) SNFF (N) Proposed



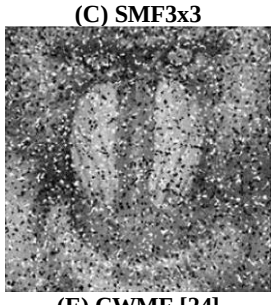
(A) Original



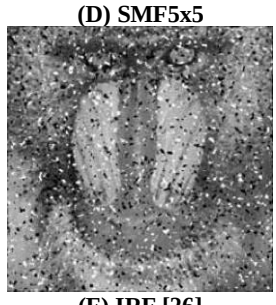
(B) Noisy

(C) SMF3x3

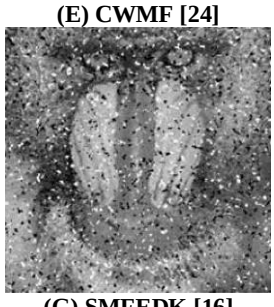
(D) SMF5x5



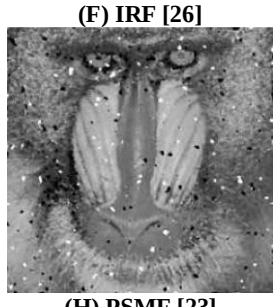
(E) CWMF [24]



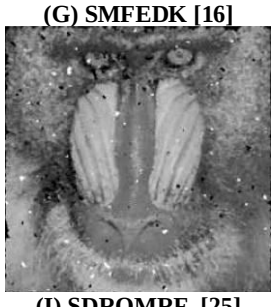
(F) IRF [26]



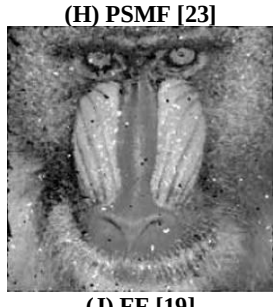
(G) SMFEDK [16]



(H) PSMF [23]

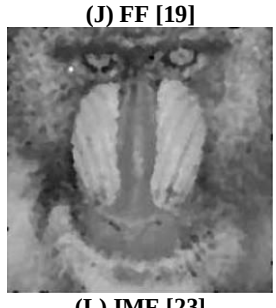


(I) SDROMRF [25]

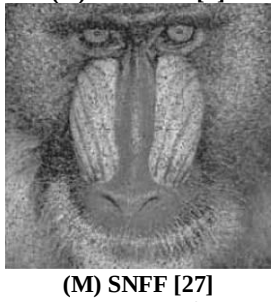


(J) FF [19]

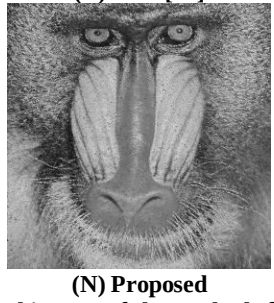
(K) ACWMR [7]



(L) IMF [23]



(M) SNFF [27]



(N) Proposed

Fig. 9 Comparison of the restored images of the methods for the Baboon by impulse noise with 50 %. (A) Original (B) Noisy (C) SMF3x3 (D) SMF5x5 (E) CWMF (F) IRF (G) SMFEDK (H) PSMF (I) SDROMRF (J) FF (K) ACWMR (L) IMF (M) SNFF (N) Proposed

The performances of methods are also evaluated by use of the objective quantitative measure which is used comparison that is called Mean Squared Error (MSE) criterion, which is mentioned in most literatures [3] [4] [7] [9] [28] [30] and defined as

$$MSE = \frac{\sum_{i=1}^{n}\sum_{j}^{m}(X(i,j)-Y(i,j))}{N \times M} \qquad (1)$$

Where $X(i, j)$ is the original image, $Y(i, j)$ is the restored image and NM is the size of image. In experiments perfect is minimize the MSE value. The MSE values listed in Table 3 for Lena and Table 4 for Baboon. In Fig. 9 were show performances of some methods as a function of noise ratio.

Table 3 Comparison of the MSE values for restoration results of the various methods in for the Lena test image

Method	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %
Noisy Image	1831.3	3734	5604.6	7374.6	9245.3	11155	12980	14869
SMF3x3	30.375	82.265	298.51	806.22	1893.3	3807	6401.9	9997.1
SMF5x5	58.34	83.44	123.81	173.02	326.19	929.55	2494.6	6051.9
CWMF [24]	76.27	259.94	727.17	1729.3	3273.4	5431.6	8350.4	11291
IRF [26]	61.15	127.84	366.24	981.06	2098.8	3918.5	6821.7	9982.4
SMFEDK [16]	33.52	96.34	299.66	892.03	1909.1	3580.6	6407.9	9482.3
PSMF [23]	54.92	85.76	132.48	272.63	647.3	1938.3	5036	9495.1
SDROMR [25]	56.79	91.21	171.88	287.33	553.76	1071.8	2452.1	5615.9
FF [19]	51.45	111.9	208.57	346.21	605.95	1114.9	2175.5	4372.3
ACWMR [7]	46.95	85.67	170.71	299.57	536.23	1007.8	2042.3	4230.2
IMF [23]	124.05	140.06	158.84	203.36	261.87	491.57	1402.7	3971.5
SNFF [27]	63.63	94.21	256.73	417.24	647.59	970.07	1375.4	1968.4
Proposed	21.273	45.785	80.415	116.85	185.84	310.75	524.82	1038.9

Table 4 Comparison of the MSE values for restoration results of the various methods in for the Baboon test image

Method	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %
Noisy Image	1804.7	3602.2	5404.3	7220.2	9050	10804	12626	14441
SMF3x3	317.28	407.18	641.76	1217.6	2313.7	4045.8	6657.6	9930.1
SMF5x5	497.33	533.95	586.69	668.11	864.83	1442.3	3040.5	6294.9
CWMF [24]	184	348.5	836.76	1795.2	3319.7	5272.3	8037.5	10960
IRF [26]	171.56	252.63	497.92	1071	2190.2	3856	6614.3	9786.6
SMFEDK [16]	86.43	161.51	398.66	914.19	1972.9	3523.1	6156.2	9312.5
PSMF [23]	108.39	147.8	205.12	335.73	736.27	1932.4	5146.6	9665.4
SDROMR [25]	171.74	224.09	312.82	429.68	692.46	1122.9	2482.9	5432.6
FF [19]	73.8	139.75	240.86	376.19	628.35	1114.4	2045.5	4217.2
ACWMR [7]	115.2	176.27	263.32	381.32	620.87	975.77	1888.5	3713.4
IMF [23]	324.1	340.02	358.99	386.6	456.02	657.79	1653.4	4082.5
SNFF [27]	52.45	97.71	154.39	243.57	369.95	544.18	786.46	1072
Proposed	50.972	103.92	159.39	231	323.4	458.19	704.88	1231

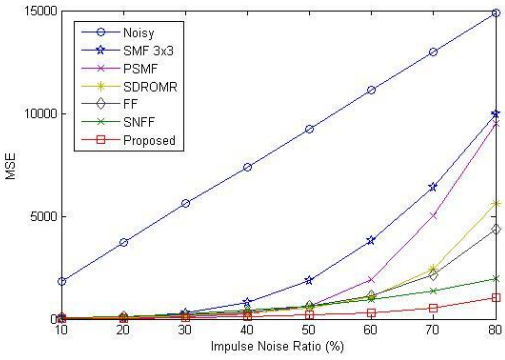


Fig. 10 Comparison of performances of some methods as a function of noise ratio for Lena image

IV. Conclusion

In this paper, we present a hybrid approach in two stages, impulse noise detection by ANFIS and impulse noise estimation by artificial neural network. The results of our approach are compared with those of some methods. Feature of two stages is tendency to achieve good results and avoid destroying uncorrupted noise pixels. Experimental results have shown the feasibility of proposed approach. A numerical measure, such as the MSE, and visual observations (Figure 8, Figure 9) have shown Convincing results. The authors in the future works will expand the method and research on other detector and the better estimator in high corrupted impulse noise.

References

[1] [1] R. C. Gonzalez, R. E. Woods, "Digital Image Processing". 2nd. Addison-Wesley Longman Publishing Co., Inc, 2001.

[2] Xu. Haixiang, Zhu Guangxi, Haoyu Peng, Desheng Wang, "Adaptive fuzzy switching filter for images corrupted by impulse noise". Pattern Recognition Letters, vol. 25, Issue 15, pp. 1657-1663, 2004.

[3] S.K. Mitra, Tian-Hu Yu, "A new nonlinear algorithm for the removal of impulse noise fromhighly corrupted images". Proceeding ISCAS '94, pages 17-20 vol.3, 1994.

[4] M.T. Yildirim, A. Basturk, M.E. Yuksel, "A Detail-Preserving Type-2 Fuzzy Logic Filter for Impulse Noise Removal from Digital Images". Fuzzy Systems Conference IEEE, London, Pages 1-6, 23-26, July 2007.

[5] M. Gabbouj, E. J. Coyle, N. C. Gallager, "An overview of median and stack filtering". Circuit Syst. and Signal Processing, vol. 11, no. 1, pp. 7-45, 1992.

[6] O. Yli-Harja, J. Astola, Y. Neuvo, "Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation". IEEE Trans. on Signal Processing, vol. 39, no. 2, pp. 395-410, 1991.

[7] T. Chen, H. R. Wu, "Adaptive impulse detection using center weighted median filters". IEEE Signal Proc. Letters, vol. 8, no. 1, pp. 1- 3, 2001.

[8] L. Yin, R. Yang, M. Gabbouj, Y. Neuvo, "Weighted median filters: A tutorial". IEEE Trans. on Circuits and Systems II, vol. 43, pp. 157-192, 1996.

[9] V. Crnojevic, V. Senk, Z. Trpovski, "Advanced impulse detection based on pixel-wise MAD". IEEE Signal Processing Letters, vol. 11, no. 7, pp. 589-592, 2004.

[10] R. H. Chan, C. Hu and M. Nikolova,"An iterative procedure for removing random-valued impulse noise". IEEE Signal Proc. Letters, vol. 11, no. 12, pp. 921-924, 2004.

[11] Aizenberg, C. Butakoff and D. Paliy, "Impulsive noise removal using threshold boolean filtering based on the impulse detecting functions". IEEE Signal Proc. Letters, vol. 12, no. 1, pp. 63-66, 2005.

[12] J. Y. Chang, J. L. Chen, "Classifier-augmented median filters for image restoration". IEEE Trans. Instrumentation and Measurement, vol. 53, no. 2, pp. 351-356, 2004.

[13] S. Q. Yuan, Y. H. Tan, "Impulse noise removal by a global-local noise detector and adaptive median filter". Signal Processing, vol. 86, no. 8, pp. 2123-2128, 2006.

[14] B. Smolka, A. Chydzinski, "Fast detection and impulsive noise removal in color images". Real-Time Imaging, vol. 11, no. 4, pp. 389- 402, 2005.

[15] S. Schulte, M. Nachtegael, V. De Witte, D. Van der Weken, E. E. Kerre, "A fuzzy impulse noise detection and reduction method". IEEE Trans. on Image Processing, vol. 15, no. 5, pp. 1153-1162, 2006.

[16] S. Zhang, M. A. Karim, "A new impulse detector for switching median filters". IEEE Signal Proc. Letters, vol. 9, no. 11, pp. 360-363, 2002.

[17] E. Besdok and M. E. Yuksel, "Impulsive noise rejection from images with Jarque-Berra test based median filter". Int. J. Electron. Commun., vol. 59, no. 2, pp. 105-109, 2005.

[18] M. E. Yuksel, E. Besdok, "A simple neuro-fuzzy impulse detector for efficient blur reduction of impulse noise removal operators for digital images". IEEE Trans. on Fuzzy Systems, vol. 12, no. 6, pp. 854-865, 2004.

[19] F. Russo, G. Ramponi, "A fuzzy filter for images corrupted by impulse noise". IEEE Signal Proc. Letters, vol. 3, no. 6, pp. 168-170, 1996.

[20] Wenbin Luo, "An efficient algorithm for the removal of impulse noise from corrupted images". AEU - International Journal of Electronics and Communications, vol. 61, Issue 8, 3, pp. 551-555, 2007.

[21] N. Alajlan, M. Kamela, E. Jernigan, "Detail preserving impulsive noise removal". Signal Processing: Image Communication, vol. 19, pp. 993- 1003, 2004.

[22] J.S.R. Jang, "ANFIS: adaptive-network-based fuzzy interface systems". IEEE Trans. Systems Man, Cybernet. 23 (3), 665–685, 1993.

[23] *Z. Wang, D. Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images". IEEE Trans. on Circuit* and Systems, vol. 46, no. 1, pp. 78-80, 1999.

[24] S.J. Ko, Y.H. Lee, "Center-weighted median filters and their applications to image enhancement". IEEE Trans. Circuits Systems II 43 (3), 157–192, 1996.

[25] Eduardo Abreu, "Signal-Dependent Rank-Ordered-Mean (SD-ROM) Filter". Nonlinear Image Processing, pp. 111-133, 2001.

[26] T. Chen, H.R. Wu, "A new class of median based impulse rejecting filters". IEEE Internat. Conf. Image Process. 1, 916–919, 2000.

[27] M.E.Yüksel, A. Ba¸stürk, "Efficient removal of impulse noise from highly corrupted digital images by a simple neuro-fuzzy operator". AEUInternat. J. Electron. Comm. 57 (3), 214–219, 2003.

[28] E. Beşdok, P. Çivicioğlu, M. Alçı, "Using an adaptive neuro-fuzzy inference system-based interpolant for impulsive noise suppression from highly distorted images". Fuzzy Sets and Systems, Volume 150, Issue 3, pp. 525-543, 16 March 2005.

[29] R. Garnett, T. Huegerich, C. Chui, W. He, "A universal noise removal algorithm with an impulse detector". IEEE Trans Image Process, 14(11):1747–54, 2005.

[30] MathWorks: MATLAB the language of technical computing, MATLAB Function Reference. New York: The MathWorks, Inc.; 2002.

[31] J. Yen, L. Wang, "Improving the interpretability of tsk fuzzy models by combining global learning and local learning" IEEE Trans Fuzzy System; 6:530–7, 1998.

[32] The USC-SIPI image database, University of Southern California [Online]. Available: http://sipi.usc.edu/services/database/Database.html , March 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Human Motion Detection and Segmentation from Moving Image Sequences

Mohiuddin Ahmad

Dept. of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
E-mail: ahmad@eee.kuet.ac.bd

Abstract - In this paper, a method for the detection and segmentation of human motion in moving image sequences is presented. For detecting motion, the intensity of each pixel is convolved with the second derivative of temporal Gaussian smoothing function or temporal Laplacian of Gaussian (LoG) filter. The zero-crossing in a single frame of the resulting function indicates the positions of moving edges. Intensity changes in time due to small illumination effects do not produce zero crossing; thus, they are not interpreted as human motion by this method. The optical flow velocity is computed by using the spatial and temporal derivatives of this function, and it is normal to the zero crossing contours. Pixels belongings to the normal velocities are back-projected in the original color image sequences to achieve a segmented color image. Experiments show that the moving object is detected correctly and achieves a good segmentation results.

I. Introduction

Human motion analysis is receiving an increasing attention from computer vision researchers. It is one of the most attractive research areas in computer vision. Human motion analysis attempts to detect, track and identify people and more generally, to interpret human behavior, from image sequences involving humans [1]. Usually human motion analysis involves human motion detection, human tracking and human behavior understanding. Human motion detection aims at segmenting regions corresponding to people from the rest of an image. It is a significant issue in human motion analysis, since the subsequent process such as tracking and activity recognition are greatly depends on it. The result of motion detection in an image is the motion segmentation, which segments the image as motion part (foreground) and stationary part (background). For simple background, it is relatively easy to detect moving objects, but for complex background, the task of detecting the accurate motion from its background is challenging. Several processes of image segmentation are briefly discussed in [1] [9].

The goal of human motion segmentation is to divide the human motion into regions that corresponds with the objects present in a scene. The goal of optical flow [2] is to examine a video sequence and calculate motion vectors correctly describe the two dimensional motion present. The motion segmentation in video sequence is known to be a significant and difficult problem, which aims at detecting regions corresponding to moving objects such as vehicles and peoples in natural scene. Detecting moving blobs provides a focus of attention for later processes such as tracking and activity analysis because only those changing pixels are considered.

Many researchers use different process of image segmentation. Every method has its own advantages and disadvantages. The studies of different process of segmentation corresponding to moving objects are cited in [9]. Briefly, they are as follows:

- Background subtraction: Motion detects in an image by differencing between current image and a reference background in a pixel-by-pixel fashion. However, it is extremely sensitive to changes in dynamic scenes due to lighting and extraneous events.
- Temporal differencing: Motion detects by pixel wise difference between some consecutive frames, such as two-frame differencing or three frame differencing
- Statistical method: Motion detects by comparing the statistics of the current background model , such as adaptive background model, background mixture model etc
- Optical flow: Motion detects by the characteristics of the flow vector of moving objects, such as flow velocity model, spatial-temporal motion model

We consider the object as human, so the task is motion segmentation from human motion detection. We mentioned that most methods for the detection of human motion in an image sequences involves the subtraction of successive frames. In the simple subtraction method, the intensities of each pixel at adjacent points in time are subtracted from each other; nonzero values in the resulting difference image indicate that something in the image has changed. It is assumed that these changes are due to motion rather than other effect, such as illumination effect.

However, the illumination can easily change that effect the image sequence even the camera is fixed in position. The illumination can change in an indoor and an outdoor image sequence. The illumination in an indoor sequence can change since the light changes at least 50 or 60 times per second due to indoor lighting; on the contrary, the image capture by a camera is 25 ~ 30 per second. Therefore, illumination has an effect on the captured image sequence. The illumination in an outdoor sequence also vary for different reasons such as the sun light change and this changes is very rapid than indoor sequence. The weather condition is another reason for changing the illumination.

Camera itself produces some random noise when the image sequence is captured. Therefore, the pixel value of one frame in an image sequence is not exactly same to the next image frame even though the camera is in fixed position. Therefore, precise detection of actual motion in an image sequence is an important issue since a good segmentation depends on precise detection of motion.

In this paper, we use a method for precise detection of human motion in an image sequence, which closely relates to the work of Duncan and Chou [4], Buxton and Buxton [3], Marr and Ullman [7]. The difference of our work is that we use only temporal Gaussian smoothing function or temporal LoG filter with optical flow technique.

In this method, the convolution of the intensity history at each pixel (x, y) point with second derivative in time of the temporal Gaussian smoothing function computes as a preprocessing stage of the system. The resulting function of Gaussian mask with the image pixel gives convoluted pixel. The convoluted pixel finds a zero crossing on the moving edges of the image sequence of human motion objects. The derivatives of the convoluted pixels are used in the optical flow constraint equation to compute the normal velocities of human motion of the zero crossing contours. Finally, a velocity threshold and median filter use to estimates the binary motion frame and segment the human motion from its static background.

The rest of the paper is organized as follows: Firstly, we give a brief introduction of the algorithm of human motion detection, the computation of normal flow and image segmentation from motion. Secondly, we show the experimental procedures, some results and discussion. Finally, we conclude this paper.

II. Algorithm

In this section, we briefly discuss the motion detection algorithm, the procedure of calculating normal flow and segmentation using motion.

A. Motion Detection

The first step in segmenting a moving object of interest from a complex background is the computation of normal component of the optical flow. As accurate estimation of visual motion is computationally expensive, a filtering technique is applied which provides an output zero-crossing image $s(x, y, t)$ by convolving the intensity history of the pixel $p(x, y, t)$ with a d'Alembertian of a temporal Gaussian filter [3]. This process is defined by Eq. (1). The width of the filter is given by $w = 2\sigma\sqrt{2}$, where σ is the standard deviation of the Gaussian function.

The convoluted pixels, $s(x, y, t)$ calculates from second derivative of this smoothing function and original image sequence, which is given by

$$s(x,y,t) = -\frac{d^2 g(t)}{dt^2} \otimes p(x,y,t) \qquad (1)$$

where,

$$g(t) = u\left(\frac{1}{2\sigma^2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{u^2 t^2}{2\sigma^2}\right)$$

u is a time scaling factor, and $\otimes$ represents the process of convolution. $s(x, y, t)$ will be zero everywhere except for the neighborhood of moving edges, where is varies continuously and linearly in the form of first derivative of Gaussian function both in time and spatial domain.

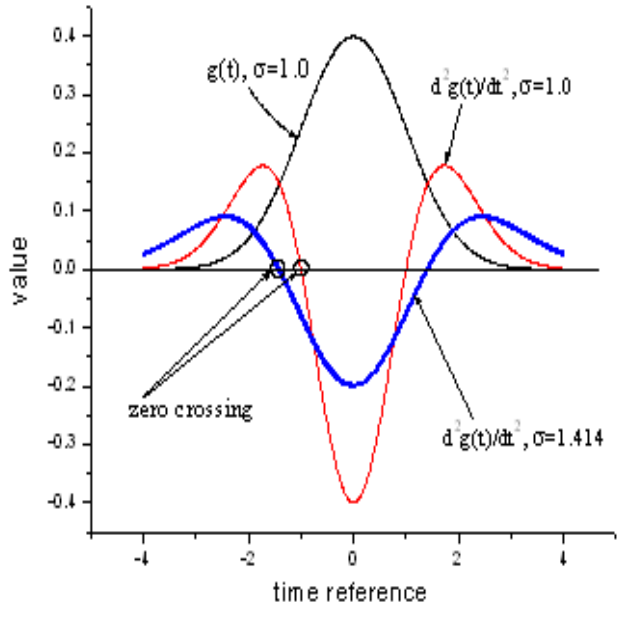


Fig. 1 Zero crossing location of the LoG filter in time reference.

It is interesting to realize that the zero crossing will occur in the location of the moving edges in spatial domain where the motion will be estimated. The zero crossing location is shown in figure 1.

The next step is to detect the zero crossing in $s(x, y, t)$ and to calculate the magnitude the velocity of each zero-

crossing pixel. The velocity is calculated using 7 successive image frames to calculate the spatial-temporal zero crossing.

B. Computation of Normal Flow

From the image brightness constancy assumption [5] [6], the gradient constraint equation selected by Horn [5] is

$$v_x \frac{\partial p}{\partial x} + v_y \frac{\partial p}{\partial y} + \frac{\partial p}{\partial t} = 0 \qquad (2)$$

Where v_x and v_y are the optical flow velocity components, and $\partial p / \partial x$, $\partial p / \partial y$ and $\partial p / \partial t$ are the spatial gradients and temporal gradient of image intensity. Now, with the output $s(x, y, t)$, the normal component of visual motion, v_n at zero crossing and the angle θ in the image plane is estimated for the middle frame of three successive frames is computed from the new gradient constraint equation or optical flow constraint (OFC) equation

$$v_x \frac{\partial s}{\partial x} + v_y \frac{\partial s}{\partial y} + \frac{\partial s}{\partial t} = 0 \qquad (3)$$

The normal component of velocity v_n at zero crossing and the angle θ in the image plane is given by

$$v_n = -u \frac{s_t}{\sqrt{{s_x}^2 + {s_y}^2}} \qquad (4)$$

$$\theta = \tan^{-1} \frac{s_y}{s_x} \qquad (5)$$

where, the derivatives $s_x = \partial s / \partial x$, $s_y = \partial s / \partial y$ and $s_t = \partial s / \partial t$ are calculated using numerical finite difference methods on values from Eq. (1). An accurate and consistent estimation of gradients are necessary for motion estimation. For estimating gradients, we use the following numerical finite difference methods [8].

$$s_x(x, y, t) = \sum_{a=-\frac{NN}{2}}^{\frac{NN}{2}} \sum_{b=-\frac{NN}{2}}^{\frac{NN}{2}} \sum_{c=-\frac{NN}{2}}^{\frac{NN}{2}} a \times s(x + a, y + b, t + c)$$

$$s_y(x, y, t) = \sum_{a=-\frac{NN}{2}}^{\frac{NN}{2}} \sum_{b=-\frac{NN}{2}}^{\frac{NN}{2}} \sum_{c=-\frac{NN}{2}}^{\frac{NN}{2}} b \times s(x + a, y + b, t + c)$$

$$s_t(x, y, t) = \sum_{a=-\frac{NN}{2}}^{\frac{NN}{2}} \sum_{b=-\frac{NN}{2}}^{\frac{NN}{2}} \sum_{c=-\frac{NN}{2}}^{\frac{NN}{2}} c \times s(x + a, y + b, t + c)$$

Subsequently, the magnitude of the normal visual motion (velocity) of the middle frame is scale in the range 0 ~ 255.

C. Segmentation using Motions

For segmenting the foreground from background, we use the motion image information. We derive the motion frame from the normal velocity at zero crossing of the image sequence. We segment the image frame based on motion associated with the original moving object (human). For segmenting the human motion with the background, we use a velocity threshold and convert the motion image to a binary motion image (BMI) frame. The binary image has some unwanted back and forth motion blobs in the image frames, which removes by filtering. In this case, we use 7×7 median filter to remove unwanted motion from the motion image frame.

After filtering, pixels belongings to the segmented velocities are back-projected in the original color frame to give a segmented color image. Back-projection is performed by sweeping through the original color image in the spatial domain and creating a new image such that only the pixels that belong to one of the segmented velocities are copied from the original image to the new image.

III. Experimental Results and Discussion

The properties of the original image pixels with the properties of $s(x, y, t)$ are discussed in [4]. We show the results of our method for some image sequences such as an indoor image sequence and an outdoor image sequence. The indoor image sequence consists of human walking motion along the corridor with complex background. The outdoor image sequence consists of a people walking in the open field with complex background. In both sequences, the camera position is stationary, so all objects are stationary except human moving. In the indoor image sequence, we use 300 frames with the image frame size is 320×240 pixels. The outdoor sequence consists of 120 frames with the frame size of 320×240 frames. The experimental results and discussions follow systematically:

Firstly, the original indoor image sequence and the outdoor image sequence are color images, which consist of rgb pixels. The intensity image derives from the color image by calculating the mean of the rgb pixels.

$$p = \frac{r + g + b}{3}$$

The original color image and the intensity image of the indoor and outdoor image sequence show in figure 2 (first

row, second row) and figure 3 (first row, second row). In these figures, we have shown two discrete frames from two distinct locations. In figures, MF means the middle frame and the numbers indicate the frame number of the image sequence; the numbers indicated in left and right denote the starting and ending frame to obtain zero crossing images.

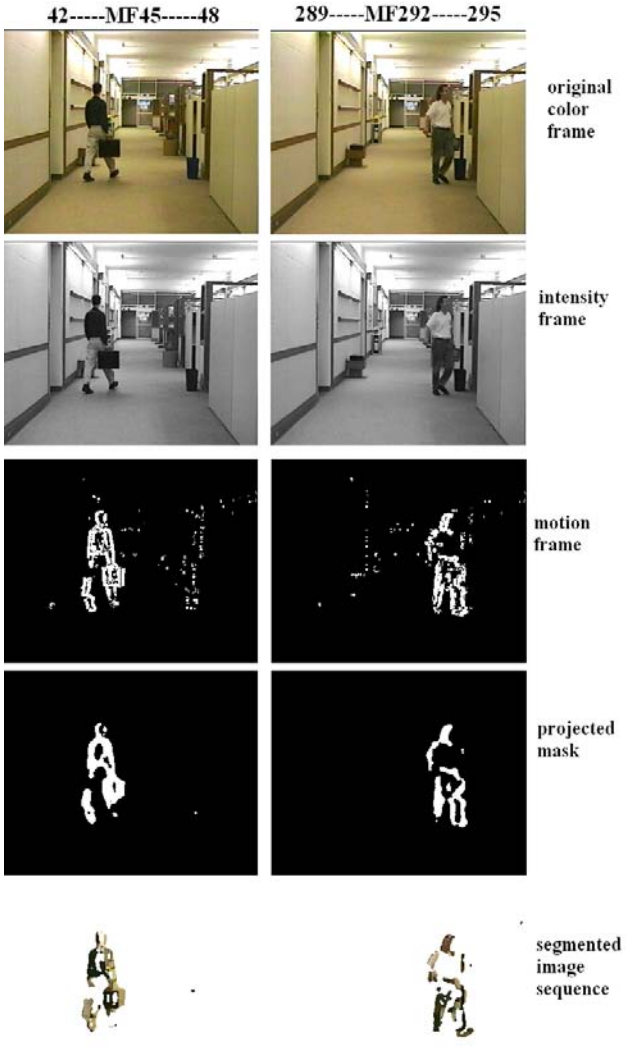


Fig. 2 Human motion detection and segmentation of an indoor image sequence. First row: original color images. Second row: Gray-scale image. Thirst row: binary motion image. Fourth row: Filtered image (back-projected image). Last row: Segmented color image.

Secondly, here $s(x, y, t)$ computes numerically by convolving the intensity history at each pixel with the second derivative in time of the temporal Gaussian. This calculates locate the spatial zero crossing positions in x, y at fixed t.

$$s(x, y, t) = m(t) \otimes p(x, y, t)$$

where

$$m(t) = u \frac{1}{\sigma\sqrt{2\pi}} \times \left(-\frac{u^2}{\sigma^2}\right)\left[1 - \frac{u^2 t^2}{\sigma^2}\right] \times \exp\left[-\frac{u^2 t^2}{2\sigma^2}\right]$$

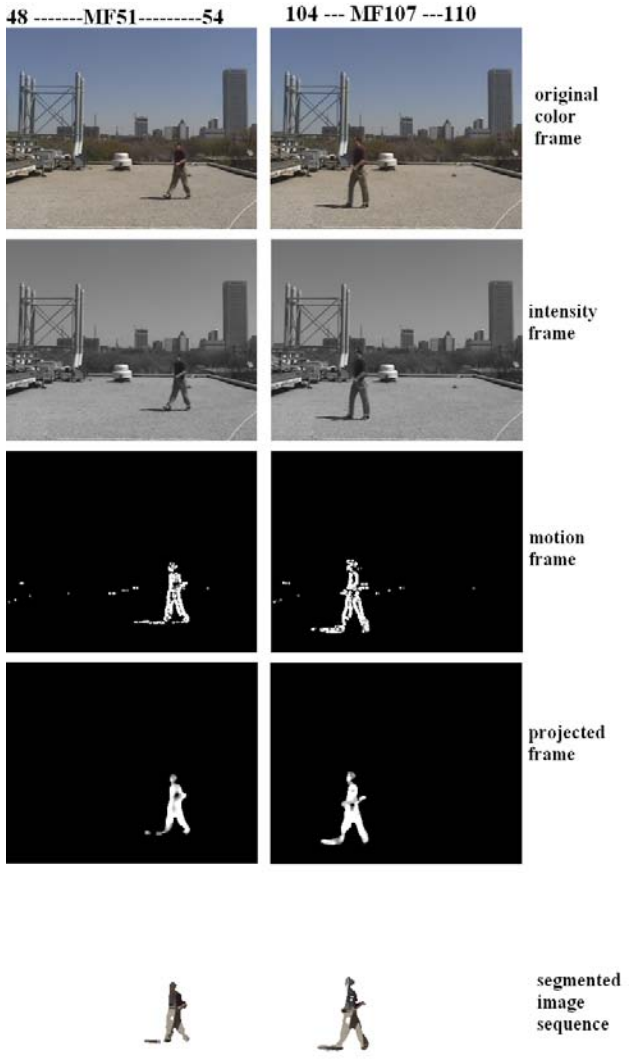


Fig. 3 Human motion detection and segmentation of an outdoor image sequence. First row: original color images. Second row: Gray-scale image. Thirst row: binary motion image. Fourth row: Filtered image (back-projected image). Last row: Segmented color image.

For our experiment, we use 7-consecutive frames to calculate the zero crossing at each pixel. The standard deviation, σ = 1.414 is considered, the time scale, u =0.98. Therefore, we get the middle frame-4 as the reference frame for a sequence of 1-2-3-4-5-6-7 frames from the image sequence; similarly, frame 5 is obtained from the sequence of 2-3-4-5-6-7-8 frames, and so on.

Thirdly, at each zero crossing pixel, the derivatives $s_x = \partial s / \partial x$, $s_y = \partial s / \partial y$, and $s_t = \partial s / \partial t$ are calculated using numerical finite difference method as per equations in previous section. The magnitude of the normal velocity at each zero crossing pixel is computed from equation (4). In this case, we use three consecutive frames in the sequence and the magnitude of the visual motion in the image plane is calculated for the middle frame. The velocity scales within 0-255 and represented as a velocity gray-scaled

frame for visualization in the portable format. The value of NN is the previous section is 3.

Fourthly, a velocity threshold for image velocity is considered for determining binary motion image. The mean velocity threshold is estimated using the following formula

$$\bar{v}_{th} = \frac{\sum_{vel=0}^{255} (vel+1) \times p_{gray}(vel)}{\sum_{vel=0}^{255} p_{gray}(vel)}$$

Where p_{gray} is the gray scale velocity image. The mean threshold value is the arithmetic mean of velocity field histogram. An estimate for the threshold is obtained by trials and final value is chosen for making the binary motion image frame. The motion image frame after using suitable threshold is shown in figure 2 (third row) and 3 (third row). The BMI has some unwanted back and forth motions with the original moving object.

Finally, we apply median filter in each frame for removing the unwanted motion blobs in the BMI frame. The size of the median filter is 7×7 and we get the maximum motion tendency, which is back-projected image, which is shown in figure 2 (fourth row) and figure 3 (fourth row). Pixels belongings to the segmented velocities are back-projected in the original color frame to give a segmented color image. Back projection is performed by sweeping through the original color image in the spatial domain and creating a new image such that only the pixels that belong to one of the segmented velocities are copied from the original image to the new image. These images are shown in figure 2 and figure 3 in the last row. In the segmented image, the human corresponding to the motion is only found because only motion due to human movement is occurred at that frame.

When an image sequence of N frames is analyzed, the convoluted reference image is found for 7-consecutive frames and the motion frame as well as the segmented image frame is image is found as the middle frame from 9-consecutive frames.

IV. Conclusions

In this paper, we presented a technique for detecting and segmenting human motion in image sequences. The appearances of spatial zero-crossing contours of in a single frame indicated the location of moving edge. Due to 1D Gaussian function, the computational cost is low. The zero crossing images were used in the optical flow equation to compute the normal velocities of the zero-crossing contours. The computation of normal velocity is insensitive to spatial and temporal variation of image intensity caused by small illumination effects. Binary motion image was found after making suitable threshold. With filtering, the unwanted motions were removed and the pixels belong to the motion back-projected to the original color frame and get a segmented color image very effectively. Our future works include the precise detection and segmentation of human motion image sequence for correctly recognition their activities and behavior.

References

[1] J. K. Aggarwal, J. K.and Q. Cai, "Human motion analyis: A review," Computer Vision and Image Understanding, vol. 73, no. 3, pp. 420-440, March, 1999.

[2] J. L. Barron, D. J. Fleet, and S.S. Beauchemin, "Performance of optical flow technique," International Journal o March, f Computer Vision, vol. 12, no. 1, pp. 43-77, 1994.

[3] B. F. Buxton and H. Buxton, "Monocular depth perception from optical flow by space time signal processing," Proc. Royal Society of London, UK, B.218, pp.27-47, 1983.

[4] J. H. Duncan and T.C. Chou, "On the detection of motion and computation of optical flow," IEEE Transactions on Patten Analysis and Machine Intelligence, vol. 14, no. 3, pp. 346-352, March, 1992.

[5] B. K. P. Horn and B.G. Schunck, "Determining optical flow,'' Artificial Intelligence 17, p.185-203, 1981.

[6] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation: an error analysis of gradient-based methods with local optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.9, p.229-244, March 1987.

[7] A. D. Marr and A. S. Ullman, "Directional selectivity and its use in early visual processing," in Proc. Royal Society of London, UK, vol. B211, pp.151-180, 1981.

[8] W. H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery, "Numerical Recipes in C: The art of Scientific Computing", Second Edition, Cambridge University press, 2002.

[9] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," Pattern Recognition, vol. 36, pp. 585-601, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A High-Frame-Rate Embedded Image-Processing System by Using a One-chip DSP

M. Fukuzawa, H. Hama, N. Nakamori, and M. Yamada

Graduation School of Science and Technology, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
E-mail: fukuzawa@kit.ac.jp

Abstract – By combining a CMOS digital image sensor with windowing function to a one-chip DSP with programmable peripherals, a high-frame-rate embedded image-processing system has been developed to measure the target position. The system can capture the VGA format image (640×480 pixels) at the rate of 83 frames/sec, exceeding the conventional video rate of 30 frames/sec. The capture rate can be easily increased by using the windowing function in which the number of pixels to be captured is decreased. When a simple algorithm is examined for measuring the coordinates of the gravity center in an image with luminous targets, the processing time is always less than the capturing time for any size of the window. Consequently, it is confirmed that the system performance is inversely proportional only to the number of pixels in the image windowed. This simple relationship without any bottleneck is useful for high-speed motion control of mobile equipments such as cargo vehicles and robots.

I. Introduction

Image processing system based on digital-signal processor (DSP) is widely used in industrial and medical applications because most sophisticated algorithms can be implemented more easily than in the FPGA and it can be executed more effectively than in the general-purpose processors. Harvey *et. al.* presented a DSP-based parallel image processing system for on-line and real-time applications [1]. Zhuo *et. al.* developed a real-time infrared image processing system based on DSP technology [2]. Coffey *et. al.* presented a development platform for real-time image processing based on the ADSP-BF533 Blackfin embedded DSP and the MicroC/OS-II real-time operating system [3]. Li *et. al.* discussed a kind of real-time image processing system designed with high-speed DSP and FPGA for target recognition and track [4]. In order to assist medical diagnosis by detecting periodical tissue-motion from ultrasonograph movie, Fukuzawa *et. al.* have also been developed several image-processing systems with various architectures: a DSP-based system for full real-time processing of a small region of interest (ROI) at the rate of 30 frames/sec (fps) for 64×64 ROI [5], a PC-based system for semi real-time processing at 5 fps for the VGA format image [6, 7], and one-dimensional processor-array system for full real-time processing at 30 fps for the VGA format image [8], a PC-based system for full real-time processing by using the Streaming SIMD extensions at 30 fps for the VGA format image [9]. Most of those systems are based on a standard PC as fixed equipment and the DSP acts as an accelerator to perform sophisticated algorithms at the rate of input video.

As well as for the fixed equipments, image-based measurement is very useful for several mobile equipments such as cargo vehicles and carrier robots in the factory because it can measure any targets if it is appeared within the vision field. In such applications, high-frame-rate of image processing is very important to control their motion at high speed. Furthermore, the size and the power dissipation of the image processing system are limited depending on the size and the power source of the mobile equipment. On the other hand, the measurement algorithm can be simplified because the target shape can be determined for convenience of the algorithm in advance. Therefore, it is effective for factory use to develop a high-frame-rate image processing system as a small unit with low power-dissipation.

In this paper, we propose a high-frame-rate image processing system by combining a CMOS digital image sensor to a one-chip DSP. The architectural features of the proposed system are discussed by comparing it with the PC-hosted system. The system performance is experimentally evaluated with an algorithm to measure the position of luminous targets.

II. Comparison of System Architectures

Here we compare two architectures of DSP-based image processing system; (A) a PC-hosted system which is a representative of the DSP-based systems mentioned above, and (B) a high-frame-rate embedded image-processing system which is proposed here.

A. PC-hosted Image Processing System

The major aim of the DSP-based systems mentioned above was to perform sophisticated algorithm at the desired rate with reasonable cost. For this purpose, some design concepts were introduced as follows:

- The system was based on a fixed PC.
- The capture rate is fixed to obtain all the pixels of the camera.

- The desired process-performance was obtained by increasing the number of DSPs.
- In addition to the process result such as measurement value, the processed image was sometimes transferred or displayed.

Figure 1 shows a block diagram of typical PC-hosted image processing system. The primary component of the system is a PC with an expansion card equipped with the DSP and the peripherals. The expansion card is essentially composed of a DSP, an external memory (static RAM in most cases), a control logic, an input video interface with an analog-to-digital converter (ADC) or a low voltage differential signalling (LVDS) interface, and an output video interface with digital-to-analog converter (DAC). In order to increase the system performance, some of the system can add several sets of the DSP and the external memory. In some cases, added DSP sets are on the other expansion card and they are connected directly through the local bus. The input video interface is connected to a CCD camera to capture the video signal repeatedly. The output video interface is connected to a display to monitor the input and/or processed images. The processing results such as the measured values can be indicated to the display or, in the factory, transmitted to a master controller for production control through a serial or LAN interface built in the PC motherboard. In general, most of the systems are designed for general-purpose with high flexibility. However, in some practical use, it becomes sometimes inevitable to adjust the process parameters depending on the targets.

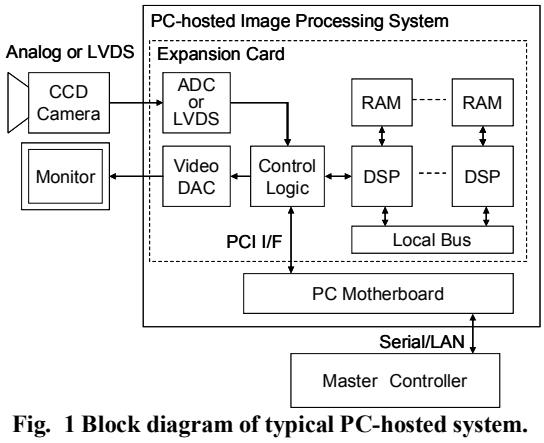


Fig. 1 Block diagram of typical PC-hosted system.

B. High-Frame-Rate Embedded Image-Processing System

We proposed here a high-frame-rate embedded image-processing system aimed to process the image as fast as that it is captured. Essential design concepts of the system are as follows:

- The system is implemented as a small unit with low power consumption.
- The capture rate can be easily increased by windowing the image to decrease the number of pixels to be captured.
- The process-performance is fixed by combining a DSP for an image sensor.
- The processed image is never transferred or displayed.

Figure 2 shows a block diagram of the proposed system. It consists of a Analog Devices ADSP-BF537 one-chip DSP, an external SDRAM of PC133 standard, and a Micron MT9V022 CMOS digital image sensor.

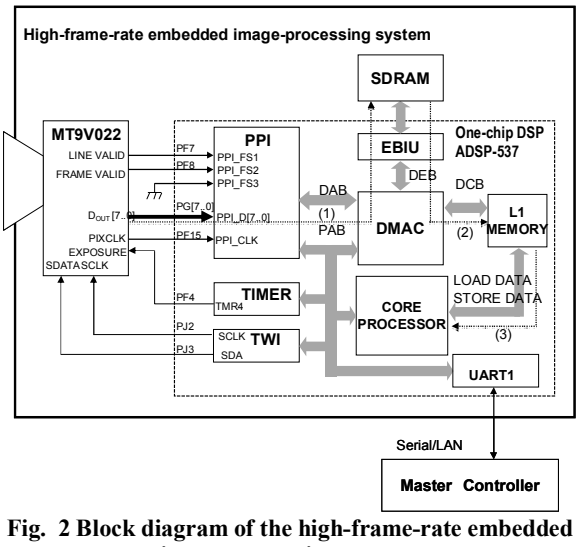


Fig. 2 Block diagram of the high-frame-rate embedded image-processing system.

The ADSP-BF537 is one-chip DSP with on-chip memory and variety of programmable peripherals. It includes the following functionalities.

- The DSP processor core contains two16-bit multiplier/accumulator units (MACs), two 40-bit arithmetic logic units (ALUs), 8-bit video ALUs, and a 40-bit shifter. It combines the MAC-based signal processing engines, RISC-like instruction set, single instruction multiple data (SIMD) instructions, and multimedia features into a single instruction set architecture. In the proposed system, the core runs with the clock of 600 MHz.
- On-chip memory includes three blocks of 64Kbyte level 1 (L1) instruction memory, 64Kbyte L1 data-memory, and 4Kbyte L1 scratchpad RAM. They can be accessed at full processor speed.
- Parallel peripheral interface (PPI) is a half-duplex, bidirectional port accommodating up to 16 bits of data. Because of a dedicated clock pin and frame sync pins, it can connect directly to parallel A/D and D/A converters, video encoders and decoders. In the proposed system, it is directly connected to the parallel LVDS video-data port of the CMOS image sensor.

- DMA controllers include twelve peripheral-DMA channels between memory and on-chip peripherals, two memory-DMA channels between memory and memory. It supports both one-dimensional (1-D) and two-dimensional (2-D) transfers. The 2-D transfer can define arbitrary row and column sizes up to 64K by 64K elements with arbitrary step sizes up to ±32K elements. The transfer can be configured not only by command registers but also by descriptors stored within the memory. The descriptor-based DMA can chain multiple DMA transfers with different configurations without interrupt to the processor. In the proposed system, the 2-D DMA transfer configured by the descriptor is used for transferring the image windowed to a certain region of the external memory.
- The external bus interface unit (EBIU) gives a direct interface to the external memories. It is compliant with the PC133 SDRAM standards. In the proposed system, it is connected to the 64MB external SDRAM.
- Two wire interface (TWI) is an interface to an inter IC (I^2C) bus. In the proposed system, it is connected to the CMOS image sensor to read and write its control registers.
- General-purpose Timer module includes eight identical 32-bit timers with the output pin. In the internal clock mode, the clock source is the processor's peripheral clock of 120MHz. In the proposed system, one of the timer output is connected to the exposure control input of the CMOS image sensor to determine the starting time of exposure precisely.
- Universal asynchronous receiver and transmitter (UART) supports the conventional serial interface such as EIA-232C, EIA-422, or EIA-485. In the proposed system, it is used to transmit the processing results to the master controller.
- Ethernet media access controller (MAC) supports the direct connection to a network. It is not used yet in the proposed system, but substitutive for the UART depending on the interface to the master controller.

The MT9V022 is 1/3-inch wide-VGA (752×480) CMOS digital image sensor. The major features are as follows:

- Square pixel and global shutter, which is suitable for image-based measurement in the mobile equipments,
- 8- or 10-bits on-chip ADC and parallel LVDS interface, which can connect to the PPI directly,
- Two-wire serial interface to access the command registers,
- Programmable windowing function in which the number of pixels to be transferred is decreased. Since the data rate is maintained in any window size, the frame rate can be increased by windowing.

Because of the architectural parallelism, the system can perform three operations concurrently: (1) a *capture* operation which includes the transmission of the image data through the LVDS parallel port from the CMOS image sensor to the PPI, and the data-transfer with the DMA from the PPI to the external SDRAM, (2) a *data-transfer* operation which includes the DMA transfer of the image data to be processed from the external SDRAM to the internal L1 memory, and (3) a *processing* operation of the image data transferred in advance to the internal L1 memory. Unlike the PC-hosted system, this system does not have any path to transfer processed image to outside except for the remote debugger interface. Figure 3 shows a picture of the prototype system.

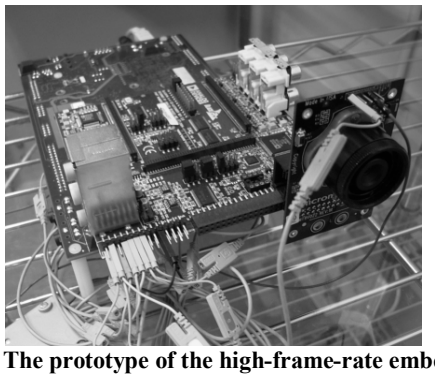


Fig. 3 The prototype of the high-frame-rate embedded image-processing system.

Figure 4 shows examples of the image windowed with 752 x 102 pixels captured by the prototype system: (a) normal image without any filter and (b) IR image taken through the IR-pass filter. The target is a front panel of a cargo vehicle with two IR-luminous markers at its left and right corners. The normal image reveals the luminous markers together with the shape of the front panel itself and something in the background. On the other hand, the IR image reveals only the luminous markers. Therefore, the IR image is drastically easy to measure the positions of the luminous markers. As *processing operation* for the proposed system, an algorithm to measure the position of luminous targets was implemented by calculating the coordinates of the gravity center in an image. This process consists of simple sum of products but is very effective to measure the position of the luminous markers in a non-textured background as shown in Fig. 4 (b).

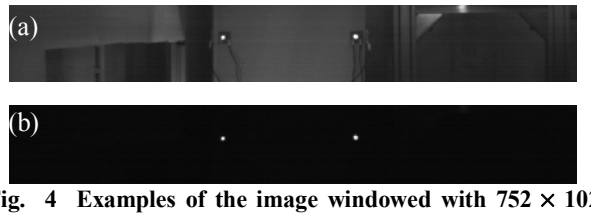


Fig. 4 Examples of the image windowed with 752 × 102 pixels captured by the prototype system: (a) normal image and (b) IR image through the IR-pass filter.

III. Evaluation of System Performance

The system performance was experimentally estimated by measuring the execution time of each system operation. In order to measure the execution time precisely without affecting the operation, the general-purpose 32-bit timer not used in any operation was applied. A small code set was added into the program so that the timer was enabled before all the operation and the value of the timer counter register was saved into the internal memory at the beginning and ending of every operation. In the experiment, three operations were sequentially repeated several times on the system. After halting the operation sequence, the values of the timer counter register were obtained through the remote debugger and consequently the averaged increment of the timer counter register was calculated for each operation. The resolution of the timer was 8 ns with the clock source of 120 MHz. The required time to save the timer count register, which consists of 8 instructions, was about 13 ns at the core clock of 600 MHz. Therefore, if the execution time was obtained in units of microsecond, it is reasonable that the timer had enough resolution and so-called 'probe effect' of code adding was negligible to measure the execution time of the operation.

Table 1 shows the execution time of the capture and the data-transfer operations measured in the VGA format image. It reveals that the capture rate attains to 83 fps at the VGA format image, which substantially exceeds the conventional video rate of 30 fps. Furthermore, the data-transfer rate exceeds 8 times of the capture's one. Since the external SDRAM has enough size to store multiple VGA images, the pixel values of the multiple frames can be transferred to the DSP's internal memory during capturing a frame. Therefore, as well as the spatial distribution of the pixel values in a single frame, the variation of pixel values in time series can be analyzed with the system.

Table 1 Execution time of the capture and data-transfer operations measured in the VGA formatted image.

Operation	Execution Time [μs/frame]	Operation Rate [frame/sec]
Capture	12026	83
Data-Transfer	1480	675

Table 2 shows the execution time measured in a image windowed with 752 x 102 pixels which is equivalent number of pixels to the QVGA format. The capture and data-transfer rates become four times as fast as those in the VGA format, which is inversely proportional to the number of pixels. By examining the execution time in several images windowed with different size, it was confirmed that the inverse proportion is maintained unless the width of the image is reduced. Therefore, the capture rate can be easily increased by using the windowing function in this system.

Table 2 Execution time of the capture and data-transfer operations measured in the image windowed with 750 × 102 pixels.

Operation	Execution Time [μs/frame]	Operation Rate [frame/sec]
Capture	3306	332
Data-Transfer	370	2702

The processing time to measure the position of luminous targets was also examined in several images windowed with different size. It was always less than the capturing time for any size of the window. Figure 4 shows the achievable frame rate of each system operation. It was revealed that the system performance was inversely proportional only to the number of pixel in the image windowed when all the operations were performed concurrently. The simple relationship between the system performance and the number of pixel without any bottleneck is useful for high-speed motion control of mobile equipments such as cargo vehicles and carrier robots.

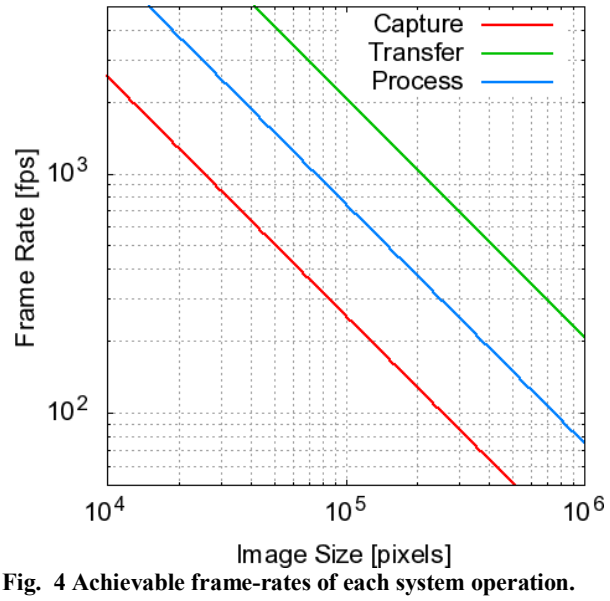


Fig. 4 Achievable frame-rates of each system operation.

IV. Conclusions

By combining a Micron MT9V9022 CMOS digital image sensor to a Analog Devices ADSP-BF537 one-chip DSP, a high-frame-rate embedded image-processing system has been developed to measure the target position. The system can capture the VGA format image (640×480 pixels) at the rate of 83 fps, exceeding the conventional video rate of 30 fps. The capture rate can be easily increased by using windowing function in which the number of pixels to be captured is decreased. When a simple algorithm is examined for measuring the coordinates of the gravity center in an image with luminous targets, the processing time is always less than the capturing time for any size of the window. Consequently, it is confirmed that the system performance is inversely proportional only to the number

of pixels in the image windowed. This simple relationship without any bottleneck is useful for high-speed motion control of mobile equipments such as cargo vehicles and robots.

References

[1] D. M. Harvey, S. P. Kshirsagar, and C. A. Hobson, "Low cost scaleable parallel image processing system", Microprocessors and Microsystems, vol. 25, no. 3, pp. 143–157, 2001.

[2] Hongyan Zhuo, Rong Zhang, Zai-Ming Li, Zhi-zhong Fu, Jian-tao Wu, Zheng-dong Li, and Yingsong Song, "Rea-time infrared image processing system based on DSP technology", in Proc. SPIE, vol. 5640, pp. 124–129, 2005.

[3] S. Coffey and J. Connell, "A customizable system for real-time image processing using the Blackfin DSProcessor and the MicroC/OS-II real-time kernel," in Proc. SPIE, vol. 5823, pp. 245–257, 2005.

[4] YanLi Han and Heng Li, "Real time image processing system," in Proc. SPIE, vol. 6625, 662517, 2008.

[5] M. Fukuzawa, M. Yamada, N. Nakamori, Y. Kitsunezuka and N. Kanamori, presented at 105th meeting of Japan Society of Medical Imaging and Information Sciences [in Japanese], 1992.

[6] M. Fukuzawa, Y. Kitsunezuka, and M. Yamada "Real time detection and display of artery pulsation in neonatal cranial ultrasonogram", Medical and biomedical engineering and computing, vol. 35, Supp. Part I, p. 403, September 1997.

[7] M. Fukuzawa, Y. Kitsunezuka, and M. Yamada, "A real-time processing system for pulsation detection in neonatal cranial ultrasonogram", Jpn. J. Appl. Phys. vol. 37, pp.3106-3109, 1998.

[8] M. Yamada, M. Fukuzawa, and Y. Kitsunezuka, "One-Dimensional Processor Array System for Fast Analysis of Tissue Motion in Ultrasonogram", Jpn. J. Appl. Phys., vol. 38, pp.3385-3387, 1999.

[9] M. Fukuzawa, M. Yamada, N. Nakamori, and Y. Kitsunezuka, "Real-time visualization of pulsatile tissue-motion in B-mode ultrasonogram for assistance in bedside diagnosis of ischemic diseases of neonatal cranium", in Proc. SPIE, vol. 6920, 69200S, March 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Modelling of the Video DCT Coefficients

M. I. H. Bhuiyan[1] and Rubaiya Rahman[2]
[1]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[2]Department of Computer Science and Engineering
Brac University, Dhaka-1216, Bangladesh
E-mail: imamul@eee.buet.ac.bd

Abstract— **In this paper, the normal inverse gaussian probability density function (PDF) is presented as a highly suitable prior for modelling the DCT coefficients of videos. The parameters of the proposed prior are estimated by minimizing the Kullback-Leibler divergence between the prior and the empirical PDF obtained from the DCT coefficients of digital videos. The effectiveness of the parameter estimation technique is demonstrated through Monte carlo tests. Experiments are carried out to study the effectiveness of the proposed prior in modelling both the full-frame and block-DCT coefficients of video data, and compare it with that of the generalized Gaussian, Laplacian and Bessel K form PDFs. It is shown that in general the normal inverse Gaussian PDF is a better model than the other PDFs.**

I. Introduction

The use of digital video is widespread now-a-days ranging from entertainment to medical applications. Various algorithms have been developed in recent times for the storage, transmission and analysis of digital video. Practical video processing techniques often use the discrete cosine transform (DCT) due to its excellent signal decorrelation property, thus enabling the representation of the signal information in terms of as few coefficients as possible, and availability of fast DCT algorithms for real time implementations [1]. For example, in MPEG-2, one of the most commonly used video coding standard, the DCT is employed for intra-frame compression. Each video frame is divided into 8×8 blocks and subsequently the blocks are subjected to the DCT. Since, the DCT has the ability to compact the signal energy into a few coefficients, most of the DCT coefficients of an image block are small and can be quantized to zero. Thus, the intra-frame signal information can be stored by using a small number of coefficients. The inter-frame information can also be decomposed using the DCT and the corresponding coefficients subsequently compressed as the intra-frame coefficients [2]. However, for efficient quantization of the DCT coefficients, it is necessary to know about the statistics of these coefficients. More specifically, the quantization step size and the corresponding decision levels can be efficiently designed by incorporating the probability distribution of these coefficients in the integrals for obtaining these values. For this purpose, it is essential to assume a certain probability distribution function (PDF) to model the DCT coefficients. Note that the knowledge about the distribution of the DCT coefficients might also be useful for designing other video processing systems such as video watermarking, since each frame can be considered as an individual image and one can exploit it to spread the hidden message in a robust and efficient manner in the frames [3].

A number of PDFs have been proposed in the literature for modelling the DCT coefficients that include the Laplacian, generalized Gaussian (GG), alpha-stable and generalized gamma PDFs [4]–[8]. In [4], the Laplacian PDF is used to model the DCT coefficients of natural images. The drawback of the Laplacian PDF is its inability to capture the tail information. In [5], a generalized Gaussian (GG) PDF is shown to be a better model than the Laplacian one. Note that the Gaussian and Laplacian PDFs are two special cases of the GG PDF. The suitability of the GG PDF in modelling the DCT coefficients is further illustrated in [6] through a detailed mathematical analysis. In [7], the generalized Gamma PDF is shown to be a more suitable prior for modelling the DCT coefficients as compared to the GG PDF. The α-stable PDF is proposed for modelling the DCT coefficients in [8] and using this, an efficient watermarking system is developed. A limitation of the α-stable PDF is that it does not have a closed-form expression, making the parameter estimation complicated especially from noisy data and increases the complexity of the parameter estimation process [9]. In fact, in [8], the Cauchy PDF, a special case of the α-stable PDF is employed to develop the watermarking system. Naturally, a more logical and desirable approach would be to use a generalized PDF that is not only highly appropriate for modelling the DCT coefficients, but also has a closed-form expression and involves less computational complexity for parameter estimation.

In this paper, a symmetric normal inverse Gaussian (SNIG) PDF is proposed for modelling the DCT coefficients corresponding to the digital video frames. The parameters of the SNIG PDF are obtained by minimizing the Kullback-Leibler (KL) divergence [10] between the SNIG PDF and that corresponding to the DCT coefficients. Monte Carlo simulations are carried out to study the efficiency of the parameter estimation technique. The effectiveness of the SNIG PDF in modelling the DCT coefficients of video frames is investigated through extensive simulations using various standard digital videos. Both block- and full-frame DCT coefficients are utilized in the simulations. The paper is organized as follows. The SNIG PDF and its properties are briefly discussed in Section II. The parameter estimation method is described in Section III. The modelling of the DCT coefficients is discussed in Section IV and some concluding remarks given in Section V.

II. THE SNIG PDF

The SNIG PDF is expressed as

$$P_X(x) = \frac{A(\delta,\alpha)K_1(\alpha\sqrt{\delta^2+x^2})}{\sqrt{\delta^2+x^2}} \tag{1}$$

where

$$K_\lambda(\xi) = \frac{1}{2}\int_0^\infty z^{\lambda-1}\exp(-\frac{1}{2}\xi(z+z^{-1}))dz \tag{2}$$

and $A(\delta,\alpha) = \frac{\delta\alpha}{\pi}\exp(\delta\alpha)$ [11]. The α parameter controls the steepness of the distribution; the larger the values of α, the higher the peak of the PDF. The increase in α also influences the shape of the tails, since with a sharper peak, the tails would become lighter. The other parameter, δ, defines the scale of the PDF and is similar to the variance parameter of a Gaussian PDF. Fig. 1 shows the variation in the shape of the SNIG distribution for various values of α.

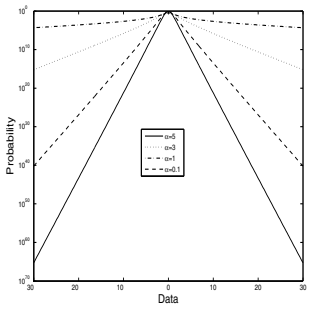


Fig. 1. The effect of changing α on the shape of the SNIG PDF. The value of δ is set to 1. The vertical axis is log-normalized.

III. SNIG PARAMETER ESTIMATION

The parameters of the SNIG PDF are estimated from the DCT coefficients by minimizing the KL divergence between the SNIG PDF and the empirical PDF corresponding to the data. The KL divergence, also known as the relative entropy, is given by

$$KL(P_{emp},P) = \int P_{emp}(x)\log_2\frac{P_{emp}(x)}{P(x)}dx \tag{3}$$

where P and P_{emp} represent the SNIG and empirical PDFs, respectively [10]. Minimizing the KL divergence is equivalent to maximizing the log-likelihood, since the latter is a negative of the sum of the KL divergence and the data entropy [10], and thus can be expected to provide unbiased estimation of the parameters asymptotically. The parameters of the SNIG PDF are obtained as

$$\widehat{\alpha},\widehat{\delta} = \arg\min_{\widehat{\alpha},\widehat{\delta}}\sum_{i=1}^{N_h} P_{emp}(x_i)\log_2\frac{P_{emp}(x_i)}{P(x_i)} \tag{4}$$

where x_i denotes the center values of the bins of a N_h-point histogram corresponding to the data, the minimization being carried out using the Nelder-Mead direct search technique [12].

TABLE I

MEAN OF THE ESTIMATED VALUES OF α AND δ. THE STANDARD DEVIATIONS ARE GIVEN IN THE PARENTHESES. A AND E STAND FOR ACTUAL AND ESTIMATED.

α		δ	
A	E	A	E
10	11.4132 (11.2283)	2	2.2817 (2.2051)
10	10.2628 (1.3339)	1	1.0255 (0.1321)
1	1.0109 (0.0560)	1	1.0102 (0.0417)
0.5	0.5020 (0.0364)	1	1.0071 (0.0399)
0.05	0.0557 (0.0122)	1	1.0099 (0.0279)

The effectiveness of the parameter estimation technique is tested using Monte Carlo simulations. For this purpose, a data set containing 10000 SNIG distributed samples is generated 1000 times. Each time, the parameters are estimated from the corresponding data. Finally, the mean and standard deviations of the estimated values are calculated. A SNIG distributed data set can be generated as [13]

$$X = R\sqrt{Y} \tag{5}$$

In (5), R and Y represent data that follow Gaussian and inverse Gaussian distributions, respectively. The corresponding PDFs are given by

$$P_R(r) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{r^2}{2}) \tag{6}$$

$$P_Y(y) = C(\varepsilon,\rho)exp\left(-\frac{\varepsilon+\rho y^2}{2y}\right) \tag{7}$$

where $C(\varepsilon,\rho) = \sqrt{\frac{\varepsilon}{2\pi y^3}}e^{\sqrt{\varepsilon\rho}}$, $\varepsilon = \delta^2$ and $\rho = \alpha^2$. Table I shows the mean and standard deviation calculated for various values of α and δ. Note that the mean values are close to the actual values. The corresponding standard deviations are small especially when the values of α are small.

IV. MODELLING OF THE DCT COEFFICIENTS

In this section, we discuss the modelling of the DCT coefficients of digital video. Each frame can be considered as a 2-D image. The DCT of a 2-D data $F(x,y)$ of size $N\times N$ is given by

$$G(u,v) = B_uB_v\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{8}$$

where B_u is $\sqrt{1/N}$ and $\sqrt{2/N}$, for $u=0$ and $u>0$, respectively. Similar conditions hold for B_v. In this paper, blocks of size 8×8 are used because of their prevalence in digital video applications. The block-DCT coefficients with index (u,v) are denoted as $C_{u,v}$. Experiments are conducted using the DCT coefficients of various standard video sequences to study the goodness-of-fit of the SNIG PDF. To make sure that the findings of the experiments are general enough, video sequences with different characteristics

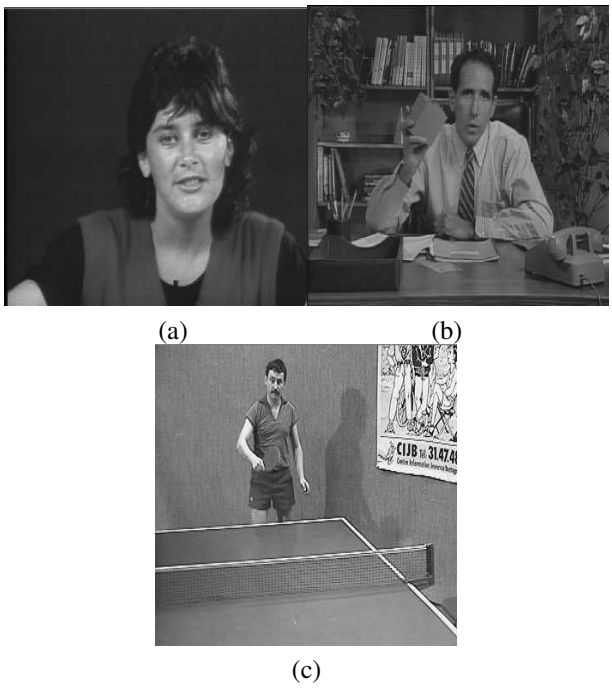


Fig. 2. Sample frames from digital video sequences used in our experiments: (a) *Miss America*, (b) *Salesman* and (c) *Tennis*.

such as high motion and low motion are used. The video sequences are collected from the website of the Center for Image Processing Research, Rensselaer Polytechnic Institute (http://www.cipr.rpi.edu/resource/sequences/index.html). Fig. 2 shows sample frames of some of the video sequences used in the experiments. For a particular video sequence, the distributions of the block-DCT coefficients of the corresponding frames are fitted with the SNIG, Laplacian, GG and Bessel K form (BKF) [15] PDFs. The parameters of the GG and BKF PDFs are obtained using the methods described in [14] and [15], respectively. The method in [14] is used for estimating the GG parameters due to its ability to provide estimates that are about as good as the maximum-likelihood ones, while incurring a considerably less computational complexity [14]. For a particular frame of a video sequence, the goodness-of-fit of the various PDFs are compared using Kolmogorov-Smirnov (K-S) distance [12]. The K-S distance is given by

$$d_{KS} = \max_{x \in \mathcal{R}} |P_F(x) - P_{Fe}(x)| \quad (9)$$

where d_{KS}, $P_{Fe}(x)$ and $P_F(x)$ denote the K-S distance, the cumulative density function (CDF) of the prior PDF and the empirical CDF, respectively. Figs. 3, 4 and 5 show the plots of values of the K-S distance for different frames of the video sequences, *Miss America*, *Salesman* and *Tennis* for various PDFs. It can be seen from these figures that the SNIG PDF, in general gives lower values of the K-S distance as compared to the other PDFs, thus indicating a better fitting to the empirical PDFs. Similar results are observed for other video sequences, but not reported for space consideration. Note that the goodness-of-fit provided by the SNIG PDF becomes better for the mid-frequency DCT coefficients. This might be useful for developing efficient video watermarking methods by embedding the hidden message in the mid-frequency coefficients [16]. The goodness-of-fit of the SNIG PDF is further illustrated in Fig. 6 that shows the plots of the empirical and the corresponding SNIG, Laplacian, GG and BKF PDFs for different frames of the *Tennis* video sequence. It is seen that the SNIG PDF matches the empirical ones more closely than the other PDFs.

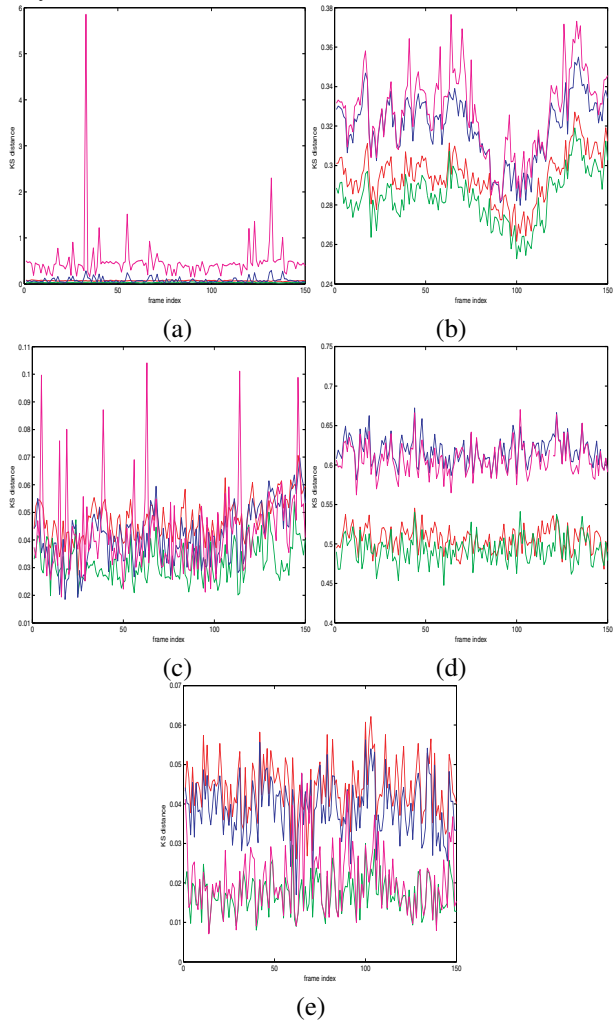


Fig. 3. Plot of the values of the K-S distance for various block-DCT coefficients of the *Miss America* video sequence: (a) C_{01}, (b) C_{10}, (c) C_{22}, (d) C_{50} and (e) C_{54}. Plots of the Laplacian, SNIG, GG and BKF PDFs are shown using red, green, blue and magenta colored lines.

We have also carried out experiments using the full-frame DCT coefficients of various video sequences. Fig. 7 shows the plots of K-S distances for various PDFs for the *Miss America*, *Salesman* and *Tennis* video sequences. It is observed from this figure that the values of the K-S distance corresponding to the SNIG PDF is smaller than those of the other PDFs for

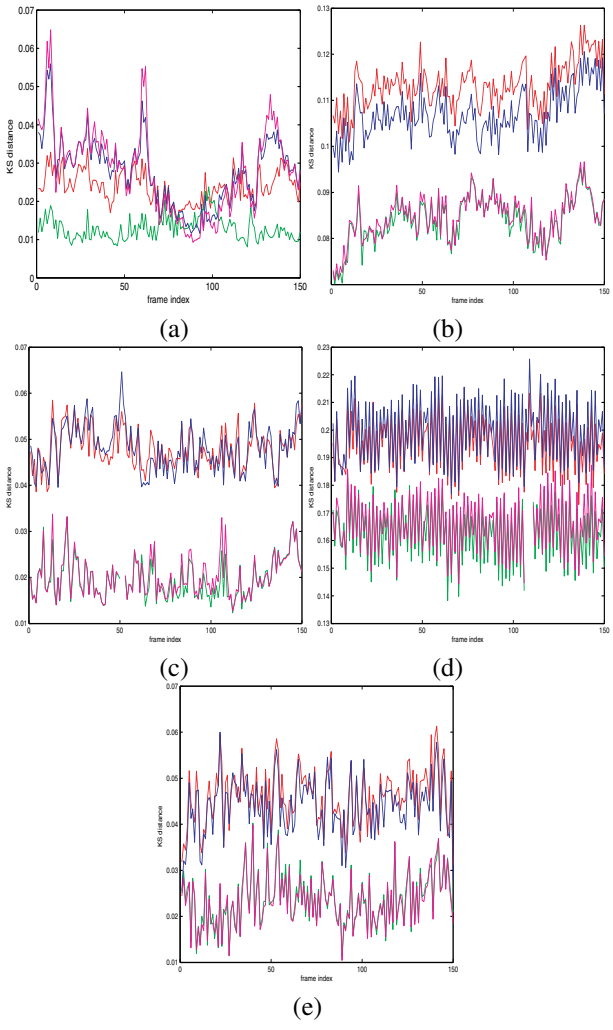


Fig. 4. Plot of the values of the K-S distance for various block-DCT coefficients of the *Salesman*: (a) C_{01}, (b) C_{10}, (c) C_{22}, (d) C_{50} and (e) C_{54}. Plots of the Laplacian, SNIG, GG and BKF PDFs are shown using red, green, blue and magenta colored lines.

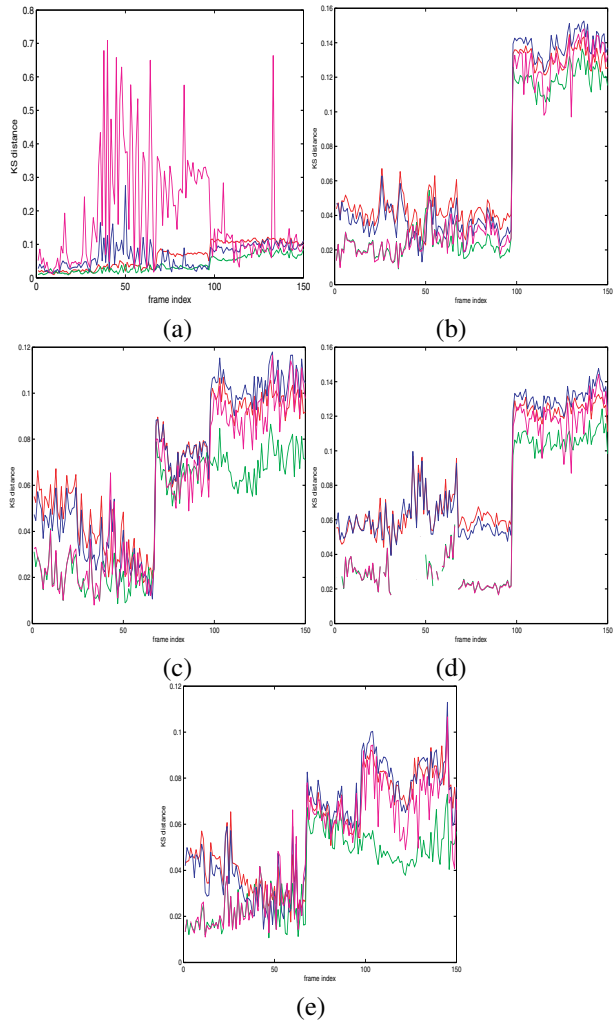


Fig. 5. Plot of the values of the K-S distance for various block-DCT coefficients of the *Tennis*: (a) C_{01}, (b) C_{10}, (c) C_{22}, (d) C_{50} and (e) C_{54}. Plots of the Laplacian, SNIG, GG and BKF PDFs are shown using red, green, blue and magenta colored lines.

different frames, thus indicating its superiority in modelling the full-frame DCT coefficients of digital video sequences. This observation is further confirmed in Table II that provides the values of the K-S distances for various video sequences for different PDFs. It can be seen that the SNIG PDF gives lower K-S distances for all the video sequences as compared to the other PDFs. The computational cost of the method for estimating the SNIG parameters is also small. For 240×352-sized video frames, the average CPU time taken (in MATLAB and on a 1.66 GHz Pentium Dual core computer) for estimating the SNIG PDF are 0.218 secs and 0.255 secs, respectively, for the block-DCT and full-frame DCT coefficients.

V. Conclusion

The DCT is widely used in various digital video processing methods. It might be possible to develop efficient video processing techniques by incorporating the statistics of the DCT coefficients of the video data in the process of development. In this paper, a symmetric normal inverse Gaussian (SNIG) PDF has been presented as a highly suitable prior for modelling the intra-frame DCT coefficients of digital videos. The parameters of the SNIG PDF has been estimated by minimizing the Kullback-Leibler (KL) divergence between the empirical and prior PDFs. Extensive simulations have been carried out to study the effectiveness of the SNIG PDF in modelling the block-DCT coefficients of digital video sequences. It has been demonstrated that the SNIG PDF is highly appropriate for modelling the block-DCT coefficients. The SNIG PDF has also been shown to be highly effective for modelling the full-frame DCT coefficients of digital videos. It is expected that the findings of this work would be helpful for researchers to develop efficient DCT-based methods for digital videos.

References

[1] T. Sikora, "Digital Video Coding Standards and Their Role in Video Cimmunications," in *Signal Processing for the Multimedia*. IOS Press, 1999.

[2] Y. Wang, J. Ostermann and Y. Zhang, *Video Processing and Communication*. USA: Prentice Hall Ltd., 2002.

[3] M. Barni *et al.*, "Capacity of full frame DCT image watermarks," *IEEE Trans. on Image Processing*, vol. 12, pp. 1450–1455, 2000.

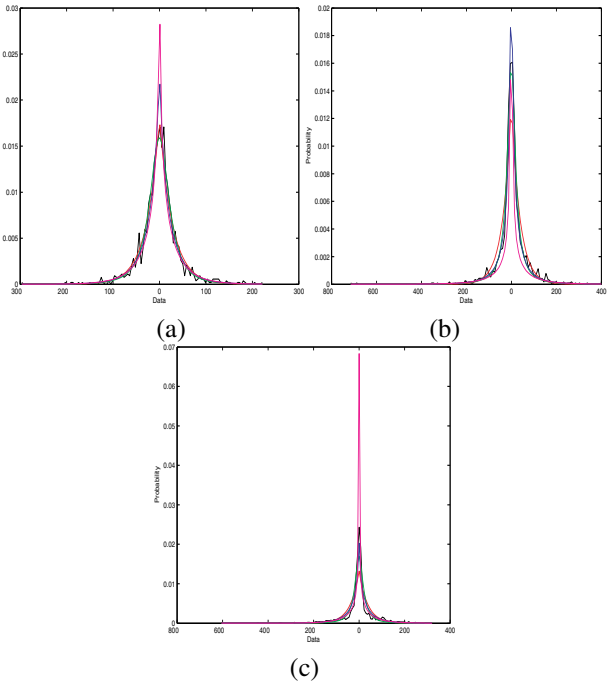


Fig. 6. Plots of the empirical, SNIG, Laplacian, GG and BKF PDFs for the C_{01} block-DCT coefficients of the *Tennis* video sequence. (a) 25^{th} frame, (b) 85^{th} frame, and (c) 105^{th} frame.

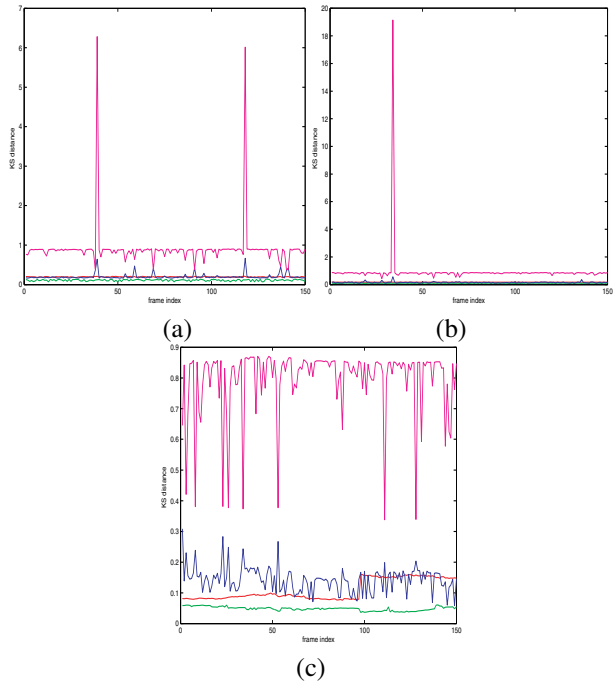


Fig. 7. Plot of the values of the K-S distance for the full-frame DCT coefficients of different video sequences : (a) *Miss America*, (b) *Salesman* and (c) *Tennis*. Plots of the Laplacian, SNIG, GG and BKF PDFs are shown using red, green, blue and magenta colored lines.

TABLE II

VALUES OF THE K-S DISTANCES FOR THE FULL-FRAME COEFFICIENTS OF VARIOUS PDFS FOR SEVERAL VIDEO SEQUENCES.

Miss America	
PDF	d_{ks}
SNIG	0.1091
Laplacian	0.1943
GG	0.2006
BKF	0.9091
Salesman	
SNIG	0.0959
Laplacian	0.1688
GG	0.1524
BKF	0.9428
Tennis	
SNIG	0.0483
Laplacian	0.1091
GG	0.1425
BKF	0.7934
Susie	
SNIG	0.1308
Laplacian	0.2271
GG	0.1241
BKF	0.8246
Football	
SNIG	0.0241
Laplacian	0.0974
GG	0.1425
BKF	0.7934

[4] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients of natural images," *IEEE Trans. on Communication*, vol. COM-31, pp. 325–328, 1983.

[5] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electronics Letters*, vol. 28, pp. 1935–1936, 1993.

[6] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, pp. 1661–1666, 2000.

[7] J. Chang *et al.*, "Image probability distribution based on generalized gamma function," *IEEE Signal Processing Letters*, vol. 12, pp. 325–328, 2005.

[8] A. Briassouli *et al.*, "Hidden messages in heavy-tails: DCT-domain watermark detection using alpha-stable models," *IEEE Transactions on Multimedia*, vol. 7, pp. 700–715, 2005.

[9] M. I. H. Bhuiyan *et al.*, "Spatially-adaptive wavelet-based method using the Cauchy prior for denoising the SAR images," *IEEE Tran. on Circuits, Systems and Video Technology*, pp. 500–507, 2007.

[10] J. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Processing Letters*, vol. 4, pp. 112–114, 1997.

[11] A. Hanssen and T. A. Oigard, "The normal inverse Gaussian distribution for heavy-tailed processes," in *Proc. IEEE-EUEASIP Workshop on Nonlinear Signal and Image processing*, 2001.

[12] W. H. Press, S. A. Teukolsky, W. T. Vellerling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*. UK: Cambridge University Press, 1999.

[13] T. H. Rydberg, "The normal inverse Gaussian Levy process: Simulation and approximation," *Communication in Statistics-Stochastic Models*, vol. 13, pp. 887–910, 1997.

[14] R. L. Joshi and T. R. Fisher, "Comparison of generalized Gaussian and Laplacian modeling in DCT image coding," *IEEE Signal Processing Letters*, vol. 2, pp. 81–82, 1995.

[15] Mohamed-Jalal Fadili and Larbi Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Trans. on Image Processing*, vol. 2, pp. 231–240, 2005.

[16] S. Bossi, F. Mapelli and R. Lancini, "Semi-fragile watermarking for video quality evaluation in broadcast scenario ," in *Proc. of ICIP*, 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Estimation of Direction of Arrival (DOA) Using Real-Time Array Signal Processing

Md. Shahedul Amin, Ahmed-Ur-Rahman, Saabah-Bin-Mahbub,
Khawza I. Ahmed[1] and Zahidur Rahim Chowdhury[1]

Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT)
Board Bazar, Gazipur-1704, Bangladesh.
[1]Department of Electrical and Electronic Engineering, United International University (UIU)
Sat Masjid Road, Dhaka, Bangladesh.
E-mail: shahedul_amin2000@yahoo.com, romel_iut@yahoo.com, saabah.mahbub@gmail.com, khawza@eee.uiu.ac.bd
and zahidur@eee.uiu.ac.bd

Abstract- Array Signal Processing (ASP) is a relatively new technique in Digital Signal Processing (DSP) with many potential applications in communication and speech processing. Direction of arrival (DOA) can be estimated using different techniques evolved with ASP. Spectral-based algorithm and subspace-based methods are implemented using two widely used softwares, MATLAB™ and National Instrument's LabVIEW™, to demonstrate the feasibility of introducing the topics in course curriculum of graduate or undergraduate program. It is observed that subspace method provides superior performance in resolving closely spaced sources. The blocks developed using LabVIEW™ can be used for processing signals obtained from data acquisition card in real time.

I. Introduction

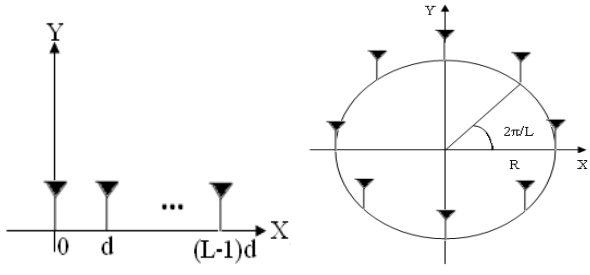
Digital Signal Processing is one of the fastest growing sectors of Electrical Engineering and many applications, available in our day to day life, have been developed using this technology. Array signal processing (ASP) is one of the techniques of DSP which has many potential applications [1]. Sensor ASP has emerged as an active area of research and is centered on the ability to analyze data collected at several sensors [1]. Such topics are not taught in undergraduate level in the most of the universities. This paper addresses method to introduce the topic and ASPs practical implications in undergraduate level. The application developed in this article may also be introduced in the laboratory as experiments in Digital Signal Processing Laboratory. The students can perform the simulation both in MATLAB and National Instruments LabVIEW. National Instrument's LabVIEW offers graphical interface for simulation with a number of advantages like – direct implementation of design from this software in DSP kit, capturing real time data using data acquisition card and easy manipulation of captured or stored data or parameters. The most common applications of array signal processing involve detecting location of acoustic signals [2] which is the focus of this paper. The sensors in this case are microphones and arrangement of microphone positions is significant. We have considered a linear array to collect signals from relatively low frequency sounds (0 to 8 kHz) coming out from a specific direction.

II. Related Terms

This section briefly introduces the relevant terms associated with array signal processing.

A. Array Signal Processing

Array signal processing is a part of signal processing that uses sensors organized in patterns or arrays to detect signals and to determine information about the signals [2]. Arrays can be arranged in a line or a circle as shown in Figure 1. Uniform linear array (ULA) where L numbers of sensors are spaced linearly with equal distance d is shown in Figure 1(a). Figure 1(b) demonstrates the uniform circular array (UCA) where L numbers of sensors are spaced circularly with equal amount of angle $2\pi / L$.



(a) (b)

Fig. 1 Array arrangements (a) Uniform linear array, (b) Uniform circular array

B. Spatial Frequency Transform

Analogous to the Discrete Fourier Transform (DFT) [2], Spatial Frequency Transform is the sampled and windowed spatial equivalent that is used to filter signal in space. The information in the space domain or wave number domain is directly related to the angle the signal is coming from relative to the ULA.

C. Spatial Aliasing

It is well known fact from the sampling theorem that aliasing occurs in the frequency domain if the signal is not sampled at high enough rate (the minimum rate is Nyquest

sampling rate given by the twice of the bandwidth of the signal). We have the same sort of considerations to take into account when we analyze the spectrum of the spatial frequency as well. The Nyquest equivalent of the sampling rate to avoid spatial aliasing implies that the distance between the sensors d should be less than or equal to the half of the minimum wavelength [3], i.e., $d \leq \lambda_{\min}/2$ where $\lambda_{\min}$ is the minimum wavelength corresponding to the maximum frequency $f_{\max}$. This is due to the fact that the velocity of sound, $v = f\lambda$ is fixed in a medium and thus, when the frequency is maximum, the wavelength is minimum.

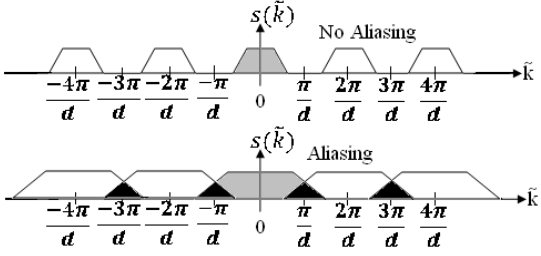


Fig. 2 Visualization of Spatial Aliasing

In Figure 2, $\tilde{k}$ is the space domain or wavenumber whereas $s(\tilde{k})$ is the spectrum of the space domain sampled signal. In the top of Figure 2 Nyquest Sampling rate is maintained and as a result there is no overlap of the spectra of the sampled signals but in the bottom of Figure 2 aliasing occurs as Nyquest criterion is not maintained.

D. Beamforming

Beamforming is the process of combining sounds or electromagnetic signals that come from only one particular direction and impinges different sensors at the receiver. Due to the coherent combining after the appropriate phase compensation at each sensor the resultant signal provides higher strength. Thus, the resultant gain of the sensor would look like a large dumbbell shaped lobe aimed in the direction of interest [3]. This important concept is used in different communication, voice and sonar applications [3].

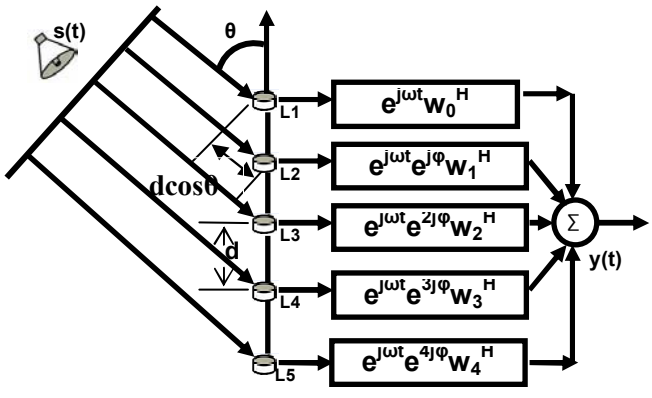
E. Direction of Arrival (DOA)

It is a process of finding the exact location of the source from where the sound is coming. There are three ways of finding the Direction of Arrival (DOA) [3]:

- Spectral-based algorithm
  1. Conventional beamformer
  2. Capon's beamformer
- Subspace-based methods
  1. Multiple signal classification (MUSIC) algorithm
  2. Extension to MUSIC algorithm
- Parametric Methods
  1. Deterministic Maximum Likelihood Method
  2. Stochastic Maximum Likelihood Method

III. System Model for Estimating DOA

It is assumed that the array is linear with L sensors and sound source is located at $\theta°$ away from the axis of the array as shown in Figure 3. A useful property of the ULA is the delay from one sensor to the next is uniform across the array because of their equidistant spacing. Planar sinusoidal sound waves are considered to avoid complexity. Trigonometry reveals that the additional distance the incident signal travels between sensors is $d\cos\theta$. Thus, the time delay between consecutive sensors is given by, $\tau = d\cos\theta / c$.



Velocity of sound, v = 330.7 m/s, f_{max} = 1600 Hz, L=5

Distance between two sensors, d = 9.9 cm, $\Phi = -\omega\tau$

Fig. 3 Application of uniform Linear Array

Let's say, the highest narrowband frequency we are interested is $f_{\max}$. To avoid spatial aliasing, we would like to limit phase differences between spatially sampled signals to π or less because phase differences above π causes incorrect time delays to be seen between received signals and we have, $2\pi f_{\max}\tau \leq \pi$ [4]. Substituting for τ, we get, $d \leq c/2f_{\max}\cos\theta$. Since, worst delay occurs for $\theta = 0°$, we obtain the fundamentally important condition to avoid spatial aliasing, $d \leq \lambda_{\min}/2$. Referring to Figure 3 for an L-element ULA, the array output vector is obtained as [4]

$$\mathbf{x}(t) = \mathbf{a}(\theta)s(t), \tag{1}$$

where $\mathbf{x}(t)$ is the array output, $s(t) = \exp(j\omega t)$ is the signal coming from the source and steering vector, $\mathbf{a}(\theta) = [1, \exp(j\phi), ..., \exp(j(L-1)\phi)]^T$ assuming the propagation delay between the source and the first sensor is normalized to unity and the phase delay between the sensors, $\phi = -\omega d\cos\theta / c$. A single signal at the DOA θ, thus results in a scalar multiple of the steering vector. If M signals impinge on an L-dimensional array from distinct DOAs $\theta_1, \theta_2, ..., \theta_M$, the output vector takes the form

$$\mathbf{x}(t) = \sum_{m=1}^{M} \mathbf{a}(\theta_m) s_m(t) \tag{2}$$

where $s_m(t)$ denotes the baseband signal waveforms from m-th source. The output equation can be put in a more compact form by defining a steering matrix and a vector of signal waveforms as [4]

$$\mathbf{A}(\theta) = [\mathbf{a}(\theta_1),...,\mathbf{a}(\theta_M)] \tag{3}$$

$$s(t) = [s_1(t),...,s_M(t)]^T \tag{4}$$

In the presence of an additive noise $\mathbf{n}(t)$, we now get the model commonly used in array processing

$$\mathbf{x}(t) = \mathbf{A}(\theta)s(t) + \mathbf{n}(t). \tag{5}$$

The methods to be presented all require $M < L$, which is therefore assumed throughout the paper.

To find DOA the idea is to "steer" the array in one direction at a time and measure the output power. The steering locations which result in maximum power yield the DOA estimates. The array response is steered by forming a linear combination of the sensor outputs [4]

$$y(t) = \sum_{l=1}^{L} w_l^* x_l(t) = \mathbf{w}^H \mathbf{x}(t) \tag{6}$$

where $\mathbf{w}$ is the weighting vector used for cancelling the phase delay between the sensors and $\mathbf{w}^H$ is the Hermitian of $\mathbf{w}$. N-samples of $y(t)$ are taken with time interval T between the samples and $t = kT$, where $k = 1,2,...,N$. The output power is measured by

$$P(\mathbf{w}) = \frac{1}{N}\sum_{t=1}^{N} |y(t)|^2 = \frac{1}{N}\sum_{t=1}^{N} \mathbf{w}^H \mathbf{x}(t)\mathbf{x}^H(t)\mathbf{w}$$

$$= \mathbf{w}^H \mathbf{R} \mathbf{w}, \tag{7}$$

where $\mathbf{R} := \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}^H(t)$. The steps for finding the DOA are shown graphically in Figure 4, which we have used in all the proposed experiments [5].

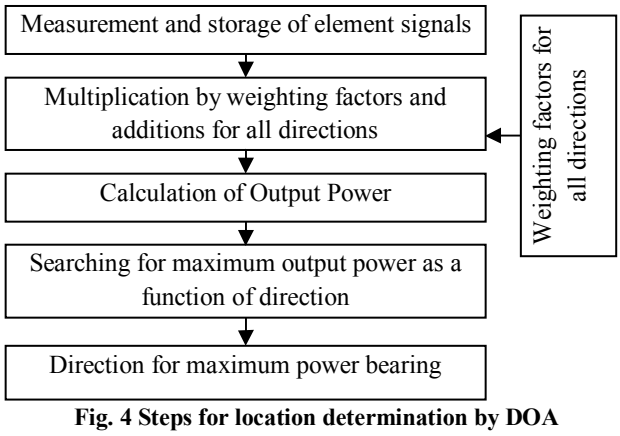


Fig. 4 Steps for location determination by DOA

IV. Experiments

In this section we would like to introduce the proposed experiments based on DOA. Here we have introduced the spectral-based methods and subspace-based methods of estimating DOA.

A. EXP. 1: Conventional Beamformer

Theory: The conventional beamformer is a natural extension of classical Fourier-based spectral analysis to sensor array data [5]. For an array of arbitrary geometry, this algorithm maximizes the power of the beamforming output for a given input signal. Let, we wish to maximize the output power from a certain direction θ. The problem of maximizing the output power is then formulated as [6],

$$\max E\{\mathbf{w}^H \mathbf{x}(t)\mathbf{x}^H(t)\mathbf{w}\} = \max\{\mathbf{w}^H E[\mathbf{x}(t)\mathbf{x}^H(t)]\mathbf{w}\}$$

$$= \max\{E|s(t)|^2 |\mathbf{w}^H \mathbf{a}(\theta)|^2 + \sigma^2 |\mathbf{w}|^2\} \tag{8}$$

where σ^2 is the noise covariance and the assumption of spatial white noise is used [6]. To obtain a non-trivial solution, the norm of $\mathbf{w}$ is constrained to $|\mathbf{w}| = 1$ when carrying out the above maximization. The resulting solution for $\mathbf{w}$ is then,

$$\mathbf{w}_{BF} = \frac{\mathbf{a}(\theta)}{\sqrt{\mathbf{a}^H(\theta)\mathbf{a}(\theta)}} \tag{9}$$

Inserting the weighting vector from Equation 9 into Equation 7, the classical *spatial spectrum* is obtained [6],

$$P_{BF} = \frac{\mathbf{a}^H(\theta)\mathbf{R}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{a}(\theta)} \tag{10}$$

MATLAB simulation: We have assumed six sensors and two sound sources located at $45°$ and $135°$ from the axis of the array. The array consists of $L = 10$ sensors arranged in the form of ULA.

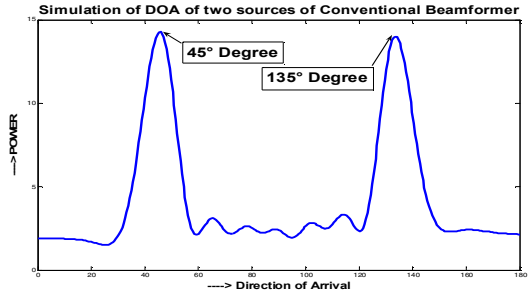


Fig. 5 DOA of two sources of Conventional Beamformer using MATLAB

Figure 5 shows the MATLAB simulation of power measurement in all the positions of space and maximum power is obtained at an angle of $45°$ and $135°$ from the axis of the array where the two sources were located. The power is measured in watts throughout the experiments.

LabVIEW Implementation:

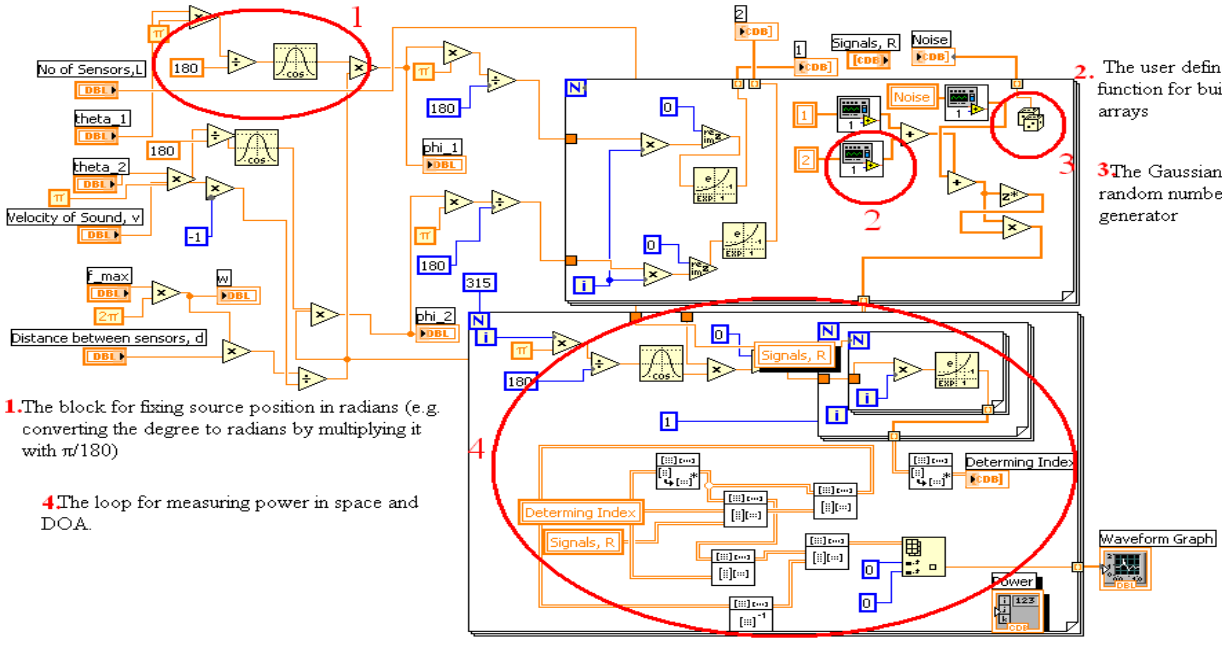


Fig. 6 Block Diagram of Conventional Beamformer in LabVIEW

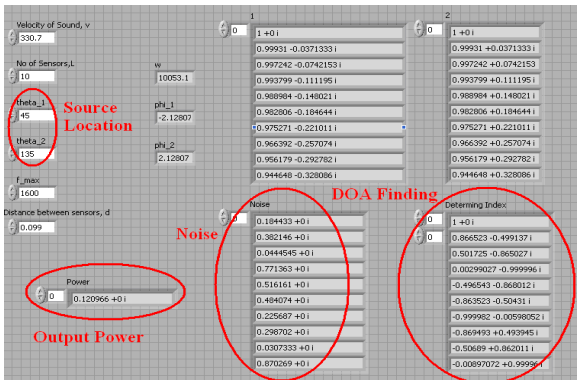


Fig. 7 Portion of Front Panel of Conventional Beamformer in LabVIEW

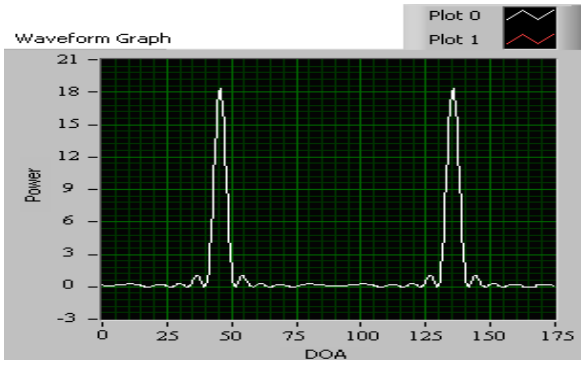


Fig. 8 Result from LabVIEW of Conventional Beamformer

LabVIEW implementation of Conventional Beamformer is shown in Figure 6 which describes graphical representation of the basic mathematical equations where some of the functional blocks are shown. Figure 7 shows some portion of the front panel of LabVIEW where important blocks are shown in the figure. The output for Conventional Beamformer in LabVIEW is shown in Figure 8 and the result is same that obtained from MATLAB as shown in Figure 5.

Limitation of Conventional Beamformer: The standard beamwidth for a ULA is $\varphi_B = 2\pi / L$, and sources whose electrical angles are closer than φ_B will not be resolved by the Conventional Beamformer, regardless of the available data quality [6].

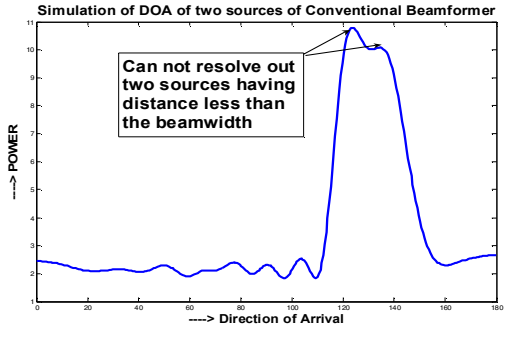


Fig. 9 DOA of two sources of having distance less than the beamwidth in Conventional Beamformer using MATLAB

A ULA of $L = 10$ sensors of half wavelength inter-element spacing has a beamwidth of $2\pi/10 = 0.63$ radians, implying that sources need to be at least $12°$ apart in order to be separated by the beamformer. We have assumed the two sources located at $124.2°$ and 135° which is less than $12°$ and thus spatial aliasing takes place. Therefore, the sensors can not resolve out two sources as desired which is shown in Figure 9.

B. EXP. 2: Capon's Beamformer

Theory: In an attempt to alleviate the limitation of the conventional beamformer, such as its resolving power of two sources spaced closer than the beamwidth, proposed modifications is given by Capon which is also known as the Minimum Variance Distorsionless Response Filter [7]. The optimization problem is proposed as,

$$Min\ P(\mathbf{w}) \text{ subject to } \mathbf{w}^H\mathbf{a}(\theta) = 1 \qquad (11)$$

where $P(\mathbf{w})$ is as defined in Equation 7. This beamformer attempts to minimize the power contributed by noise and any signals coming from other directions than θ, while maintaining a fixed gain in the "look direction θ" like as a sharp spatial bandpass filter. The optimal $\mathbf{w}$ can be found using the technique of Lagrange multipliers, resulting in [8]

$$\mathbf{w}_{CAP} = \frac{\mathbf{R}^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{R}^{-1}\mathbf{a}(\theta)} \qquad (12)$$

Inserting the above weight into Equation 7 leads to the following "*spatial spectrum*" [8]

$$P_{CAP}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{R}^{-1}\mathbf{a}(\theta)} \qquad (13)$$

MATLAB simulation: Here we consider two sources at 135° and 124.2° away from the axis of the array and we see that it resolves out them perfectly and its MATLAB Simulation is shown in Figure 10.

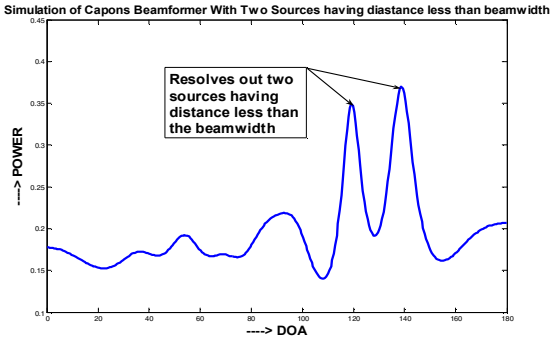


Fig. 10 Simulation of DOA for Capon's Beamformer with two sources having distance less than beamwidth

LabVIEW Implementation: Block diagram for Capon's beamformer is pretty similar to that of Conventional Beamformer except the power calculation loop as mentioned in Figure 6. It is shown in Figure 11. The results from LabVIEW of Capon matches with the results obtained from MATLAB simulation shown in Figure 10.

C. EXP. 3: MUSIC Algorithm

Theory: The main features of MUSIC are [8]:

- Its properties are directly related with the Eigen-structure of the covariance matrix.
- Unlike others, MUSIC was originally presented as a DOA estimator.

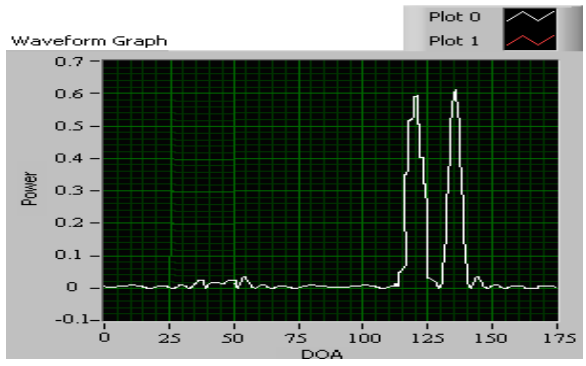


Fig. 11 Result from LabVIEW of Capon's Beamformer

- It is a frequency estimation technique.
- It reduces noise to a great extent.

The signal parameters which are of interest are spatial in nature and thus require the cross covariance information among the various sensors, i.e., the spatial covariance matrix given by [8]

$$\mathbf{R} = E\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \mathbf{A}E\{s(t)s^H(t)\}\mathbf{A}^H + E\{\mathbf{n}(t)\mathbf{n}^H(t)\} \qquad (14)$$

with

$$E\{s(t)s^H(t)\} = \mathbf{P} \qquad (15)$$

is the source covariance matrix and

$$E\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma^2\mathbf{I} \qquad (16)$$

is the noise covariance matrix. To allow for unique DOA estimates, the array is usually assumed to be unambiguous; that is, any collection of M steering vectors corresponding to distinct DOAs $\theta(k), k = 1,\ldots,M$ forms a linearly independent set of $\{\mathbf{a}(\theta(1)),\ldots,\mathbf{a}(\theta(M))\}$ and $\mathbf{P}$ has full rank [9]. In practice, an estimate $\mathbf{R}$ of the covariance matrix is obtained and its eigenvectors are separated into the signal and noise subspace as

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I} = \mathbf{U}\Lambda\mathbf{U}^H \qquad (17)$$

with $\mathbf{U}$ unitary and $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_L\}$, a diagonal matrix of real eigenvalues ordered such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L > 0$. It is observed that any vector orthogonal to $\mathbf{A}$ is an eigenvector of $\mathbf{R}$ with the eigenvalue σ^2 [9]. There are $L - M$ linearly independent such vectors. Since, the remaining eigenvalues are all larger than σ^2, we can partition the eigenvector pairs into noise eigenvector (corresponding

to eigenvalues $\lambda_{M+1} = \ldots = \geq \lambda_L = \sigma^2$) and signal eigenvectors (corresponding to eigenvalues $\lambda_1 \geq \ldots \geq \lambda_M > \sigma^2$). Hence we can write [10]

$$\mathbf{R} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^H + \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{U}_n^H, \qquad (18)$$

where $\mathbf{\Lambda}_n = \sigma^2 \mathbf{I}$. Since all noise eigenvectors are orthogonal to $\mathbf{A}$, the columns of $\mathbf{U}_s$ must span the range space of $\mathbf{A}$ whereas those of $\mathbf{U}_n$ span its orthogonal complement. The projection operators onto this signal and noise subspaces are defined as [10]

$$\mathbf{\Pi} = \mathbf{U}_s \mathbf{U}_s^H = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \qquad (19)$$

$$\mathbf{\Pi}^{\perp} = \mathbf{U}_n \mathbf{U}_n^H = \mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \qquad (20)$$

Thus, MUSIC "Spatial Spectrum" is defined as

$$P_M(\theta) = \frac{\mathbf{a}^H(\theta)\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{\Pi}^{\perp}\mathbf{a}(\theta)} \qquad (21)$$

MATLAB simulation: This experiment assumes two sources at $124.2°$ and $135°$ which was taken in the previous experiments. The output obtained from MATLAB Simulation of MUSIC is shown in Figure 12. From the figure it is apparent that through MUSIC sensors are able to detect sources having distance less than beamwidth.

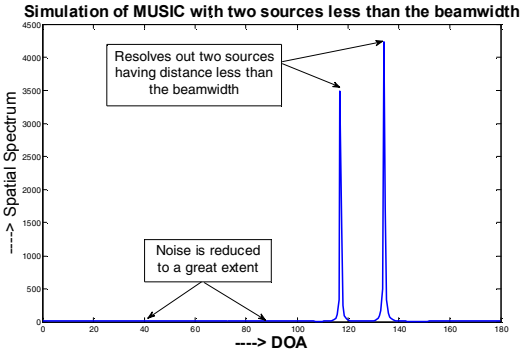


Fig. 12 Simulation of MUSIC

We get much more improved result here since the noise is reduced to a very good extent and also it works as a very sharp *spatial bandpass filter* [7].

LabVIEW Implementation: The block diagram is also similar with the Figure 6 except the power calculation loop due to the difference in power calculation method. Figure 13 shows the output from LabVIEW of MUSIC algorithm and no difference exists between Figure 12 and 13.

Results: From the above Three experiments it is apparent that MUSIC is the best algorithm for finding the DOA.

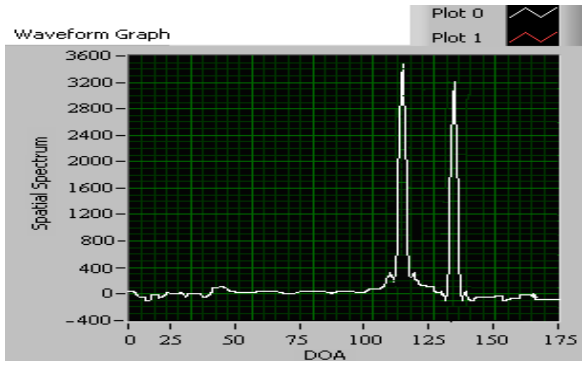


Fig. 13 Result of MUSIC in LabVIEW

V. Conclusion

In this paper, the theory of Array Signal Processing is introduced for voice signals to find the location of source. It can be inferred that estimates of an arbitrary location of signal source can be performed with moderate accuracy if the data collection time is sufficiently long or the SNR is adequately high, and the signal model is sufficiently accurate. A hardware implementations using LabVIEW data acquisition card that would use the blocks developed in LabVIEW in real time is in progress. Since the proposed set of experiments augments the practical implication of ASP in addition to the theoretical understanding, it will be beneficial for the undergraduate students if it is introduced in the Digital Signal Processing Laboratory experiments.

Reference

[1] Monson H. Hayes, *Statistical Digital Signal Processing and Modeling,* John Wiley and Sons, INC., 1996.

[2] Henry Stark, John W. Woods, *Probability and Random Processes with Application to Signal Processing,* Pearson Education, 2002.

[3] Hamid Krim and Mats Viberg, "Two Decades of Array Signal Processing", IEEE Signal Processing Magazine, pp. 67-90, July, 1996.

[4] Claiborne McPheeters, James Finnigan, Jeremy Bass and Edward Rodriguez,"Array Signal Processing: An Introduction" Version 1.6: Sep 12, 2005.

[5] A.Leshem and A.J.van der Veen, *"Direction-of-Arrival estimation for constant modulus signals,"* IEEE Trans. Signal Processing; vol. 47, pp.3125-3129, Nov 1999.

[6] Alan V. Oppenheim, Alan S. Willsky and S. Hamid Nawab, *Signals and Systems,* Pearson Education, 2004.

[7] Robert F. Coughlin and Frederick F. Driscoll, *Operational Amplifiers and Linear Integrated Circuits,* Prentice-Hall of India Private Limited, 2006.

[8] John G. Proakis and Dimitris G. Manolakis, *Digital Signal Processing Principles, Algorithms and Applications,* Prentice-Hall of India Private Limited, 2006.

[9] Frank Ayres, JR, *Theory and Problems of Matrices,* McGraw-Hill Book Company, New York, 1974.

[10] J.J.Shynk and R.P.Gooch, "The constant modulus array for co-channel signal copy and direction finding, "IEEE Trans. Signal Processing; vol. 44, pp.652-660, Mar.1996.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Least-Squares Optimal Variable Step-Size LMS for Nonblind System Identification with Noise

M. A. Wahab [1], M. Adel Uzzaman [1], M. S. Hai [1], M. A. Haque [1], and M. K. Hasan [1,2]
[1] Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.
[2] East West University, Dhaka, Bangladesh.
E-mail: {arifulhoque, khasan}@eee.buet.ac.bd

Abstract— This paper proposes a least-square optimal variable-step-size (LSVSS) least-mean-square (LMS) adaptive algorithm for nonblind identification of single-input single-output (SISO) finite impulse response systems. It is shown that the well-known normalized LMS (NLMS) and the LSVSS-LMS algorithms are mathematically equivalent for the noise-free case. The derivation of LSVSS is then extended for noisy measurements. The convergence analysis of the LSVSS-LMS is also presented. The performance of the proposed method is compared with conventional robust variable-step-size LMS algorithms. Experimental results demonstrate improved performance of the proposed algorithm for nonblind system identification in both stationary and nonstationary environments.

I. Introduction

The computational simplicity and operational stability of the LMS algorithm have made it attractive in various adaptive signal processing applications such as echo cancelation, dereverberation, video conferencing, mobile and hands-free telephony [1]. The convergence speed and stability of this algorithm, however, are critically dependent on the adaptation step-size. A small value of the step-size increases the convergence time while a large value increases the excess mean-square error (MSE). A too large value might cause instability of the algorithm. These contradictory requirements can only be resolved by using a variable step-size LMS algorithm [2], [3]. The main idea of these algorithms is to use a function of process parameters to detect the distance of the adaptive filter weights from the optimal ones and use a small step-size when it is small and a large step-size otherwise. The performance of the LMS algorithm can be significantly improved by using such a logic in selecting the adaptation step-size.

In recent years, several adaptive step-size LMS algorithms have been reported in the literature by many researchers. In [4], the step-size is adjusted by sign changes of successive samples of the gradient. A technique for achieving high initial speed of convergence and small steady-state misadjustment using variable step-size has been introduced in [5]. It is very similar to the unsigned-VSSLMS (UVSSLMS) algorithm [6]. In [7], the step-size is adjusted using the squared instantaneous error. A fast LMS (FLMS) adaptive filter with three step-sizes derived from three output errors has been reported in [8]. Instead of using a common step-size parameter for all the filter taps, a multiple step-size algorithm has been described in [9]. The step-sizes are adjusted based on the consecutive sign changes of the adaptive filter update term. These approaches perform better than the conventional fixed step-size LMS algorithm in a low noise environment. Since, practical environments are most often noisy, the usefulness of any adaptive algorithm is evaluated by its performance in presence of noise. In a noisy environment, the error includes the noise and thus it makes the step-size to be a function of noise while using the algorithms in [7]-[9]. As a consequence, near convergence of the adaptive filter, the adaptive step-size will still be large due to the presence of noise term. This results in large misadjustment due to the large fluctuations around the optimum. A different approach to adapt the step-size is used in [10]. Here, the step-size of the adaptive filter is changed according to a gradient descent algorithm to reduce the squared estimation error during each instant. In [6], the emphasis is given on tracking performance of the signed-VSSLMS (SVSLMS) and unsigned-VSLMS (UVSLMS) algorithms. In the modified variable step-size LMS (MVSS-LMS) algorithm [3], the step-size is adjusted using the time averaged estimate of the autocorrelation of $e(n)$ and $e(n-1)$. The motivation behind using the autocorrelation of $e(n)$ and $e(n-1)$ instead of $e^2(n)$ as in [7] is to minimize the measurement noise effect on the calculation of the adaptive step-size. This method works fine as long as the autocorrelation of noise terms, i.e., $v(n)$ and $v(n-1)$ can be neglected. In practical situations, measurement noise is not truly uncorrelated. Due to this fact, at low SNR the noise effect in the calculation of adaptive step-size cannot be ignored. A new class of gradient adaptive step-size LMS denoted as c-VSSLMS has been reported in [2]. Instead of using the autocorrelation between the error as stated above, this algorithm exploits the autocorrelation between the gradients $\nabla J(n)$ and $\nabla J(n-1)$ for improving robustness to noise. The convergence and stability of such algorithms are, however, highly dependent on the tunable parameters in the algorithm as we show later in the results section.

In this paper, we propose a novel technique for deriving an expression for the optimal variable step-size of the LMS algorithm. The variable step-size that is derived here is optimal in the least-squares (LS) sense which minimizes the distance between the true and next estimate of the coefficient vector given the present estimate. The proposed method results in an exact variable step-size even in presence of measurement noise. We call it exact because it leads to a convergence similar to that of the noise-free case.

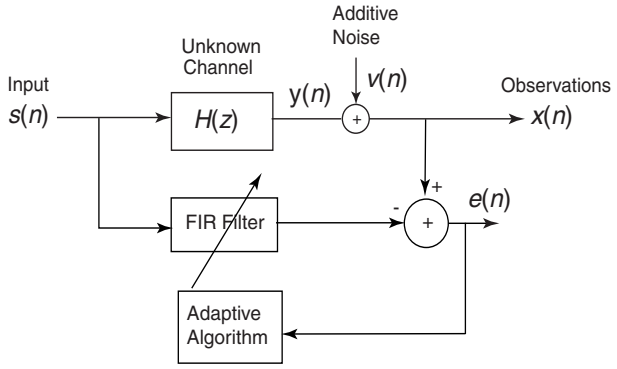


Fig. 1. Block diagram of a nonblind SISO system identification setup.

The practical implementation of the exact equation, however, is not possible. Nevertheless, it gives us a structure of the adaptation step-size to be used in the noisy case. An approximate realization of the LS variable step-size (LSVSS) is also proposed to make the algorithm useful in practical applications. The convergence analysis of the LSVSS-LMS is presented for completeness of the algorithm. Finally, the performance of the algorithm is evaluated and compared with other recently reported techniques [2], [3] both in stationary and nonstationary environments.

II. Problem Formulation

The input-output relationship and noisy observations of a SISO system in Fig. 1 are given by the following equations:

$$y(n) = s(n) * h(n) = \sum_{k=0}^{L-1} h(k)s(n-k) \quad (1)$$

$$x(n) = y(n) + v(n) \quad (2)$$

where $s(n)$, $y(n)$, $x(n)$, $v(n)$ and $h(n)$ denote, respectively, the input, output, the noisy output, observation noise, and impulse response of the system. It is also assumed that $v(n)$ is uncorrelated with $s(n)$. Using matrix notation, (1) can be written as

$$y(n) = \mathbf{h}^T(n)\mathbf{s}(n) \quad (3)$$

where $\mathbf{h}(n) = [h(n)\ h(n-1)\ \cdots\ h(n-L+1)]^T$ and $\mathbf{s}(n) = [s(n)\ s(n-1)\ \cdots\ s(n-L+1)]^T$.

A nonblind system identification algorithm estimates $\mathbf{h}(n)$ from the input-output observations $s(n)$ and $x(n)$, $n = 0, 1, \cdots, N-1$, respectively.

III. The LMS-type Algoritms: A Brief Review

The LMS-type algorithms minimize the following cost function:

$$J(n) = e^2(n) \quad (4)$$

where the error $e(n)$ is defined as

$$e(n) = x(n) - \widehat{\mathbf{h}}^T(n)\mathbf{s}(n) = y(n) - \widehat{\mathbf{h}}^T(n)\mathbf{s}(n) + v(n). \quad (5)$$

Now, the gradient of the cost function can be obtained as

$$\nabla J(n) = -2e(n)\mathbf{s}(n) \quad (6)$$

A. LMS Algoritm

The update equation for LMS algorithm [11] is given by

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + 2\mu e(n)\mathbf{s}(n) \quad (7)$$

where μ is the fixed positive scalar adaptation step-size which controls the speed of convergence, final misalignment, and stability of the algorithm. We call this basic algorithm the fixed step-size LMS (FSS-LMS).

B. NLMS Algoritm

The NLMS update equation [11] is expressed as

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + 2\frac{\mu}{||\mathbf{s}(n)||^2}e(n)\mathbf{s}(n) \quad (8)$$

where $||\cdot||$ denotes the l_2 norm. The convergence of the algorithm is guaranteed for $0 < \mu < 1$ and the speed of convergence is fastest at $\mu = \frac{1}{2}$ for white noise input [12]. This algorithm has some distinct advantages over the LMS algorithm. It provides potentially-faster convergence speeds than the LMS algorithm for both correlated and whitened input data [11], [13], [14] and shows stable behavior for a known range of parameter values independent of the input data frames. The disadvantage of the NLMS algorithm, however, is the requirement of a minimum of one additional multiplication, division and addition over the LMS algorithm.

IV. The LSVSS-LMS Algorithm

In this section, we derive a novel variable step-size LMS algorithm for adaptive applications. Unlike the conventional heuristic approaches, an expression for the adaptive step-size $\mu(n)$ is derived in the least-squares (LS) sense following our previous work on SIMO blind channel identification [15].

We define the cost function as

$$J_\mu(n+1) = [\mathbf{h} - \widehat{\mathbf{h}}(n+1)]^T[\mathbf{h} - \widehat{\mathbf{h}}(n+1)] \quad (9)$$

The cost function is formulated such that $\widehat{\mathbf{h}}(n+1)$ approximates the true $\mathbf{h}$ in the LS sense for the given $\widehat{\mathbf{h}}(n)$. The LMS update equation for adaptive step-size $\mu(n)$ can be written as

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) - \mu(n)\nabla J(n) \quad (10)$$

Substituting (10) into (9), we obtain

$$J_\mu(n+1) = [\mathbf{h} - \widehat{\mathbf{h}}(n) + \mu(n)\nabla J(n)]^T [\mathbf{h} - \widehat{\mathbf{h}}(n) + \mu(n)\nabla J(n)] \quad (11)$$

Rearranging and simplifying, we get

$$J_\mu(n+1) = \mathbf{h}^T\mathbf{h} - 2\mathbf{h}^T\widehat{\mathbf{h}}(n) + 2\mu(n)\mathbf{h}^T\nabla J(n) + \widehat{\mathbf{h}}(n)^T\widehat{\mathbf{h}}(n) - 2\mu(n)\widehat{\mathbf{h}}(n)^T\nabla J(n) + \mu^2(n)||\nabla J(n)||^2 \quad (12)$$

To obtain the optimal step-size $\mu(n)$, the partial derivative of $J_\mu(n+1)$ is equated to zero. Setting, $\partial J_\mu(n+1)/\partial\mu(n) = 0$ gives

$$\mu_{opt}(n) = \frac{\widehat{\mathbf{h}}^T(n) - \mathbf{h}^T}{||\nabla J(n)||^2}\nabla J(n) \tag{13}$$

Substituting (6) into (13) and simplifying, the LSVSS step-size can be obtained as

$$\mu_{opt}(n) = \frac{1}{2}\frac{e^2(n) - e(n)v(n)}{e^2(n)||\mathbf{s}(n)||^2} \tag{14}$$

For the noise-free case, i.e. when $v(n) = 0$, (14) reduces to

$$\mu_{opt}(n) = \frac{1}{2}\frac{1}{||\mathbf{s}(n)||^2} \tag{15}$$

In a real environment, the effect of noise cannot be avoided and thus practical implementation of (14) is inevitable. If convergence of the LMS algorithm is assumed, then the estimation error $e(n)$ converges to the measurement noise $v(n)$. In the initial period of adaptation, the effect of noise on the update may be neglected as the signal gradient will dominate. We, therefore, propose an estimator for the measurement noise $v(n)$ as

$$\widehat{v}(n) = g(n)e(n) \tag{16}$$

where $g(n)$ is defined as

$$g(n) = 1 - \exp\left(-\frac{\alpha\sigma_v^2}{\sigma_s^2(\rho\bar{e}^2(n) - \sigma_v^2)}\right), \quad \rho \geq 1 \tag{17}$$

where σ_s^2 and σ_v^2 denote, respectively, the input and noise signal power. The values of α and ρ are adjusted to make a compromise between the desired level of steady state misadjustment and the required tracking capabilities of the algorithm. The average value of the error square, $\bar{e}^2(n)$, is calculated as

$$\bar{e}^2(n) = \beta\bar{e}^2(n-1) + (1-\beta)e^2(n), \quad 0 \leq \beta \leq 1 \tag{18}$$

and σ_s^2 can be replaced by $||\mathbf{s}(n)||^2/L$ for simplicity. Now, the steady-state MSE is defined as [16]

$$E\{e^2(\infty)\} = \bar{J}_{min} + \bar{J}_{ex}(\infty) \tag{19}$$

where $\bar{J}_{min} = E\{v^2(n)\}$ is the minimum value of the MSE and $\bar{J}_{ex}(\infty)$ denotes the excess MSE. Substituting, $E\{v^2(n)\} = \sigma_v^2$, where σ_v^2 denotes the noise power, we can rewrite (19) as

$$E\{e^2(\infty)\} = \sigma_v^2 + \bar{J}_{ex}(\infty) \tag{20}$$

Note that from (17) $g(n) \simeq 0$ during the initial period of adaptation when the error is very large and $g(n) = 0$ for $\sigma_v^2 = 0$, i.e. for noise-free condition. On the other hand, for $\rho = 1$, when $\bar{e}^2(n) \to \sigma_v^2$ then $g(n) \to 1$ thus satisfying the desirable conditions. A practical implementation of the LSVSS in (14) using (16) results in the following suboptimal variable step-size:

$$\tilde{\mu}_{opt}(n) = \frac{1}{2}\frac{e^2(n) - e(n)\widehat{v}(n)}{e^2(n)||\mathbf{s}(n)||^2} \tag{21}$$

Clearly, (21) can be implemented in practical systems with the noise power, σ_v^2, estimated using any standard technique.

A. Relationship between LSVSS LMS and NLMS

From (10), (14), (16) and (21), we get

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) + \frac{1-g(n)}{||\mathbf{s}(n)||^2}e(n)\mathbf{s}(n) \tag{22}$$

Now comparing (8) and (22), we see that the two algorithms are equivalent for $\mu = 0.5$ for noise-free case. Under noisy condition, they are equivalent for $\mu = 0.5(1 - g(n))$. Thus we can view the proposed LSVSS-LMS algorithm as the generalized version of the NLMS algorithm in that, unlike the NLMS, it incorporates the noise effect in the adaptation step-size.

V. The Convergence Analysis of LSVSS-LMS

The convergence analysis of the LSVSS-LMS algorithm is similar to that of the NLMS algorithm for the noise-free case but in the noisy case it is necessary to show that $0 < E\{\tilde{\mu}_{opt}(n)\} < 2/\text{tr}(\mathbf{R})$ for the mean weight vector convergence and $0 < E\{\tilde{\mu}_{opt}^2(\infty)\}/E\{\tilde{\mu}_{opt}(\infty)\} \leq 2/(3\text{tr}(\mathbf{R}))$ for the MSE convergence, where $\text{tr}(\mathbf{R})$ is defined as [1]

$$\text{tr}(\mathbf{R}) = \sum_{k=0}^{L-1} E\left\{|s(n-k)|^2\right\} = L\sigma_s^2 \tag{23}$$

where σ_s^2 denotes the input signal power.

We first demonstrate that $E\{\tilde{\mu}_{opt}(n)\}$ satisfies the conditions for mean weight vector convergence. Substituting (16) and (17) into (21), we get

$$\tilde{\mu}_{opt}(n) = \frac{1}{2||\mathbf{s}(n)||^2}\exp\left(-\frac{\alpha\sigma_v^2}{\sigma_s^2(\rho\bar{e}^2(n) - \sigma_v^2)}\right) \tag{24}$$

Taking the statistical expectation of (24), we obtain

$$E\{\tilde{\mu}_{opt}(n)\} = \frac{1}{2}E\left\{\frac{1}{||\mathbf{s}(n)||^2}\exp\left(-\frac{\alpha\sigma_v^2}{\sigma_s^2(\rho\bar{e}^2(n) - \sigma_v^2)}\right)\right\} \tag{25}$$

Considering uniform distribution of $\tilde{\mu}_{opt}(n)$, we can write (25) as

$$\begin{aligned} E\{\tilde{\mu}_{opt}(n)\} = {} & \frac{(\rho\bar{e}^2(n) - \sigma_v^2)(\rho E_{max} - \sigma_v^2)}{2\alpha L\sigma_v^2[(\rho E_{max} - \sigma_v^2) - (\rho\bar{e}^2(n) - \sigma_v^2)]} \\ & \times\left\{\exp\left(-\frac{\alpha\sigma_v^2 L}{\text{tr}(\mathbf{R})(\rho E_{max} - \sigma_v^2)}\right)\right. \\ & \left. - \exp\left(-\frac{\alpha\sigma_v^2 L}{\text{tr}(\mathbf{R})(\rho\bar{e}^2(n) - \sigma_v^2)}\right)\right\} \end{aligned} \tag{26}$$

In the simplified results of (26), we have assumed $E\left\{1/||\mathbf{s}(n)||^2\right\} \cong 1/\text{tr}(\mathbf{R})$. The maximum value of $E\{\tilde{\mu}_{opt}(n)\}$ occurs when $\bar{e}^2(n) \to E_{max}$ and is given by

$$E\{\tilde{\mu}_{opt}(n)\}_{max} = \frac{1}{2\text{tr}(\mathbf{R})}\exp\left(-\frac{\alpha\sigma_v^2 L}{\text{tr}(\mathbf{R})(\rho E_{max} - \sigma_v^2)}\right) \tag{27}$$

Therefore, from (27) it is obvious that $0 < E\{\tilde{\mu}_{opt}(n)\} < 2/\text{tr}(\mathbf{R})$. The convergence of the LSVSS-LMS algorithm in the weight vector mean [1] is thus ensured.

Next, we demonstrate that $E\{\tilde{\mu}_{opt}^2(\infty)\}/E\{\tilde{\mu}_{opt}(\infty)\}$ satisfies the condition for MSE convergence. Using (20)

and (26), the steady-state value of $E\{\tilde{\mu}_{opt}(n)\}$ can be calculated as

$$
\begin{aligned}
&E\{\tilde{\mu}_{opt}(\infty)\} \\
&= \frac{1}{2\alpha L\sigma_v^2}\frac{((\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty))(\rho E_{max}-\sigma_v^2)}{[(\rho E_{max}-\sigma_v^2)-((\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty))]} \\
&\quad\times\left\{\exp\left(-\frac{\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})(\rho E_{max}-\sigma_v^2)}\right)-\right. \\
&\quad\left.\exp\left(-\frac{\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})((\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty))}\right)\right\} \qquad (28)
\end{aligned}
$$

Neglecting $(\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty)$ with respect to $(\rho E_{max}-\sigma_v^2)$, we obtain

$$
\begin{aligned}
E\{\tilde{\mu}_{opt}(\infty)\} &\approx \frac{(\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty)}{2\alpha L\sigma_v^2} \\
&\quad\exp\left(-\frac{\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})(\rho E_{max}-\sigma_v^2)}\right) \qquad (29)
\end{aligned}
$$

Squaring both sides of (24) and following the similar procedure to evaluate the steady-state value of $E\{\tilde{\mu}_{opt}^2(n)\}$, we obtain

$$
\begin{aligned}
&E\{\tilde{\mu}_{opt}^2(\infty)\} \\
&= \frac{1}{8\alpha L\sigma_v^2\mathrm{tr}(\mathbf{R})}\frac{((\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty))(\rho E_{max}-\sigma_v^2)}{[(\rho E_{max}-\sigma_v^2)-((\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty))]} \\
&\quad\times\left\{\exp\left(-\frac{2\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})(\rho E_{max}-\sigma_v^2)}\right)-\right. \\
&\quad\left.\exp\left(-\frac{2\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})((\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty))}\right)\right\} \qquad (30)
\end{aligned}
$$

Simplifying (30) as before, we get

$$
\begin{aligned}
E\{\tilde{\mu}_{opt}^2(\infty)\} &\approx \frac{(\rho-1)\sigma_v^2+\rho\bar{J}_{ex}(\infty)}{8\alpha L\sigma_v^2\mathrm{tr}(\mathbf{R})} \\
&\quad\exp\left(-\frac{2\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})(\rho E_{max}-\sigma_v^2)}\right) \qquad (31)
\end{aligned}
$$

Dividing (31) by (29), yields

$$
\frac{E\{\tilde{\mu}_{opt}^2(\infty)\}}{E\{\tilde{\mu}_{opt}(\infty)\}} = \frac{1}{4\mathrm{tr}(\mathbf{R})}\exp\left(-\frac{\alpha\sigma_v^2 L}{\mathrm{tr}(\mathbf{R})(\rho E_{max}-\sigma_v^2)}\right) \qquad (32)
$$

Therefore, from (32) it is obvious that $0 < \frac{E\{\tilde{\mu}_{opt}^2(\infty)\}}{E\{\tilde{\mu}_{opt}(\infty)\}} < 2/(3\mathrm{tr}(\mathbf{R}))$. The MSE convergence of the LSVSS-LMS algorithm is thus ensured.

VI. Simulation Results

In this section, we present computer simulation results to investigate the effectiveness of the proposed LSVSS-LMS algorithm for system identification with noise. The source signal in all the experiments was considered to be Gaussian white noise unless otherwise stated.

The performance indices used for measurement of improvement is the normalized projection misalignment (NPM) defined as [17]

$$
\begin{aligned}
\mathrm{NPM}(n) &= 20\log_{10}\left(\frac{||\boldsymbol{\Upsilon}(n)||}{||\mathbf{h}||}\right)\ \mathrm{dB}, \\
\boldsymbol{\Upsilon}(n) &= \mathbf{h}-\frac{\mathbf{h}^T\widehat{\mathbf{h}}(n)}{\widehat{\mathbf{h}}^T(m)\widehat{\mathbf{h}}(n)}\widehat{\mathbf{h}}(n) \qquad (33)
\end{aligned}
$$

where $||\cdot||$ is the l_2 norm, and the mean-square-error (MSE).

To demonstrate performance of the proposed LSVSS-LMS algorithm, it is implemented for both stationary and nonstationary environments in a system identification setup. In the Examples 1 and 2 of stationary environment, we implemented the proposed LSVSS-LMS algorithm with parameters $\alpha = 3$, $\beta = 0.993$ and $\rho = 1.5$. The noise power (σ_v^2) is assumed to be known in the tests. The performance of this algorithm is compared with the NLMS, MVSS-LMS [3] and c-VSSLMS [2] algorithms. The parameters' of these algorithms are selected so as to produce a comparable level of final misadjustment. Moreover, our choice of the parameters is also guided by the recommended values in the corresponding references. Experiments with these techniques have shown that their performance is highly dependent on the selection of certain parameters in the algorithms and furthermore, the optimal choice of these parameters is highly data dependent. This fact would severely limit the usefulness of such algorithms in practical applications. In presenting results, we focus on three main concerns: stability, maximum possible convergence speed and the desired level of steady state misadjustment.

A. Example 1: Random Channel, Medium SNR (15dB)

In this experiment, a channel with $L = 64$ random time-invariant coefficients was used. The impulse responses were generated using the 'randn' function of MATLAB. Without loss of generality, the adaptive filter coefficients were initialized by zeros. The MVSS-LMS algorithm was implemented with $\alpha_{MVSSLMS} = 0.97$, $\beta_{MVSSLMS}$ and $\gamma_{MVSSLMS} = 1\times10^{-6}$. For the c-VSSLMS algorithm we used $\alpha_{c-VSSLMS} = 0.88$ and $\rho_{c-VSSLMS} = 1\times10^{-7}$. The SNR for this experiment was set to 15 dB.

The results of this experiment is depicted in Fig. 2. We observe from this figure that the steady-state NPM of the NLMS algorithm is about -15dB as expected. The steady-state NPM of MVSS-LMS is smaller than the NLMS though its speed of NPM convergence is not satisfactory. It can be also seen from Fig. 2 that both the initial speed of convergence and steady-state value of NPM for the c-VSSLMS algorithm are better than in case of the MVSS-LMS. The best performance, however, is obtained from the proposed LSVSS-LMS algorithm both in terms of speed of convergence and steady-state value of NPM.

The mean step-size as shown in Fig. 3 shows that its initial values for both the LSVSS-LMS and NLMS are higher than those of the MVSS-LMS and c-VSSLMS algorithms.

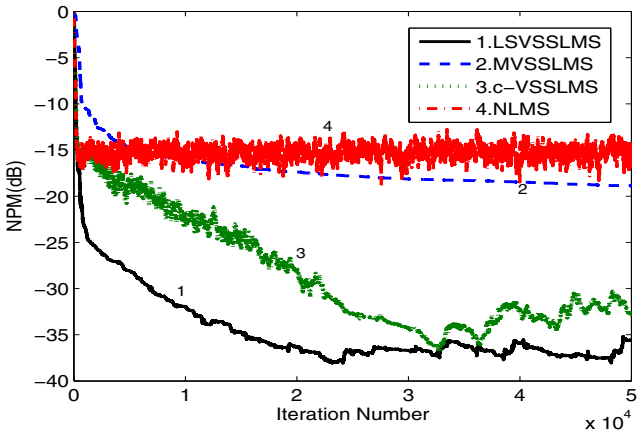


Fig. 2. Comparative performance results using NPM at SNR=15 dB for random channel.

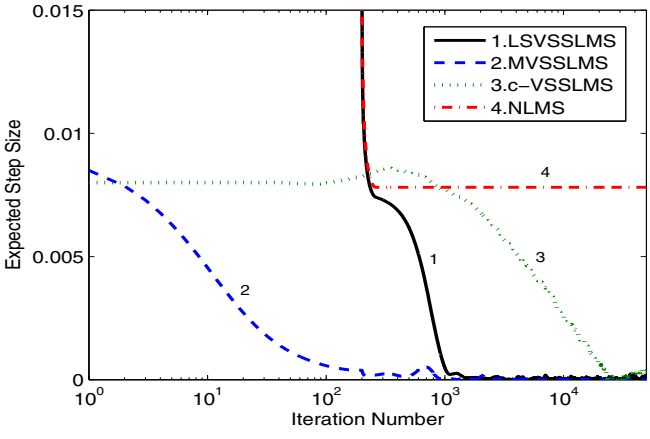


Fig. 3. Comparative performance results using mean step size at SNR=15 dB for random channel.

This is the main reason behind the better speed of convergence of the LSVSS-LMS and NLMS than it is for the other two approaches. Except for the NLMS, the final value of the step-sizes are very similar for the VSS algorithms.

A comparison between the given noise power σ_v^2 and steady-state value of the mean-square-error found from the simulation is shown in Table 1. The theoretical and simulation results of the quantities $E\{\tilde{\mu}_{opt}(n)\}$ and $E\{\tilde{\mu}_{opt}^2(n)\}$ are also given in this table. As shown in Table 1, the steady-state value of the mean-square-error approaches the noise power as expected. Also note that there is a small deviation between the theoretical and the corresponding simulation results. The cause of this discrepancy may be imparted to the assumption of uniform statistical distribution of $\tilde{\mu}_{opt}(n)$ in the theoretical computation.

B. Example 2: Random Channel, Low SNR (5dB)

The random channel we used in this test was the same as in Example 1. The MVSS-LMS algorithm was implemented with $\alpha_{MVSSLMS} = 0.85$, $\beta_{MVSSLMS}$ and $\gamma_{MVSSLMS}$ were as in Example 1. For the c-VSSLMS algorithm, $\alpha_{c-VSSLMS}$ was as in Example 1, and $\rho_{c-VSSLMS}$ was set to 1×10^{-6}. The SNR for this example was 5 dB.

It can be observed from Fig. 4 that the steady-state value

Table 1
Comparison between the analytical and simulation results

σ_v^2	MSE
18.2410	18.4079
$E\{\tilde{\mu}_{opt}(n)\}$ for $n = 40000$	
Theoretical Results	Simulation Results
4.8323×10^{-4}	9.5270×10^{-5}
$E\{\tilde{\mu}_{opt}^2(n)\}$ for $n = 40000$	
Theoretical Results	Simulation Results
4.9378×10^{-7}	2.8069×10^{-7}

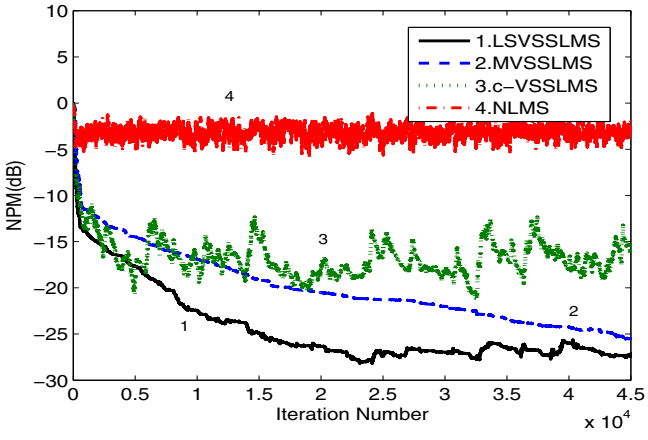


Fig. 4. Comparative performance results using NPM at SNR=5 dB for random channel.

of NPM for the NLMS is around -5 dB whereas the steady-state value of NPM is much smaller than this value for the c-VSSLMS and MVSS-LMS algorithms. In this high noise and random channel environment as well, both the initial speed and steady-state value of NPM are better for the proposed LSVSS-LMS algorithm. If we forcibly set a high value of initial step-size for the MVSS-LMS and c-VSSLMS algorithms to obtain high initial speed of convergence, though not recommended in the two papers, then they become unstable.

C. Example 3: Tracking Performance, Medium SNR (15dB)

In this example, we study the performances of the adaptive algorithms when the unknown system changes abruptly. Here, we have implemented the proposed LSVSS-LMS algorithm with parameters $\alpha = 0.5$, $\beta = 0.993$ and $\rho = 1$ which is found to be a good choice in a nonstationary environment. The MVSS-LMS algorithm was implemented with $\alpha_{MVSSLMS} = 0.93$, $\beta_{MVSSLMS} = 0.995$ and $\gamma = 1 \times 10^{-4}$. These parameters satisfy (29) in [3]. For the c-VSSLMS algorithm, we used $\alpha_{c-VSSLMS} = 0.9991$ and $\rho_{c-VSSLMS} = 3.16 \times 10^{-7}$. The steady-state behavior of the proposed LSVSS-LMS, MVSS-LMS and c-VSSLMS algorithms does not depend on the initial value of the step-size. Even if we set a very

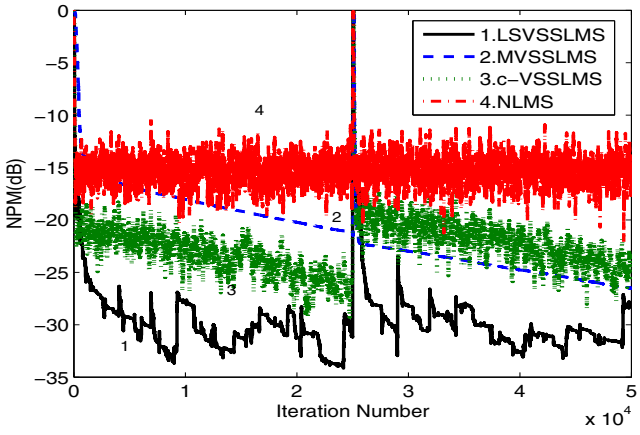


Fig. 5. Comparative tracking performance results using NPM at SNR=15 dB for random channel.

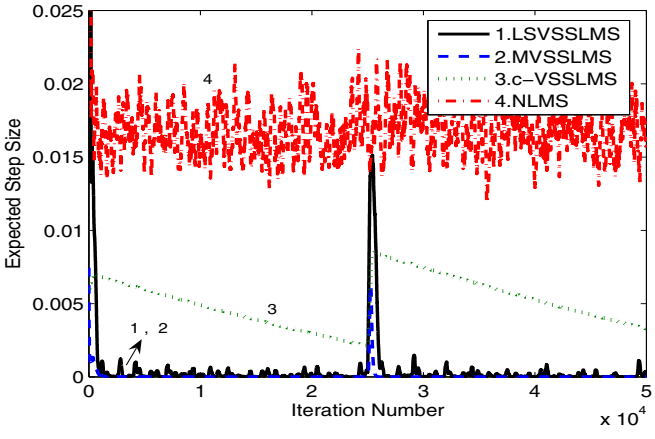


Fig. 6. Comparative tracking performance results using mean step size at SNR=15 dB for random channel.

small value of initial step-size, it increases very quickly to some peak value and then decreases again. If the adaptive filter uses a variable step-size algorithm with the step-size being a function of error, tracking can be performed for abrupt change in coefficients of the unknown system. In this test, we have used a random channel. Here, we obtain the abrupt change in the channel by multiplying all the coefficients of the filter by -1 after 25000 iterations.The NPM and mean step-size are plotted in Figs. 5 and 6, respectively. As can be observed, in case of the proposed algorithm the mean step-size increases very quickly immediately after the environment changes. This implies that the algorithm is able to track abrupt changes in the operating environments. It also proves the sensitivity of the mean step-size of the LSVSS-LMS to the average error energy. The NPM results in Fig. 5 also reveal that the tracking capability of the proposed algorithm is the best among all other algorithms studied in this experiment. Since, the mean step-size of the NLMS algorithm changes only with the input signal energy, its mean step-size does not response to the abrupt change in the channel.

VII. Conclusions

A least-squares variable step-size LMS (LSVSS-LMS) algorithm has been proposed for nonblind identification of finite impulse response systems with noise. The convergence analysis for the proposed LSVSS-LMS has been also presented in this paper. The results of our experimental tests have demonstrated better robustness of the proposed method to additive noise in comparison to other reported methods. The most attractive feature of this algorithm is its stability in both stationary and nonstationary environments. Though we need to tune some of the parameters in (17) to get a fast convergence and small steady-state misadjustment, the proposed algorithm is less sensitive to this parameters' settings for its stability, where a kind of acute parameter tuning played a vital role for the stability of other reported algorithms in the noisy case.

References

[1] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.

[2] W.-P. Ang and Farhang-Boroujeny, "A new class of gradient adaptive step-size lms algorithm," *IEEE Trans. on Signal Processing*, vol. 49, no. 4, pp. 805–810, Apr. 2001.

[3] T. Aboulnasr and K. Mayyas, "A robust variable step-size lms-type algorithm: analysis and simulations," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 631–639, Mar. 1997.

[4] R. H. et al., "A variable step size (VSS) algorithm," *IEEE Trans. Acoust.,Speech, Signal Processing*, vol. ASSP-34, pp. 499–510, June 1986.

[5] J. B. Evans and B. Liu, "Variable step size methods for the LMS adaptive algorithm," in *IEEE Int. Symp Circuits Syst. Proc.*, 1987, pp. 422–425.

[6] B. Farhang-Boroujeny, "Variable-step-size LMS algorithm : new developments and experiments," in *IEE proc. -Vis. Image Signal Processing*, vol. 141, Oct. 1994, pp. 311–317.

[7] R. H. Kwong and E. W. Jhonston, "A variable step-size lms algorithm," *IEEE Trans. Signal Processing*, vol. 40, no. 7, pp. 1633–1642, July 1992.

[8] J. S. Soo and K. K. Pang, "A multi step size (MSS) frequency domain adaptive filter," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp. 115–121, June 1991.

[9] J. B. Evans, P. Xue, and B. Liu, "Analysis and implementation of variable step size adaptive algorithms," *IEEE Trans. Signal Processing*, vol. 41, no. 8, pp. 2517–2535, June 1993.

[10] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing*, vol. 41, no. 6, pp. 2075–2087, June 1993.

[11] D. T. M. Slock, "On the convergence behavior of the lms and the normalized lms algorithms," *IEEE Trans. on Signal Processing*, vol. 41, no. 9, pp. 2811–2825, Sept. 1993.

[12] W. B. M. et al., "Adaptive filters with individual adaptation of parameters," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 677–685, July 1986.

[13] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated gaussian data," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 680–691, July 1988.

[14] S. C. Douglas and T. H.-Y. Meng, "Normalised data nonlinearities for LMS adaptation," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 42, no. 6, pp. 2517–2535, June 1994.

[15] N. D. Gaubitch, M. K. Hasan, and P. A. Naylor, "Generalized optimal step-size for blind multichannel lms system identification," *IEEE Signal Processing Lett.*, vol. 13, no. 10, pp. 624–627, Oct. 2006.

[16] B. W. et al., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," in *Proc. IEEE*, vol. 64, Aug. 1976, pp. 1151–1162.

[17] D. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Lett.*, vol. 5, no. 7, pp. 174–176, July 1998.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Gate C-V Characteristics of Si MOSFETs with Uniaxial Strain Along <110> Direction

Md. Itrat Bin Shams[1], Quazi Deen Mohd Khosru[1], and Anisul Haque[2]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh.
[2]Department of Electrical and Electronic Engineering, East West University, Dhaka-1212, Bangladesh.
E-mail: itrat@eee.buet.ac.bd

Abstract – Gate C-V characteristics of nMOSFETs fabricated on 100 silicon and subjected to a uniaxial strain applied along the <110> direction are studied. MOS electrostatics is calculated by solving self-consistent Schrödinger-Poisson equations including wave-function penetration into the gate dielectrics. It is observed that the gate capacitance increases in strong inversion with strain, but is relatively unaffected by strain in depletion region. This is due to the changes in the electron quantization mass and density of states effective mass with strain. We have also found that the effect of strain on the gate capacitance is not sensitive to changes in the substrate doping density.

I. Introduction

MOS devices have entered the nanometer regime in the past decade [1]. 65 nm node has been commercially reached in 2007. Further scaling of MOS devices is facing several technological challenges. Nonconventional substrates are recently introduced to increase device speed and reliability. Biaxially strained Si MOSFETs have been widely studied for enhanced mobility of charge carriers [2, 3]. Uniaxially strained Si MOSFETs [4] have become more popular in recent days. Uniaxial strain also enhances carrier mobility [5]. The uniaxial straining process is simpler and more compatible with the conventional CMOS technology. By uniaxial strain different types of strain can be applied for pMOS and nMOS devices in a cMOS circuit. Mobility enhancement due to uniaxial strain has extensively been studied [6, 7], but gate C-V characteristics of uniaxially strained MOSFETs have received less attention even though the C-V characteristics is an important tool for characterizing MOS devices.

In this work a self-consistent quantum mechanical simulator is developed to model gate C-V characteristics of nMOS devices with uniaxially strained silicon channel. Uniaxial tensile stress is applied in the <110> direction. <110> direction is taken for its wide practical use in 100 silicon as the transport direction. Tensile stress is commonly applied to nMOSFETs for mobility enhancement. Proper band splitting and effective mass change due to strain are considered. The model is used to study how uniaxial tensile strain affects the gate C-V characteristics of nMOS devices. Physical insights into the effects of strain are also discussed.

II. Theory

A. Band splitting due to uniaxial strain

For relaxed 100 silicon, two separate valleys of conduction bands (Δ_2 and Δ_4) are at a degenerate state. Uniaxial stress removes this degeneracy and there is a splitting between the two valleys. Uniaxial stress also introduces a hydrostatic shift of the mean position of the conduction band. Similarly heavy hole and light hole bands also become nondegenerate under uniaxial stress. Fig. 1 shows the evolution of the energy bands under a uniaxial tensile strain.

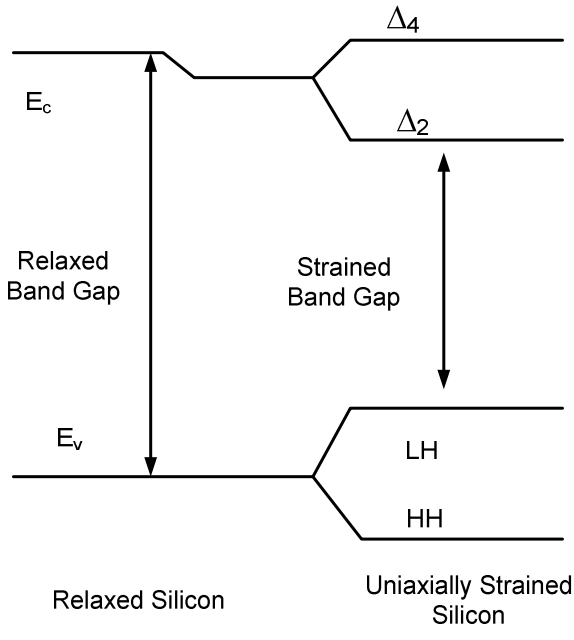


Fig. 1 Band splitting of relaxed 100 silicon under uniaxial tensile stress. Δ_2 and Δ_4 are the two conduction band valleys, LH and HH are for light hole and heavy hole, respectively.

Expressions for splitting of conduction band under a tensile stress along the <110> direction are described in [8] as,

$\Delta E_c^i - \Delta E_c^0 = -\frac{1}{3}(S_{11} - S_{12})\Xi_u X$, for Δ_2 band,

$\Delta E_c^i - \Delta E_c^0 = \frac{1}{6}(S_{11} - S_{12})\Xi_u X$, for Δ_4 band,

$\Delta E_c^0 = \Delta E_{g0} - |E_{\varepsilon\varepsilon}| - \frac{1}{3}(S_{11} - S_{12})\Xi_u X$,and

$\Delta E_{g0} = (\Xi_d + \frac{1}{3}\Xi_u - a)(S_{11} + 2S_{12})X$.

Here,
ΔE_c^0 ≡ Hydrostatic band shifting of the conduction band,
ΔE_c^i ≡ Band splitting for the *i*th valley,
ΔE_{g0} ≡ Change in the energy band gap,
X ≡ Applied stress,
$S_{11} = 8.63\times10^{-12}$ N^{-1}m^2,
$S_{12} = -2.13\times10^{-12}$ N^{-1}m^2,
$\Xi u = 8.6$ eV, and
$\Xi_d + \frac{1}{3}\Xi_u - a = 3.8$ eV.

$2|E_{\varepsilon\varepsilon}|$, the splitting between light hole and heavy hole bands at the band edge, is given by

$$|E_{\varepsilon\varepsilon}| = \frac{1}{2}\left[b^2(S_{11} - S_{12})^2 + 3\left(\frac{d}{2\sqrt{3}} S_{44}\right)^2 \right]^{1/2} |X|$$

Here,
$b = 2.4$ eV,
$d = 5.3$ eV,
$S_{44} = 12.49\times10^{-12}$ N^{-1}m^2.

B. Effective mass variation

Uniaxial strain causes the curvatures of the energy band structures to change. As effective masses depend on the curvature of energy bands, uniaxial stress changes effective masses. We follow the expressions given in [9] to model the variation of effective masses with uniaxial strain applied along <110> direction.

$m_x = 0.918 + 0.0236X^2$
$m_y = 0.196 - 0.016X$
$m_z = 0.196 + 0.029X$
$m_{di1} = \sqrt{m_y m_z}$
$m_{z1} = m_x$
$m_{di2} = \sqrt{m_x m_y}$
$m_{z2} = m_z$

Here, X is the applied stress in GPa. Subscript 1 stands for Δ_2 valley and subscript 2 stands for Δ_4 valley.

C. Self-consistent solver

A self-consistent Schrödinger-Poisson solver is developed to simulate gate C-V characteristics [10]. In the first part, Schrödinger's equation is solved using the logarithmic derivative technique of the retarded Green's function [11]. Open boundary condition is used at the silicon-gate-oxide interface to include penetration of wave functions into the gate dielectric. Our open boundary condition, that considers zero electric field deep inside the gate electrode, as well as deep inside the bulk silicon, incorporates wave function penetration effect naturally without introducing any unphysical artifact. In the second part, Poisson's equation is solved for the combined silicon-oxide regions to include the charge in gate-oxide region due to wave function penetration. Finite difference method is applied and non-uniform grid spacing is used to improve computational efficiency.

III. Results

Fig. 2 shows the gate C-V characteristics of nMOSFETs with relaxed and uniaxially strained silicon channels. Three different stress levels along the <110> direction are considered. Here gate oxide thickness T_{ox} = 2 nm, doping density $N_a = 10^{18}$ cm^{-3}. Flatband voltage is assumed to be zero in all calculations.

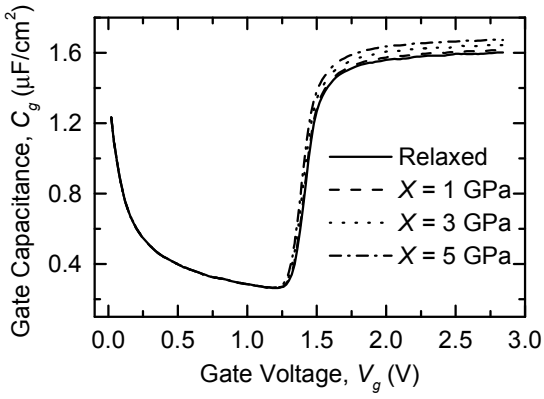


Fig. 2 Gate C-V characteristics of relaxed and uniaxially strained Si MOSFETs.

It is observed that uniaxial strain along the <110> direction has no effect on the gate capacitance in the depletion region and has a small effect in the weak inversion region. There is no shift in the flatband voltage due to uniaxial strain. However the gate capacitance in strong inversion increases under strain. Gate capacitance in strong inversion is increased by 2.7 % when 3 GPa stress is applied. It is also observed that uniaxial strain has a small effect on the threshold voltage V_T. This is in agreement with [12] where it is shown that compared to the biaxially strained MOSFETs, the threshold voltage shift is much smaller in uniaxially strained MOSFETs.

To investigate the factors influencing changes in gate capacitance under strain, we calculate the depletion width, band bending due to depletion charges and the total semiconductor charges as functions of the gate voltage for relaxed and strained silicon channel in Figs. 3, 4 and 5, respectively. We find that uniaxial strain has no effect on the depletion properties. The depletion width, consequently the depletion band bending and the depletion charge remain unaffected by strain. The increase in the total charge due to strain in strong inversion, as observed in Fig. 5, is due to increased charge in the inversion region.

Inversion capacitance ($dQ_{inv}/d\varphi_s$) as a function of gate voltage is shown in Fig. 6. It is clearly seen that C_{inv} increases with increase in uniaxial stress for a given gate

voltage in strong inversion. Strain affects the inversion charges in two ways. First, strain modifies the bandgap E_g and second, strain changes the effective masses of electrons. For MOSFETs fabricated on 100 silicon surface, uniaxial strain increases the quantization effective mass. Consequently the energies of the quasi-bound states are lowered relative to the conduction band minima. For a given gate voltage, the density of the inversion charges is increased, leading to an increase in C_{inv}.

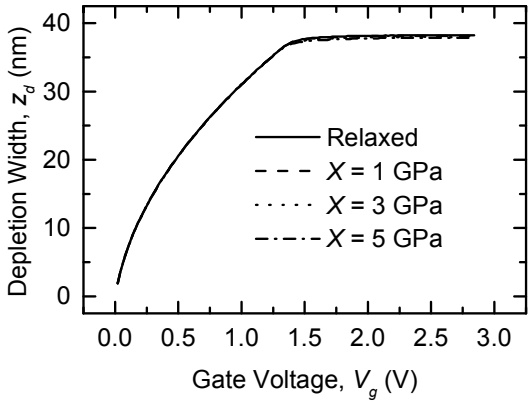


Fig. 3 Depletion width as a function of V_g for relaxed and strained Si MOSFETs.

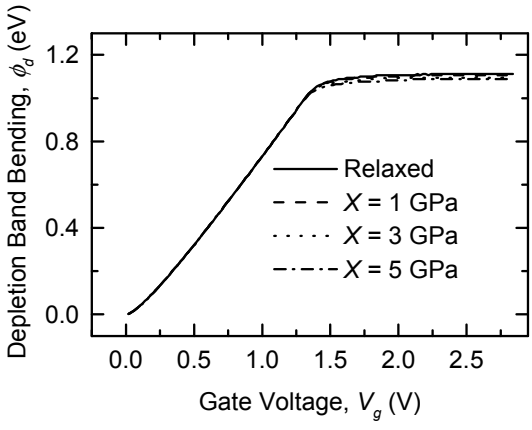


Fig. 4 Band bending due to depletion charges for relaxed and strained Si MOSFETs.

In Fig. 7 we show the C-V characteristics of strained MOSFETs calculated with (i) considering only the bandgap correction (ii) considering both the bandgap and the effective mass corrections. When modification of effective mass is neglected, strain has negligible effect on the gate C-V characteristics. The result shows that for the stress level considered in this work, the increase in the gate capacitance with uniaxial strain is primarily due to the increase in quantization and density of states effective masses and the bandgap modification does not play an important role.

Effect of doping density on C-V characteristics of strained Si devices is shown in Fig. 8. Here doping density level is 5×10^{17} cm^{-3}. Gate capacitance in strong inversion is increased by 2.6 % for 3 GPa stress. This change is nearly the same as that observed in Fig. 2 for N_a = 10^{18} cm^{-3}. Figs 2 and 8 suggest that the effect of

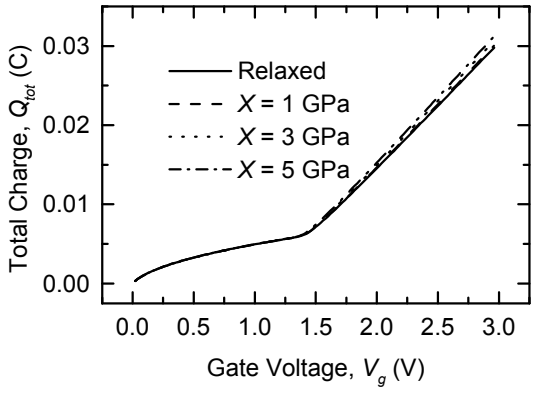


Fig. 5 Semiconductor total charge versus V_g for relaxed and strained Si MOSFETs.

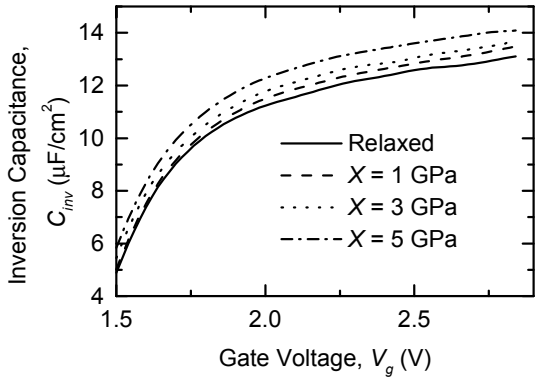


Fig. 6 Inversion capacitance for relaxed and strained Si MOSFETs.

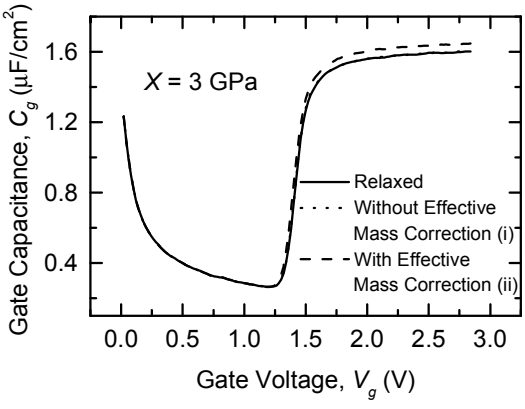


Fig. 7 C-V characteristics with and without considering effective mass modification by strain.

uniaxial strain on the gate C-V characteristics is not very sensitive to changes in the substrate doping density.

IV. Conclusion

We have developed a self-consistent quantum mechanical simulator to investigate the effects of uniaxial strain along the <110> direction on the gate C-V characteristics of MOSFETs fabricated on 100 silicon surface. Changes in both the bandgap and the electron effective masses are included in the model. We show that strain increases the inversion capacitance due to change in quantization and density of states effective masses. However for the stress levels considered in the work, depletion properties are unaffected by strain. It is also found that the effect of strain does not depend on the substrate doping density.

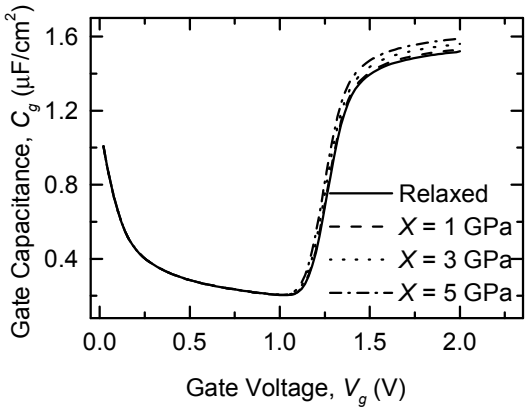


Fig. 8 Gate C-V characteristics of strained and relaxed Si nMOSFETs for substrate doping density N_a = 5x10^{17} cm^{-3}.

References

[1] "International Technology Roadmap for Semiconductors," http://www.itrs.net/common/2004Update.htm.

[2] J. Welser, J. L. Hoyt, and J. F. Gibbons, "NMOS and PMOS Transistors Fabricated in Strained Silicon/ Relaxed Silicon-Germanium Structures," *IEDM Tech. Dig.*, pp. 1000-1002, 1992.

[3] F. M. Buffler and W. Fichtner, "Scaling and Strain Dependence of Nanoscale Strained-Si p-MOSFET Performance," *IEEE Trans. on Electron Devices*, vol. 50, no. 12, pp. 2461–2466, 2000.

[4] S. E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z. Ma, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, and Y. El-Mansy, "A Logic Nanotechnology Featuring Strained-Silicon," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 191-193, 2004.

[5] K. Uchida, T. Krishnamohan, K.C. Saraswat, and Y. Nishi, "Physical Mechanism of Electron Mobility Enhancement in Uniaxial Stress Engineering in Ballastic Regime," *IEDM Tech. Dig.*, pp. 129 – 132, 2005.

[6] F. Rochette, M. Cassé1, M. Mouis, D. Blachier, C. Leroux, B. Guillaumot, G. Reimbold, and F. Boulanger, "Electron Mobility Enhancement in Uniaxially Strained MOSFETs: Extraction of the Effective Mass Variation," *Proceeding of the 36th European Solid-State Device Research Conference,* pp. 93-96, 2006.

[7] E. Ungersböek, S. Dhar, G. Karlowatz, H. Kosina, and S. Selberherr, "Physical Modeling of Electron Mobility Enhancement for Arbitarily Strained Silicon," *J. Comput. Electron.*, vol. 6, pp. 55-58, 2007.

[8] I. Baslev, "Influence of Uniaxial Stress on the Indirect Absorption Edge in Silicon and Germanium," *Phys. Rev.,* vol. 143, no. 2, pp. 636-647, 1966.

[9] S. Dhar, E. Ungersböck, H. Kosina, T. Grasser, and S. Selberherr, "Electron Mobility Model for 110 Stressed Silicon Including Strain-Dependent Mass," *IEEE Trans. on Nanotechnology*, vol. 6, no. 1, pp. 97-100, 2007.

[10] A. Haque and M. Z. Kauser, "A comparison of Wavefunction Penetration Effects on Gate Capacitance in Deep Submicron n- and p-MOSFETs," *IEEE Trans. on Electron Devices*, vol. 49, no. 9, pp. 1580–1587, 2002.

[11] A. Haque and A.N. Khondker, "An Efficient Technique to Calculate the Normalized Wave Functions in Arbitrary One-Dimensional Quantum Well Structures," *J. Appl. Phys.*, vol. 84, no. 10, pp. 5802–5804, 1998.

[12] J. Lim, S.E. Thompson, and J.G. Fossum, "Comparison of Threshold-Voltage Shifts for Uniaxial and Biaxial Tensile-Stressed n-MOSFETs," *IEEE Electron Device Lett.*, vol. 25, no. 11, pp. 731-733, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Inversion Layer Properties of <110> Uniaxially Strained Silicon n-Channel MOSFETs

Samia Nawar Rahman, Hasan Mohammad Faraby[*], Md Manzur Rahman, Md. Quamrul Huda, and Anisul Haque[**]

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.
[*]Department of Electrical and Electronic Engineering United International University, Dhaka, Bangladesh.
[**]Department of Electrical and Electronic Engineering, East West University, Dhaka, Bangladesh.
E-mail:thesisgroup.buet@gmail.com

***Abstract*--This paper discusses the influence of <110> uniaxial tensile stress on some of the inversion layer properties of (100) silicon n-channel MOSFETs. Quantum mechanical calculations are performed assuming Airy function approximation holds. Uniaxial tensile strain lowers the eigen-energies and increases the occupation of the ground state. Average inversion layer penetration is also decreased. The change in the surface electric field due to strain is insignificant.**

I. Introduction

Strain improves the performance of the MOSFETs. Mobility enhancement is the main advantage of applying strain [1]. The mobility enhancement is due to the decrease in the effective mass in the Δ2 valley and the scattering of electrons [2]. The change in effective mass also lowers the eigen-energies which enhances electron occupancy in the ground state. Strain can be applied either biaxially or uniaxially. The process of applying uniaxial strain is technologically easier and more cost effective than biaxial strain [3]. But, the inversion layer properties of uniaxially strained silicon nMOS devices have not been studied in detail so far. Some of the main properties have been discussed in this paper.

Quantum mechanical analysis of the uniaxially strained silicon nMOS devices are performed using the Airy function approximation. Relationship between <110> stress and strain is also derived.

II. Stress - Strain Relationship

The generalized and simplified Hooke's law for cubic crystal is [4],

$$\begin{pmatrix}\in_1\\ \in_2\\ \in_3\\ \in_4\\ \in_5\\ \in_6\end{pmatrix} = S\times\begin{pmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\\ \sigma_{23}\\ \sigma_{13}\\ \sigma_{12}\end{pmatrix}$$

Here,

$$S=\begin{pmatrix} S_{11} & S_{12} & S_{12} & 0 & 0 & 0\\ S_{12} & S_{11} & S_{12} & 0 & 0 & 0\\ S_{12} & S_{12} & S_{11} & 0 & 0 & 0\\ 0 & 0 & 0 & S_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & S_{44} & 0\\ 0 & 0 & 0 & 0 & 0 & S_{44}\end{pmatrix}$$

Where S is the compliance matrix, S_{ij} are the tensor components. Here, S_{11} = 7.681 TPa^{-1}, S_{12} = -2.138 TPa^{-1} and S_{44} = 12.56 TPa^{-1} are material constants [5]. For a uniaxial stress σ applied along <110> direction, the strain tensor components can been calculated as,

$$\in_1=\in_2=(S_{11}+S_{12})\times\sigma/2 \quad (1)$$

$$\in_3=S_{12}\times\sigma \quad (2)$$

$$\in_6=S_{44}\times\sigma/2 \quad (3)$$

The valence band edge shift [6] is given by,

$$\Delta E_v = a(\in_1+\in_2+\in_3)\pm\left[b^2(\in_2-\in_3)^2+\left(\frac{d^2}{4}\right)\in_6{}^2\right]^{\left(\frac{1}{2}\right)} \quad (4)$$

The conduction band edge shift [6] for i^{th} valley is given by,

$$\Delta E_c^{(i)} = \Xi_d(\in_1+\in_2+\in_3)+\Xi_u\in_i \quad (5)$$

Here Ξ_d and Ξ_u are deformation potentials for conduction band given in TABLE I. Putting (1), (2), (3) in (4) and (5), ΔE_c, ΔE_v and therefore ΔE_g can be calculated. This change in bandgap energy affects the inversion layer properties of silicon n-channel MOSFETs.

III. Inversion Layer Properties

The Schrödinger Equation is solved for the analysis of MOS inversion layer. The electronic wave function [7] is given by

$$\Psi(x,y,z) = \xi_i(z)e^{i\gamma z}e^{ik_1x+ik_2y}$$

where k_1 and k_2 are measured relative to the band edge, γ depends on k_1 and k_2, the envelope function $\xi_i(z)$ is obtained [1] from the solution of the following equation:

$$d^2\xi_i/dz^2 + (2m_3/\hbar^2)[E_i + e\Phi(z)]\xi_i(z) = 0$$

Here m_3 is the effective mass in the direction perpendicular to the oxide-semiconductor interface and E_i is the eigen-energy of the i^{th} bound state in the same direction. N_i is the carrier concentration in the i^{th} sub-band and is given by,

$$N_i = (n_{vi}m_{di}KT/\pi\hbar^2)F_0[(E_F - E_i)/KT] \quad (6)$$

Where, n_{vi} is the valley degeneracy, m_{di} is the density of states effective mass per valley, E_F is the Fermi Energy, $F_0(x) = \ln(1+e^x)$, is the function used in (6).

The depletion layer thickness z_d is given by,

$$z_d = [2k_{sc}\varepsilon_0\Phi_d/eN_a]^{1/2} \quad (7)$$

The surface electric field F_s is given by,

$$F_s = e(N_{inv} + N_{depl})/k_{sc}\in_0 \quad (8)$$

$$N_{depl} = z_d N_a \quad (9)$$

where, k_{sc} is the dielectric constant, N_{depl} is the number of charges per unit area in the depletion layer and $N_{inv} = \sum N_i$ is the total number of charges per unit area in the inversion layer, N_a is the net doping density.

IV. Calculations

We performed all our calculations assuming (100) silicon nMOSFETs which has a doping density of 10^{15} cm^{-3} in the bulk. We varied the inversion layer charge density from 10^8cm^{-2} to 10^{12}cm^{-2}. We used the effective masses and other parameters as given in TABLE II. <110> stress of 3.5GPa is applied over the gate of the nMOSFET. The total band bending at the surface, Φ_s is the summation of the band bending due to the depletion layer, Φ_d and the band bending due to the inversion layer, Φ_{inv}. $e\Phi_f$ is the difference in energy between intrinsic Fermi level E_{fi} and E_f.

At the moment of strong inversion

$$\Phi_f = (KT/e)\ln(N_a/n_i), \quad \Phi_d = 2\Phi_f \quad (10)$$

$$\Phi_{inv} = eN_{inv}z_{av}/k_{sc}\in_0 \quad (11)$$

Here z_{av} is the average penetration of the inversion layer from the surface. From [7], using Airy function approximation, the eigen-energies are given by,

$$E_i = (\hbar^2/2m_3)^{1/3}[3/2\pi eF_s(i+3/4)]^{2/3} \quad (12)$$

The average penetration of the carriers in the i^{th} sub-band z_i from the surface is measured as

$$z_i = 2E_i/3eF_s \quad (13)$$

$$E_f = \Phi_d \times e/2 \quad (14)$$

The effective mass changes due to strain [8] according to

$$m_3 = 0.914 + 0.0236 \times \sigma^2 \quad (15)$$

Here m_3 denotes the out of plane effective mass of $\Delta 2$ valley.

From (10), Φ_d, was found. From Φ_d we calculated the depletion layer width, z_d and then N_{depl}. From N_{inv} and N_{depl} the surface electric field (8) was obtained. Using F_s and the new effective mass we calculated eigen-energies (12). Using the new eigen-energies we calculated the average penetration z_i by (13) and the carrier concentration, N_i by (6). Values of m_{di} and n_{vi} were taken from TABLE II. z_{av} was calculated using the following equation,

$$z_{av} = \frac{\sum_i z_i N_i}{\sum_i N_i} \quad (16)$$

Strain changes the bandgap [6] by the following equation,

$$\Delta E_g(\sigma) = -6.19\varepsilon \quad (17)$$

Here $\Delta E_g(\sigma)$ is the change in band-gap due to stress. ε is the strain due to the applied stress. This change in band-gap causes the change in the intrinsic carrier concentration (n_i) which is calculated by ,

$$n_i = \sqrt{N_c N_v}\exp(-(E_g + \Delta E_g(\sigma))/2kT) \quad (18)$$

This changed n_i is used in (10) to calculate Φ_f and from (10) we get the band-bending due to depletion layer. From this we get z_d by (7) and N_{depl} from (9). Thus we get the new depletion layer carrier concentration including the effect of strain. Using (11) & (16) we calculated Φ_{inv}. By adding Φ_d and Φ_{inv} we get the total band bending at the surface due to strain.

V. Results and Discussions

When uniaxial stress is applied to silicon, bandgap narrowing occurs. Smaller bandgap causes increase in intrinsic carrier concentration of strained silicon. The Fermi level moves towards the conduction band. So, to achieve inversion we will need less gate voltage. This is evident from Fig.1.

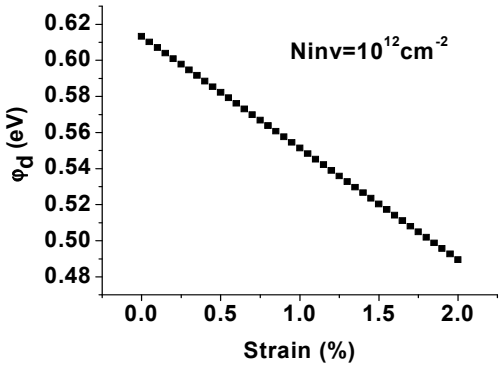


Fig. 1 Effect of strain on band bending for constant level of inversion.

Inversion is expected at a lower level of bending in presence of strain. To achieve inversion of the same strength, band bending at the surface will be lower for strained silicon than that of unstrained silicon. So, depletion width also decreases. This decreases the depletion charge density. So surface field decreases, but the amount of decrease is insignificant. Therefore, decrease in eigen-energy due to surface field is negligible but increase in effective mass decreases the eigen-energies significantly which can be seen from Fig. 2. The same trend was found for higher eigen-energies as well.

Strain also changes the Fermi Level, E_f. As a result energy difference between E_f and the ground state gets reduced. So the Fermi level will be closer to the ground state in the strained case than in the unstrained case. This phenomenon can be seen in Fig. 3.

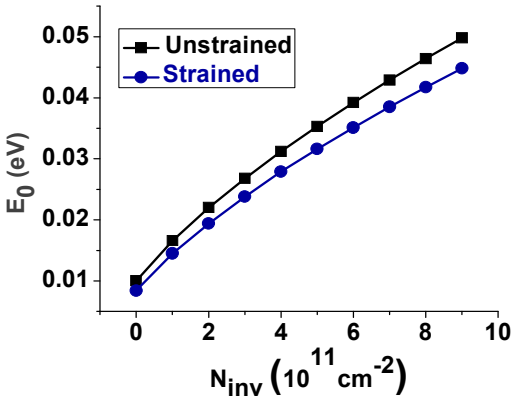


Fig. 2 Ground state energy (E_0) as a function of N_{inv}.

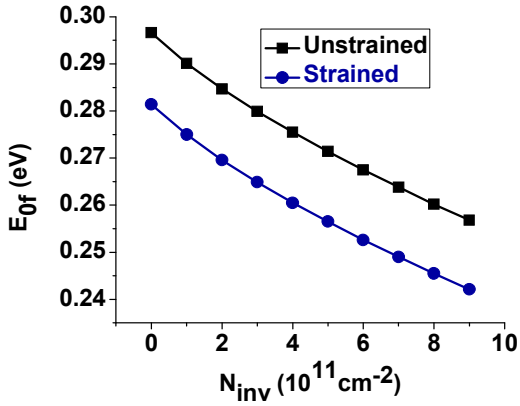


Fig. 3 Ground state energy (E_0) with respect to Fermi level (E_f) as a function of N_{inv}.

As the ground state energy decreases, more and more electrons gather in the lowest state. So the electron occupancy increases. This is seen from Fig. 4.

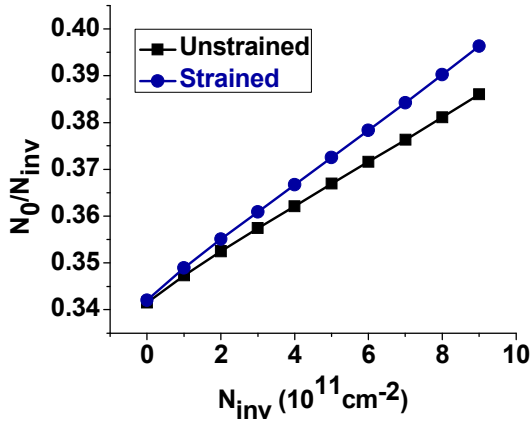


Fig. 4 Electron occupancy of the ground state (N_0/N_{inv}) as a function of N_{inv}.

For a fixed N_{inv}, the electron occupancy in the higher states decreases. This is seen in Fig. 5 for the E_1 state.

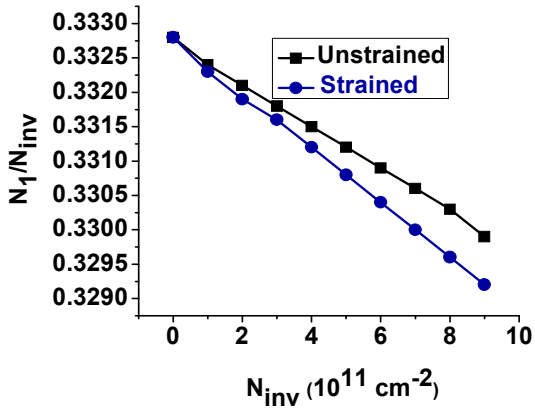


Fig. 5 Electron occupancy of first eigen state (N_1/N_{inv}) as a function of N_{inv}.

As the depletion width decreases the average penetration decreases with decrease in E_i as seen from (13) which is also evident from Fig. 6.

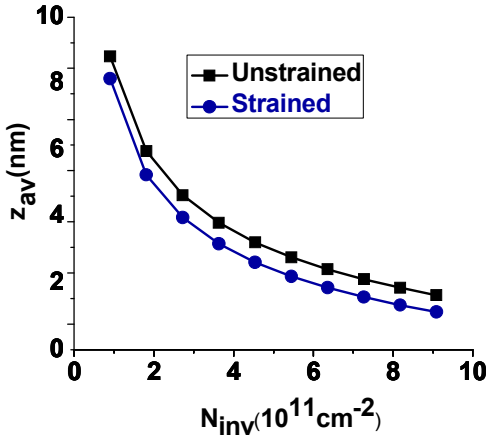


Fig. 6 Average inversion layer penetration as a function of N_{inv}.

VI. Conclusions:

In conclusion, the effect of <110> uniaxial stress on silicon nMOSFET inversion properties has been investigated and quantified. The results obtained from the Airy Function approximation qualitatively describe the trends of the effect of strain. Smaller band bending, larger occupancy of ground state and smaller depth of penetration is expected in uniaxially strained silicon.

Table 1 Values of Deformation Potential constants [6]

Stress Type	Deformation Potential Constants [eV]		ΔE_g [eV]
[110] Uniaxial	$\Xi_d = 1.13$	$a_v = 2.46$	-6.19ε
	$\Xi_u = 9.16$	$b_v = -2.35$	

Table 2 Various effective masses and constants [7]

Surface		(100)	
Valleys		Lower	Higher
Degeneracy	n_v	2	4
Density of States mass per valley	m_d / m_o	0.190	0.417
Dielectric Constant	k_{sc}	11.7	
Permittivity	$k_{sc} \in_0$	1.04×10^{-10}F/m	

Acknowledgements

We would like to express our sincere gratitude to Mr. Itrat Bin Shams, Lecturer, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET) Dhaka, for his constructive advice, criticism and encouragement during our research work.

References:

[1] F. Rochette, M. Cassé, M. Mouis, D. Blachier, C. Leroux, B. Guillaumot, G. Reimbold, and F. Boulanger, "Electron mobility enhancement in uniaxially strained MOSFETs: Extraction of the effective mass variation," in Proc. Solid-State Device Research Conference, 2006. (ESSDERC 2006) the 36th European, pp. 93-96, September 2006.

[2] E. Ungersboeck, V. Sverdlov, H. Kosina, and S. Selberherr, "Electron inversion layer mobility enhancement by uniaxial stress on (001) and (110) oriented MOSFETs," in Proc. The 2006 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD '06), pp. 43-46, September 2006.

[3] Scott E. Thompson, Guangyu Sun, Youn Sung Choi, and Toshikazu Nishida, "Uniaxial-Process-Induced Strained-Si: Extending the CMOS Roadmap," IEEE Transactions on electron Devices, vol. 53, no. 5, pp.1010-1020, May 2006.

[4] John. P. *Loehr, Physics of Strained Quantum Well Lasers,* Kluwer Academic Publishers, Boston/Dordrecht/London.

[5] Nextnano3, '1D strained silicon' [http://www.wsi.tumuenchen.de/nextnano3/tutorial/1D tutorial_s trained_Si.htm].

[6] J. S. Lim, S. E. Thompson, and J. G. Fossum, "Comparison of Threshold-Voltage shifts for Uniaxial and Biaxial Tensile -stressed n –MOSFETs," IEEE Electron Device Letters, Vol. 25, No. 11, pp. 731-733, November 2004.

[7] F. Stern, 'Self-Consistent Results for n- type Si Inversion Layers', Physical Review B, Vol. 5, No. 12, pp. 4891-4899, June 1972.

[8] E. Ungersboeck, S. Dhar, G. Karlowatz, H. Kosina, and S. Selberherr, "Physical modeling of electron mobility enhancement for arbitrarily strained silicon," Journal of Computational Electronics, Vol. 6, pp. 55-58, December 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Analytical Surface Potential Model For Pocket Implanted Sub-100 nm n-MOSFET

Muhibul Haque Bhuyan[1,2] and Quazi D. M. Khosru[2]

[1]Department of Electronics and Telecommunication Engineering
Daffodil International University, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
E-mail: muhibulhb@yahoo.com

Abstract - This paper presents an analytical surface potential model for pocket implanted sub-100 nm n-MOSFET. The model is derived by solving the Poisson's equation in the depletion region at the surface with the appropriate boundary conditions at source and drain. The model includes the effective doping concentration of the two linear pocket profiles at the source and drain sides of the device. The model also incorporates the drain and substrate bias effect below and above threshold conditions. The simulation results show that the derived surface potential model has a simple compact form that can be utilized to study and characterize the pocket implanted advanced ULSI devices.

I. Introduction

As the channel length of MOSFETs is scaled down to deep-submicrometer or sub-100 nm regime, we observe the reduction of threshold voltage with the reduction of channel length due to the charge sharing between the drain/source region and the channel [1]. Also, the off-state leakage current increases due to sensitivity of the source/channel barrier to the drain potential or drain induced barrier lowering (DIBL). This effect is known as short-channel effect (SCE). This effect arises as a result of two dimensional potential distribution and high electric fields in the channel region [2]. It can be reduced or can be even reversed (then it is called reverse short channel effect or RSCE) by locally raising the channel doping near source and drain junctions. RSCE was originally observed in MOSFETs due to oxidation-enhanced-diffusion [3] or implant-damage-enhanced diffusion [4] which are very difficult to control. Lateral channel engineering utilizing halo or pocket implant [5-9] surrounding drain and source regions is effective in suppressing short channel effects. The halo or pocket implant can be either symmetrical [10] or asymmetrical [11] with respect to source or drain. Reported circuit applications include a 256 M-bit DRAM [12] and mixed-signal processor [13]. In fact, this pocket implant technology is found to be very promising in the effort to tailor the short-channel performances of deep-submicron as well as sub-100 nm MOSFETs although careful tradeoffs need to be made between minimum channel length and other device electrical parameters [6].

Solution of the Poisson's equation in the depletion region of the MOSFETs is an important step in order to determine the surface potential. Numerical device simulators like MEDICI [14] can produce most accurate solutions of the Poisson's equation. But analytical models that have been used for MOSFET device design, take less time for device simulation. It also provides device physics insight [15]. Analytical model shown in [6] does not satisfy the boundary conditions and device simulation results of MEDICI. Analytical model in [15] assume a step profile of pocket doping. Besides, Gaussian profile [16] and hyperbolic cosine function [17] were assumed for the pocket profile to derive the threshold voltage equations. But it has been observed that linear profile of pocket doping produce better results for threshold voltage [18]. The previous works are on the threshold voltage modeling of pocket implanted MOS devices.

In this paper, an analytical model that can predict the surface potential of the sub-100 nm pocket implanted n-MOSFET is derived assuming the linear profile of pocket doping. Here the 1-D pocket profile across the channel has been transformed to an effective doping concentration expression, which is used in the 1-D Poisson's equation to derive the model applying the appropriate boundary conditions at the source and drain. Simulation results show that the model predicts surface potential very well for various device and pocket profile parameters as well as various bias conditions, and also satisfies the boundary conditions. It proves the validity and usefulness of our proposed model for circuit simulation of next generation ULSI devices.

II. Surface Potential Model

The pocket implanted n-MOSFET structure shown in Fig. 1 is considered in this work and assumed co-ordinate system is shown at the right side of the structure. All the device dimensions are measured from the oxide-silicon interface. In the structure, the junction depth (r_j) is 25 nm. The oxide thickness (t_{ox}) is 2.5 nm, and it is SiO_2 with fixed oxide charge density of 10^{11} cm^{-2}. Uniformly doped p-type Si substrate is used with doping concentration of $N_{sub} = 4.2\times10^{17}$ cm^{-3} with pocket implantation both at the source and drain side with peak pocket doping concentration of 1.75×10^{18} cm^{-3} and pocket length from 25 nm, and the source or drain doping concentration of 9.0×10^{20} cm^{-3}.

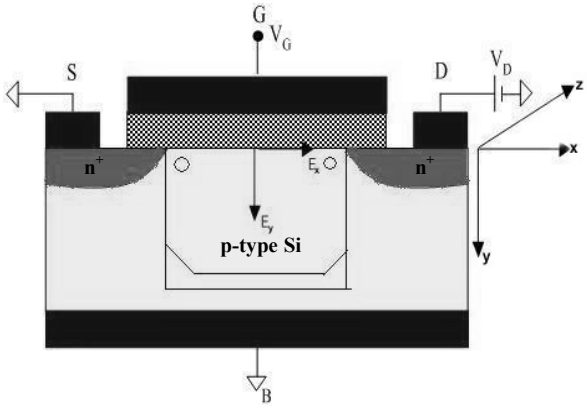


Fig. 1 Pocket implanted n-MOSFET Structure

The pocket implantation, which causes the Reverse Short Channel Effect (RSCE), is done by adding impurity atoms both from the source and drain edges. It is assumed that the peak pocket doping concentration, N_{pm} gradually decreases linearly towards the substrate level concentration, N_{sub} with a pocket length, L_p from both the source and drain edges. The basis of the model of the pocket profile is to assume two linear doping profiles from both the source and drain edges across the channel as shown in Figs. 2-3. The pocket parameters, N_{pm} and L_p, play important role in determining the RSCE.

At the source side, the pocket profile is given as:

$$N_s(x) = -\frac{N_{pm} - N_{sub}}{L_p}x + N_{pm}$$

$$\therefore N_s(x) = N_{sub}\frac{x}{L_p} + N_{pm}\left(1 - \frac{1}{L_p}x\right) \qquad (1)$$

At the drain side, the pocket profile is given as:

$$N_d(x) = \frac{N_{pm} - N_{sub}}{L_p}\left[x - \left(L - L_p\right)\right] + N_{sub}$$

$$\therefore N_d(x) = N_{sub}\left(\frac{L}{L_p} - \frac{1}{L_p}x\right) + N_{pm}\left(1 - \frac{L}{L_p} + \frac{1}{L_p}x\right) \qquad (2)$$

where x represents the distance across the channel. Since the pile-up profile is due to the direct pocket implant at the source and drain side, it is assumed symmetric at both sides.

With these two conceptual pocket profiles, the profiles are integrated mathematically along the channel length and then divided by it to derive an average effective concentration as in equation (3):

$$N_{eff} = \frac{1}{L}\int_0^L \left[N_s(x) + N_d(x) + N_{sub}\right]dx \qquad (3)$$

Putting the expressions of $N_s(x)$ and $N_d(x)$ from equations (1) and (2) in equation (3) to obtain equation (4) for the effective doping concentration along the channel.

$$N_{eff} = N_{sub}\left(1 - \frac{L_p}{L}\right) + \frac{N_{pm}L_p}{L} \qquad (4)$$

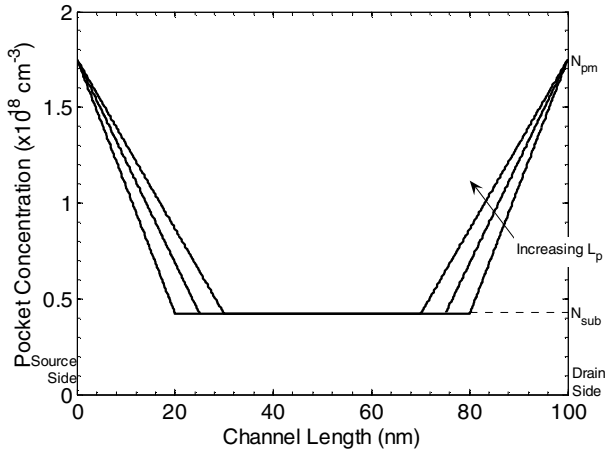


Fig. 2 Simulated pocket profiles at the surface for different pocket lengths, L_p = 20, 25 and 30 nm. Peak pocket concentration, $N_{pm} = 1.75\times10^{18}$ cm^{-3}.

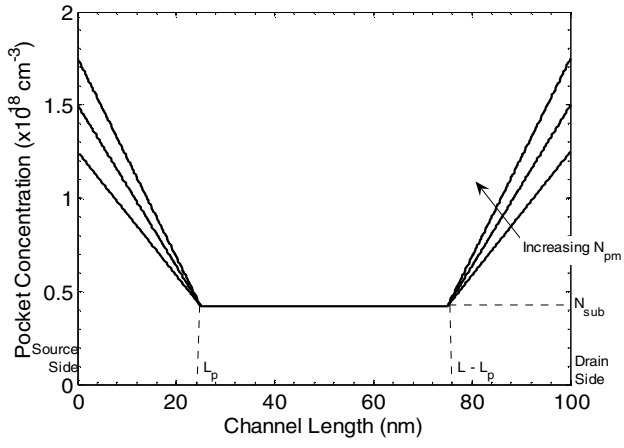


Fig. 3 Simulated pocket profiles at the surface for various peak pocket concentrations, $N_{pm} = 1.25\times10^{18}$, 1.5×10^{18} and 1.75×10^{18} cm^{-3}. Pocket length, L_p = 25 nm

When L_p << L for long channel device then pocket profile has no effect on the surface potential, But when L_p is comparable with L for short channel device then there is a effect of pocket profile on the surface potential.

In the depletion region, by applying Gauss's law, we can write the following equation incorporating the effect of substrate bias and effective doping concentration along the channel

$$\varepsilon_s\frac{X_D}{\eta}\frac{d\mathrm{E}}{dx} + \varepsilon_{ox}\frac{V_{GS} - V_{BS} - V_{FB} - \psi_s(x)}{t_{ox}} = qN_{eff}X_D \qquad (5)$$

where E, ε_s, ε_{ox}, V_{GS}, V_{BS}, V_{FB}, $\Psi_s(x)$, t_{ox} and q is are the electric field, dielectric permittivity of Si and oxide, gate and substrate bias, flat band voltage, surface potential, oxide thickness and electronic charge. η is a fitting parameter, and it is assumed 1 everywhere. The depletion layer thickness, X_D is given by the expression given in equation (6).

$$X_D = \sqrt{\frac{2\varepsilon_s(\varphi_F - V_{BS})}{qN_{eff}}} \quad (6)$$

where φ_F is the Fermi potential due to pocket implantation and is given as follows

$$\varphi_F = \frac{kT}{q}\ln\frac{N_{eff}}{n_i} \quad (7)$$

The lateral electric field is defined as

$$\mathrm{E} = \frac{d\psi_s(x)}{dx} \quad (8)$$

Therefore, from equation (5) we can write

$$\varepsilon_s \frac{X_D}{\eta}\frac{d^2\psi_s(x)}{dx^2} + \varepsilon_{ox}\frac{V_{GS} - V_{BS} - V_{FB} - \psi_s(x)}{t_{ox}} = qN_{eff}X_D \quad (9)$$

Now we assume the following boundary conditions:

1. At $x = 0$, i.e. at the source side, the surface potential is $\psi_s(0) = \varphi_{bi} - V_{BS}$.

2. At $x = L$, i.e. at the drain end, the surface potential is $\psi_s(L) = \varphi_{bi} - V_{BS} + V_{DS}$.

After solving the 2nd order differential equation (9) using the above two boundary conditions we get the desired complete analytical expression for the surface potential as follows

$$\psi_s(x) = \frac{c_1}{\sinh\sqrt{\frac{a_0}{a_2}}L}\sinh\sqrt{\frac{a_0}{a_2}}(L - x) + \frac{c_1 + V_{DS}}{\sinh\sqrt{\frac{a_0}{a_2}}L}\sinh\sqrt{\frac{a_0}{a_2}}x - \frac{b_1}{a_0} \quad (10)$$

where the parameters a_0, a_2, b_1 and c_1 are given by the following expressions

$$a_0 = \frac{\varepsilon_{ox}}{t_{ox}}, \quad a_2 = \frac{\varepsilon_s}{\eta}X_D$$

$$b_1 = qN_{eff}X_D - \frac{\varepsilon_{ox}}{t_{ox}}(V_{GS} - V_{BS} - V_{FB})$$

$$c_1 = \varphi_{bi} - V_{BS} + \frac{b_1}{a_0}$$

The solution is obtained by finding the transient solution and particular integral and then adding them together.

III. Results and Discussion

In order to verify the analytical surface potential model for the pocket implanted n-MOSFET, different types of simulations were performed. Figure 4 shows the variation of surface potential along the channel for different drain biases. It has been observed that as the drain bias increases surface potential increases at the drain side whereas it remains constant at the source side. It proves the validity of our assumed boundary conditions while deriving the model.

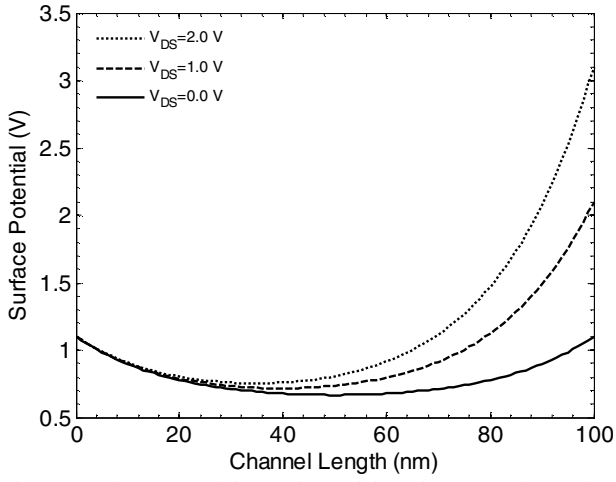


Fig. 4 Surface potential vs. channel length curves for various drain biases with channel length, L = 100 nm and substrate bias, V_{BS} = 0.0 V

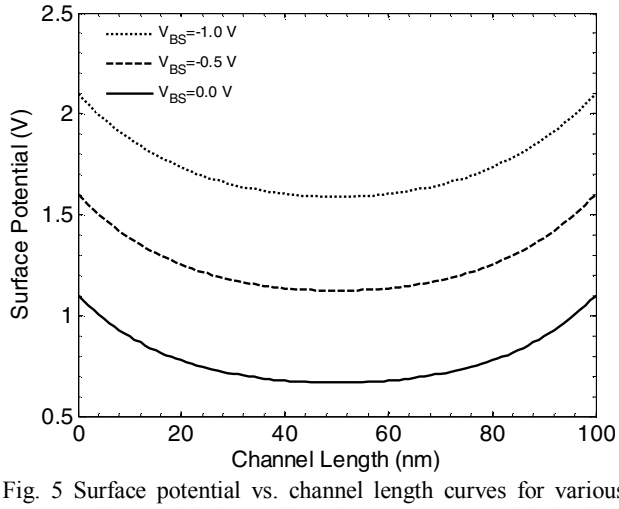


Fig. 5 Surface potential vs. channel length curves for various substrate biases with channel length, L = 100 nm and drain bias, V_{DS} = 0.0 V

From Fig. 5, it is observed that as the substrate bias increases in the negative direction keeping drain bias constant at 0 V, both sides of the curve shifts upward. Since when the substrate bias increases depletion charge increases thereby increasing the surface potential.

From Fig. 6, it is found that as the channel lengths are decreased from 100 nm to 50 nm the surface potential is same as in Fig. 4 for V_{DS} = 1.0 V at both sides of the device. But the potential minimum shifts upward direction as the channel lengths are decreased down. This occurs due to the widening of the depletion region under the gate.

Figure 7 shows the variation of surface potential along the channel for different oxide thicknesses with zero substrate bias and drain bias, V_{DS} = 1.0 V. It is observed that as the oxide thickness decreases the potential minimum

increases near the source side. But near the drain, opposite phenomenon is observed. When oxide thickness decreases oxide capacitance increases. This increases the surface charge and hence surface potential for a fixed bias condition. Since at the drain side with the reduction of the oxide thickness, the oxide capacitance increases as well as the off-state current increases thus the potential at the drain side decreases. Hence DIBL effect will be more pronounced as the oxide thickness is decreased.

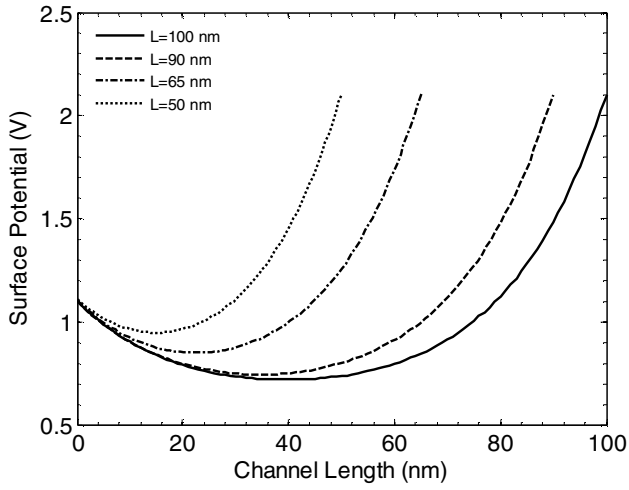


Fig. 6 Surface potential curves along the channel for various channel lengths with substrate bias, V_{BS} = 0.0 V and drain bias, V_{DS} = 1.0 V

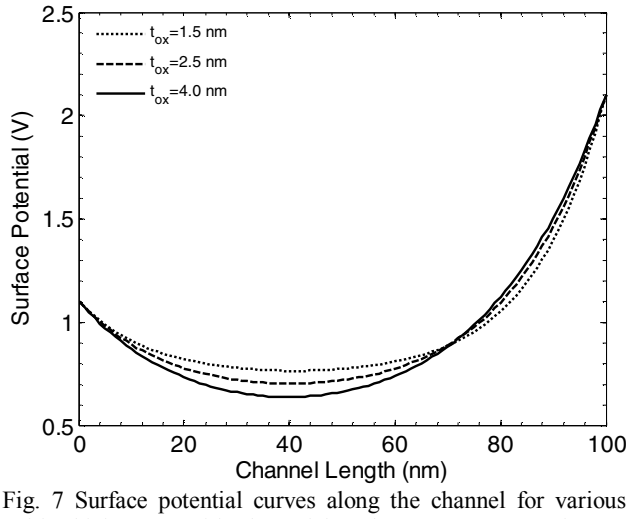


Fig. 7 Surface potential curves along the channel for various oxide thicknesses with channel length, L = 100 nm, substrate bias, V_{BS} = 0.0 V and drain bias, V_{DS} = 1.0 V

Figure 8 shows that the surface potential increases with the decreasing pocket lengths. Since when the pocket length decreases the effective doping concentration also decreases and thus depletion charge decreases. But the boundary value remains the same as expected.

Figure 9 shows the surface potential variation with the position of the channel for different peak pocket doping concentration. It is observed the surface potential increases as the peak pocket doping concentration increases. This is due to the increase of effective carrier concentration along the channel. The results are shown for zero substrate bias and drain bias of 1 V.

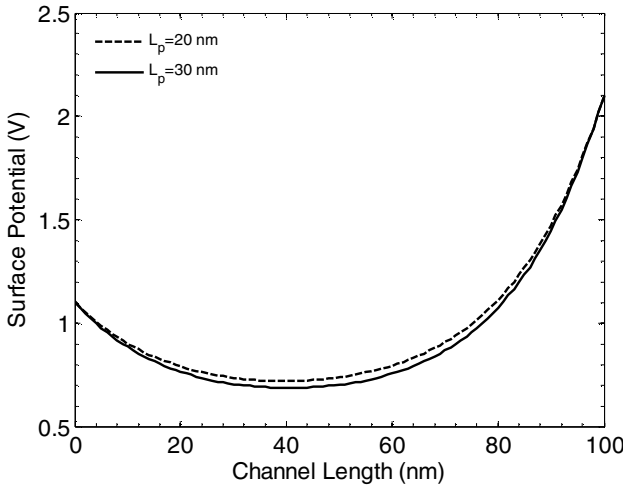


Fig. 8 Surface potential curves along the channel for various pocket lengths with channel length, L = 100 nm, substrate bias, V_{BS} = 0.0 V and drain bias, V_{DS} = 1.0 V

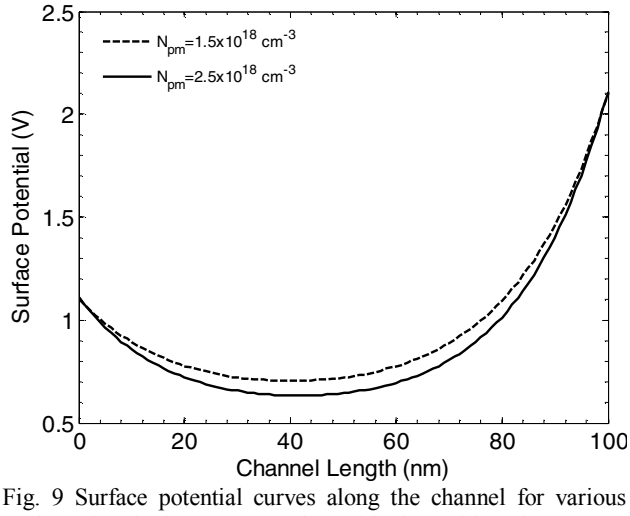


Fig. 9 Surface potential curves along the channel for various peak pocket doping concentration with channel length, L = 100 nm, substrate bias, V_{BS} = 0.0 V and drain bias, V_{DS} = 1.0 V

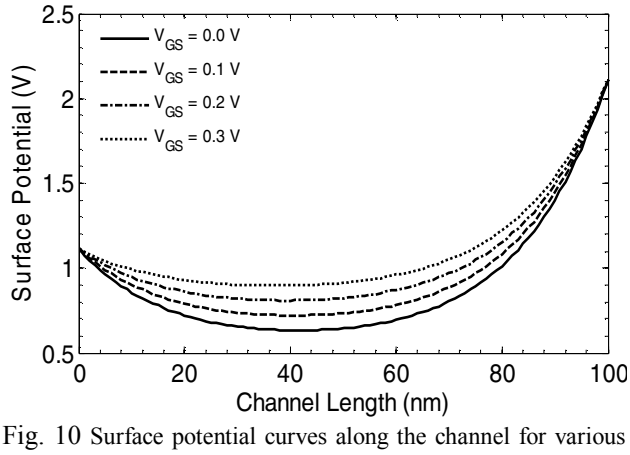


Fig. 10 Surface potential curves along the channel for various peak gate voltages below the threshold voltage with channel length, L = 100 nm, substrate bias, V_{BS} = 0.0 V and drain bias, V_{DS} = 1.0 V

Figure 10 shows the variation of surface potential along the channel for different gate voltages below the threshold voltage with channel length of 0.1 μm. It is observed from this figure that as the gate voltage is increased the peak of

potential minimum shifts upward without changing the boundary values. The reason for increase of the surface potential can be attributed to the increase of the depletion layer charge with gate bias.

IV. Conclusion

An analytical surface model for ultra thin oxide and sub-100 nm pocket implanted n-MOSFET has been developed incorporating the substrate and drain bias dependence. The model is developed assuming two linear pocket profiles along the channel at the surface of the MOS device from the source and drain edges. The effect of changing the device and pocket profiles parameters on the surface potential have been studied using the proposed model. The simulated results show that the proposed model satisfies the boundary conditions as well as predicts the surface potential down to 50 nm channel length. It shows the expected results when the various parameters are changed. Hence this model efficiently determines the surface potential of scaled pocket n-MOSFETs having channel lengths in sub-100 nm regime.

References

[1] S. M. Sze, "Physics of Semiconductor Devices," 2nd Edition, John Wiley & Sons, New York, ch. 8, 1981.

[2] M. Miura-Mattausch, M. Suetake, H. J. Mattausch, S. Kumashiro, N. Shigyo, S. Oganaka and N. Nakayama, "Physical modeling of the reverse short channel effect for circuit simulation," IEEE Trans. on Electron Devices, vol. 48, pp. 2449-2452, Oct. 2001.

[3] M. Orlowski, C. Mazure and F. Lau, "Submicron short channel effects due to gate reoxidation induced lateral interstitial diffusion," IEEE IEDM Tech. Digest, p. 632, 1987.

[4] M. Nishida and H. Onodera, "An anomalous increase of threshold voltage with shortening the channel lengths for deeply boron-implanted n-channel MOSFETs," IEEE Trans. on Electron Devices, vol. 48, pp. 1101, 1981.

[5] K. Y. Lim and X. Zhou, "Modeling of Threshold Voltage with Non-uniform Substrate Doping," in Proc. of the IEEE Int'l Conf. on Semiconductor Electronics (ICSE'98), Malaysia, pp. 27-31, 1998.

[6] B. Yu, C. H. Wann, E. D. Nowak, K. Noda and C. Hu, "Short Channel Effect improved by lateral channel engineering in deep-submicrometer MOSFETs," IEEE Transactions on Electron Devices, vol. 44, pp. 627-633, Apr. 1997.

[7] B. Yu, H. Wang, O. Millic, Q. Xiang, W. Wang, J. X. An and M. R. Lin, "50 nm gate length CMOS transistor with super-halo: Design, process and reliability," IEDM Tech. Digest, pp. 653-656, 1999.

[8] K. M. Cao, W. Liu, X. Jin, K. Vasant, K. Green, J. Krick, T. Vrotsos and C. Hu, "Modeling of pocket implanted MOSFETs for anomalous analog behavior," IEEE IEDM Tech. Digest, pp. 171-174, 1999.

[9] Y. S. Pang and J. R. Brews, "Models for subthreshold and above subthreshold currents in 0.1 μm pocket n-MOSFETs for low voltage applications," IEEE Transactions on Electron Devices, vol. 49, pp. 832-839, May 2002.

[10] J. Tanaka, S. Kimura, H. Noda, T. Toyabe, and S. Ihara, "A sub-0.1 μm grooved gate MOSFET with high immunity to short channel effects," IEEE IEDM Tech. Digest, p. 537, 1993.

[11] T. N. Buti, S. Ogura, N. Rovedo, K. Tobimatsu, and C. F. Codella, "Asymmetrical halo source GOLD drain (HS-GOLD) deep-half-micron n-MOSFET design for reliability and performance," IEEE IEDM Tech. Digest, pp. 617, 1989.

[12] A. Chatterjee, J. Liu, P. K. Mozumder, M. Rodder and I. C. Chen, "Pass transistor designs using pocket implant to improve manufacturability for 256-Mbit DRAM and beyond," IEEE IEDM Tech. Digest, p. 87, 1994.

[13] H. Chen, J. Zhiao, C. S. Teng, L. Moberly and R. Lahri, "Submicron large-angle-tilt-implanted drain technology for mixed signal applications,"IEEE IEDM Tech. Digest, p. 91, 1994.

[14] "MEDICI ver. 1999.2," Avant! Corporation, Fremont, CA, USA, 1999.

[15] Y. S. Pang and J. R. Brews, "Analytical subthreshold surface potential model for pocket n-MOSFETs," IEEE Transactions on Electron Devices, vol. 49, pp. 2209-2216, December 2002.

[16] X. Zhou, K. Y. Lim and D. Lim, "Physics-Based threshold voltage modeling with Reverse Short Channel Effect," Journal of Modeling and Simulation of Microsystems , Vol. 2, No. 1, pp. 51-56, 1999.

[17] X. Zhou, K. Y. Lim and D. Lim, "A general approach to compact threshold voltage formulation based on 2-D numerical simulation and experimental correlation for deep-submicron ULSI technology development," IEEE Trans. on Electron Devices, vol. 47, no. 1, pp. 214-221, Jan. 2000.

[18] M. H. Bhuyan, F. Ferdous and Q. D. M. Khosru, "A threshold voltage model for sub-100 nm pocket implanted n-MOSFET," in Proc. of the 4th International Conference on Electrical and Computer Engineering (ICECE 2006), Dhaka, December 19-21, 2006, pp. 522-525.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Linear Pocket Profile Based Threshold Voltage Model For Sub-100 nm n-MOSFET Incorporating Substrate and Drain Bias Effects

Muhibul Haque Bhuyan[1,2] and Quazi D. M. Khosru[2]

[1]Department of Electronics and Telecommunication Engineering
Daffodil International University, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
E-mail: muhibulhb@yahoo.com

Abstract - This paper presents a threshold voltage model of pocket implanted sub-100 nm n-MOSFETs incorporating the drain and substrate bias effects using two linear pocket profiles. Two linear equations are used to simulate the pocket profiles along the channel at the surface from the source and drain edges towards the center of the n-MOSFET. Then the effective doping concentration is derived and is used in the threshold voltage equation that is obtained by solving the Poisson's equation in the depletion region at the surface. Simulated threshold voltages for various gate lengths fit well with the experimental data already published in the literature. The result is compared with two other pocket profiles used to derive the threshold voltage models of n-MOSFETs. The comparison shows that the linear model has a simple compact form that can be utilized to study and characterize the pocket implanted advanced ULSI devices.

I. Introduction

The conventional threshold voltage model is derived for the homogeneous doping concentration [1]. As the channel length of MOSFETs is scaled down to deep-submicrometer or sub-100 nm regime, short-channel effects have been observed [2]. The short channel effects arise as results of two dimensional potential distribution and high electric fields in the channel region. Lateral channel engineering utilizing halo-pocket implant [3-7] surrounding drain and source regions is effective in suppressing short channel effects. An extension of the homogeneous model to the nonhomogeneous impurity pileup in the vertical direction has been reported previously [2-3, 8]. However, the reported model cannot be extended further to the pocket implantation, where inhomogeneity along the channel is the main cause for the reverse short channel effect (RSCE) [9]. A strong reverse short channel effect suppresses the short channel effect on threshold voltage of the MOSFET [10]. Threshold voltage model for pocket implanted MOSFETs for circuit simulation does not describe the sub-100 nm case [10]. In this work, we propose a threshold voltage model capable of describing the threshold voltage for the gate length down to 50 nm.

Advanced MOSFETs are non-uniformly doped as a result of complex process flow. Therefore, one of the key factors to model threshold voltage (V_{th}) accurately is to model its non-uniform doping profile. Here the lateral 1-D pocket profile across the channel has been transformed to an effective doping concentration expression that can be applied directly to the V_{th} expression incorporating V_{th} shift due to short channel effect to suppress this effect. Besides, drain and substrate bias effects have also been incorporated. Here a short channel threshold voltage equation is used for the case of pocket implanted n-MOSFET where exponential dependence of on channel length and a linear dependence on drain and substrate biases is observed [11] for various device and pocket profile parameters. Gaussian profile [12] and hyperbolic cosine profile [13] found in the literature for the threshold voltage model, are compared with the linear model. Simulation results using these two profiles along with the proposed linear profile show that the threshold voltage model is better for the linear profile. Besides, experimental data already published in the literature [4] fits well with our simulated data for various gate lengths. It proves the validity and usefulness of our proposed model of the threshold voltage for circuit simulation.

II. Threshold Voltage Model

The pocket implanted n-MOSFET structure shown in Fig. 1 is considered in this work and assumed co-ordinate system is shown at the right side of the structure.

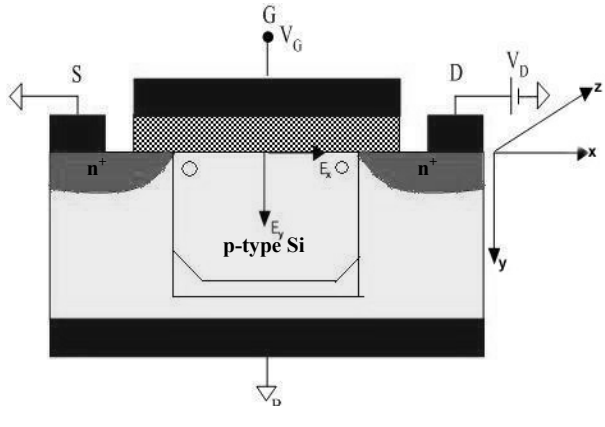


Fig. 1 Pocket implanted n-MOSFET Structure

All the device dimensions are measured from the oxide-silicon interface. In the structure, the junction depth (r_j) is 25 nm. The oxide thickness (t_{ox}) is 2.5 nm, and it is SiO_2 with fixed oxide charge density of 10^{11} cm^{-2}. Uniformly doped p-type Si substrate is used with doping concentration of N_{sub} = 4.2×10^{17} cm^{-3} with pocket implantation both at the source and drain side with peak pocket doping concentration of 1.5×10^{18} cm^{-3} and pocket length from 20 to 30 nm, and source or drain doping concentration of 9.0×10^{20} cm^{-3}.

The model of the conventional bulk n-MOSFET exhibits drastic reduction of the threshold voltage (ΔV_T) from the long channel value beyond 100 nm. This is known as short channel effect. A group of analytical models, known as "charge-sharing" models, are found in the literature. But their accuracy is limited [11]. In [11], a model is presented that solves the two-dimensional Poisson equation analytically, and predicts ΔV_T accurately as a function of drain bias (V_D), substrate bias (V_{BS}), channel length (L), oxide thickness (t_{ox}) and substrate concentration (N_{sub}). This model is then transformed to short channel n-MOSFET assuming the step doping profile along the vertical direction of the channel.

To preserve the long channel threshold voltage behavior for the short channel device, pocket implantation, which causes reverse short channel effect (RSCE), is done by adding donor atoms both from the source and drain edges. The peak pocket doping concentration (N_{pm}) gradually decreases towards the substrate level concentration, N_{sub} with a pocket length (L_p) from both the source and drain edges. The basis of the model of the pocket is to assume two linear doping profiles from both the source and drain edges across the channel as shown in Figs. 2-3. The pocket parameters, N_{pm} and L_p, play important role in determining the RSCE.

At the source side, the pocket profile is given as:

$$N_s(x) = -\frac{N_{pm} - N_{sub}}{L_p}x + N_{pm}$$

$$\text{or, } N_s(x) = N_{sub}\frac{x}{L_p} + N_{pm}\left(1 - \frac{1}{L_p}x\right) \qquad (1)$$

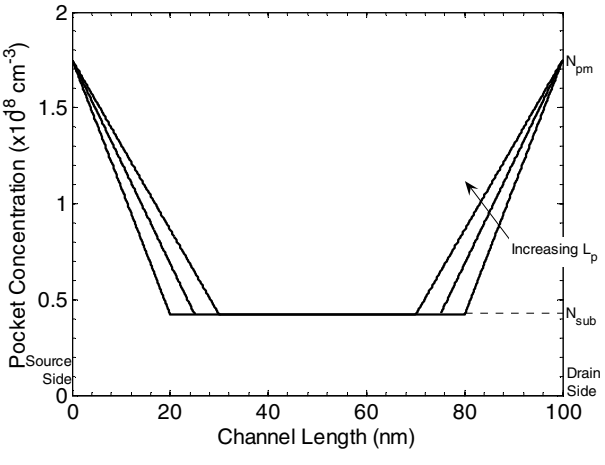


Fig. 2 Simulated pocket profiles at the surface for different pocket lengths, L_p = 20, 25 and 30 nm. Peak pocket concentration, $N_{pm} = 1.75\times10^{18}$ cm^{-3}.

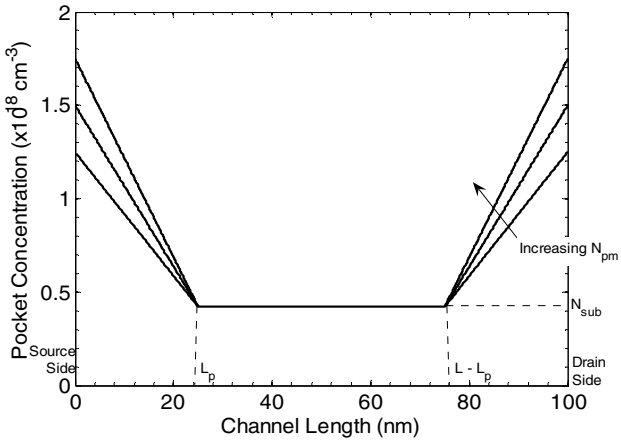


Fig. 3 Simulated pocket profiles at the surface for various peak pocket concentrations, $N_{pm} = 1.25\times10^{18}$, 1.5×10^{18} and 1.75×10^{18} cm^{-3}. Pocket length, L_p = 25 nm

At the drain side, the pocket profile is given as:

$$N_d(x) = \frac{N_{pm} - N_{sub}}{L_p}\left[x - \left(L - L_p\right)\right] + N_{sub}$$

$$\text{or, } N_d(x) = N_{sub}\left(\frac{L}{L_p} - \frac{1}{L_p}x\right) + N_{pm}\left(1 - \frac{L}{L_p} + \frac{1}{L_p}x\right) \qquad (2)$$

where x represents the distance across the channel. Since the pile-up profile is due to the direct pocket implant at the source and drain side, it is assumed symmetric at both sides.

With these two conceptual pocket profiles, the profiles are integrated mathematically along the channel length and then divided by it to derive an average effective concentration as in equation (1):

$$N_{eff} = \frac{1}{L}\int_0^L \left[N_s(x) + N_d(x) + N_{sub}\right]dx \qquad (3)$$

Putting the expressions of $N_s(x)$ and $N_d(x)$ from in equation (3) the effective doping concentration is obtained as follows:

$$N_{eff} = N_{sub}\left(1 - \frac{L_p}{L}\right) + \frac{N_{pm}L_p}{L} \qquad (4)$$

When $L_p << L$ for long channel device then pocket profile has no effect on V_{th}, But when L_p is comparable with L then pocket profile affects V_{th}.

In [14], the threshold voltage expression was obtained by solving the 1-D Poisson equation and then applying Gauss's law, but that model did not incorporate the effect of substrate and drain biases. The threshold voltage model derived in [11] incorporated the effects of substrate and drain biases on threshold voltage for vertically non-uniform doping profile. Based on that concept, we derived a threshold voltage model for our proposed pocket doping profile along the channel. This model incorporates the effective doping concentration of our linear pocket profiles given in equation (4) to derive the threshold voltage equations and hence we obtain the V_{th} expression given in equation (5).

$$V_{th} = V_{th,L} + \gamma_B \left[\frac{N_{sub}}{N_{eff}} (2\varphi_F) - V_{BS} \right]^{\frac{1}{2}} - \gamma_A \frac{N_{sub}}{N_{eff}} (2\varphi_F)^{\frac{1}{2}} - \frac{6t_{ox}}{d_1} \left[2(\varphi_{bi} - V_{BS}) + V_{DS} \right] \exp\left(-\frac{\pi L}{4d_1} \right) \quad (5)$$

where $V_{th,L}$ is the long channel threshold voltage, the second and the third parts include the threshold voltage due to both the effect of substrate bias and effective doping concentration, the fourth part incorporates the effect of drain and substrate biases as well as the short channel effect. The long channel threshold voltage, $V_{th,L}$ for the pocket implanted MOSFET is given by the following expression:

$$V_{th,L} = V_{FB} + 2\varphi_F + \gamma_A (2\varphi_F)^{\frac{1}{2}} \quad (6)$$

where V_{FB} is the flat band voltage. From simulation it is found -0.9316V. φ_{FA}, γ_A and γ_B are Fermi potential due to pocket implantation, threshold sensitivity due to back bias for effective doping concentration along the channel and body factor corresponding to bulk doping respectively and are given as follows:

$$\varphi_F = \frac{kT}{q} \ln \frac{N_{eff}}{n_i} \quad (7)$$

$$\gamma_A = \frac{\left(2q\varepsilon_{Si} N_{eff}\right)^{\frac{1}{2}}}{C_{ox}} \quad (8)$$

$$\gamma_B = \frac{\left(2q\varepsilon_{Si} N_{sub}\right)^{\frac{1}{2}}}{C_{ox}} \quad (9)$$

The depth (where band bending of $2\varphi_F$ occurs) of the pocket doping vertical to the channel and the built-in potential at the source or drain to channel junction are given by the following equations respectively

$$d_1 = \left(\frac{2\varepsilon_{Si}}{qN_{eff}} \right)^{\frac{1}{2}} (2\varphi_F)^{\frac{1}{2}} \quad (10)$$

$$\varphi_{bi} = \frac{kT}{q} \ln \frac{N_{sd} N_{eff}}{n_i^2} \quad (11)$$

where N_{sd} is the source or drain doping concentration and n_i is the intrinsic carrier concentration of Si.

III. Results and Discussion

The simulated V_{th} vs. L curve has been drawn for this new model is shown in Figs. 4-5 for different drain and substrate biases. It was shown in [14] that with the increasing pocket lengths and peak pocket concentration the peak of this curve increases and the onset of threshold voltage (V_{th}) roll-up happens at a longer channel length and also the onset of V_{th} roll-off happens at a shorter channel length. This result exhibits strong reverse short channel effect (RSCE) with the increased L_p and N_{pm}. If these are increased further by keeping the other parameters constant then V_{th} roll-off starts to vanish exhibiting only RSCE.

From Fig. 4, it is observed that as the drain bias increases, both RSCE and SCE occur at longer channel length due to the drain induced barrier lowering (DIBL). As channel length becomes shorter DIBL effect is more pronounced. Higher drain bias makes the threshold voltage negative. Since at shorter channel length, electric field is very high and it lowers the potential barrier that separates it from the adjacent diffused junction. Therefore, due to the presence of high channel doping even at negative gate voltage drain current starts to flow.

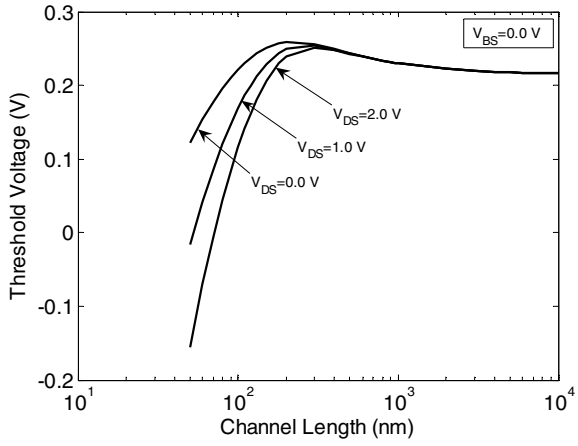


Fig. 4 Threshold voltage vs. gate length curves for various drain biases at zero substrate bias

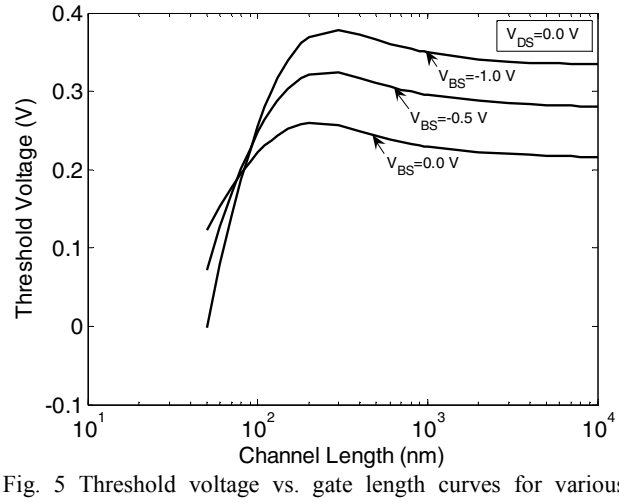


Fig. 5 Threshold voltage vs. gate length curves for various substrate biases at zero drain bias

From Fig. 5, it is found that as the substrate bias increase in the negative direction the threshold voltage increases. This is due to the increment of depletion charge under the gate. Also, with increasing magnitude of V_{BS}, RSCE occurs at longer channel length (L), V_{th} roll-off becomes steeper and V_{th}-L curve crosses the zero-substrate bias curve at a shorter channel length. This is because of the linear dependence of V_{DS} and V_{BS} on V_{th} and exponential dependence of L on V_{th}. As V_{BS} becomes more negative RSCE starts to diminish.

Figs. 6-7 show the variation of threshold voltage with the drain bias for different substrate biases of V_{BS} = 0.0 V and -1.0 V respectively with channel length as a parameter. It is observed that as the drain bias increases threshold voltage decreases. As the channel length shrinks, this effect becomes more prominent. For longer channel device, lateral electric field is less than the transverse electric field. Thus for low drain bias diffusion current dominates over drift current. Hence threshold voltage

does not deviate too much from low drain bias to high drain bias. But for shorter channel device, lateral electric field becomes stronger at low drain bias too. Hence drift current increases at low drain bias thereby larger threshold voltage deviation is observed from low to high drain bias.

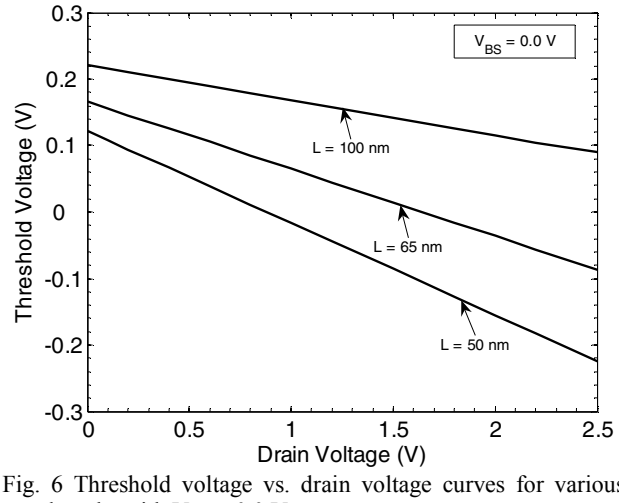


Fig. 6 Threshold voltage vs. drain voltage curves for various gate lengths with V_{BS} = 0.0 V

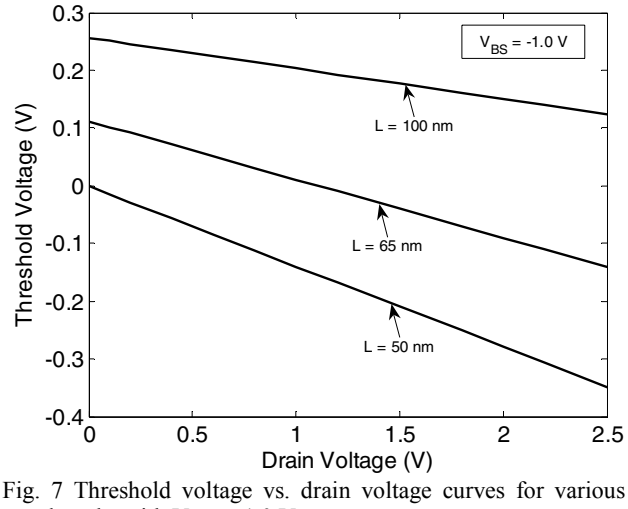


Fig. 7 Threshold voltage vs. drain voltage curves for various gate lengths with V_{BS} = -1.0 V

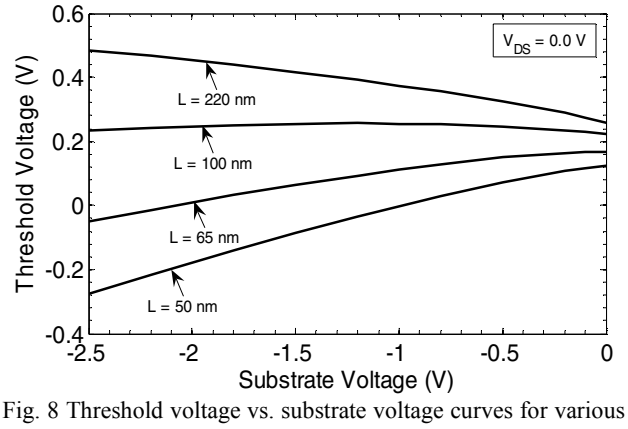


Fig. 8 Threshold voltage vs. substrate voltage curves for various gate lengths with V_{DS} = 0.0 V

Fig. 8 shows the variation of threshold voltage with the substrate bias for different channel lengths at zero drain bias. It is seen that as V_{BS} increases in the negative direction threshold voltage increases for long channel lengths and decreases for short channel lengths, i.e. in the sub-100 nm regime, and near 100 nm channel lengths threshold voltage is insensitive to substrate bias. This phenomenon happens due to the pocket implantation. When substrate bias increases in the negative direction in the long channel device, depletion layer charge increases due to the increase of depletion layer width that causes the threshold voltage to increase. But in the short channel device, threshold voltage decreases due to the increment of minority carriers at the surface when the substrate bias increases in the negative direction.

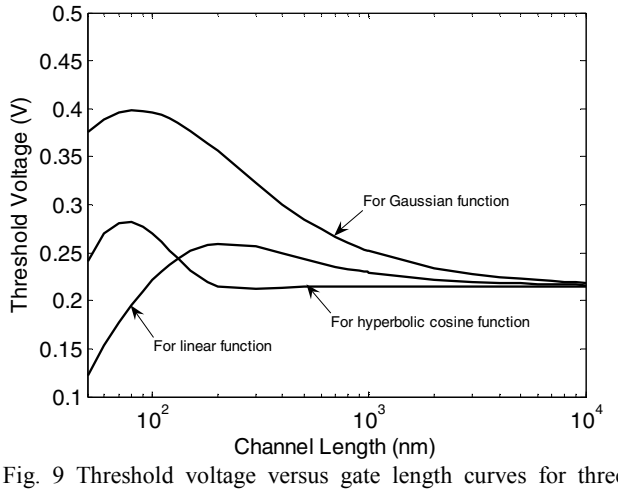


Fig. 9 Threshold voltage versus gate length curves for three different pocket based on linear, Gaussian and hyperbolic cosine functions for $N_{pm} = 1.75\times10^{18}$ cm^{-3}, L_p = 25 nm and $N_{sub} = 4.2\times10^{17}$ cm^{-3}

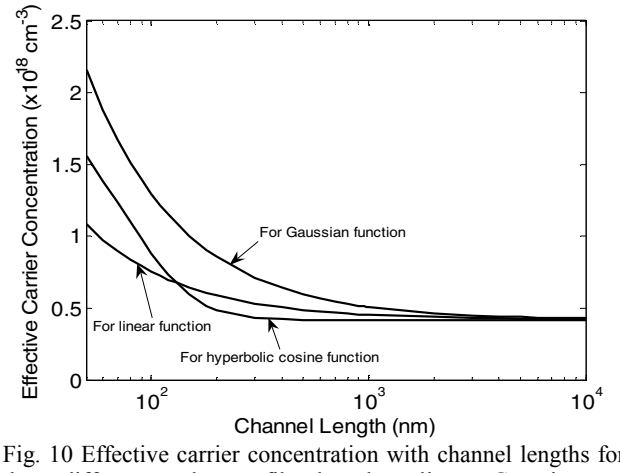


Fig. 10 Effective carrier concentration with channel lengths for three different pocket profiles based on linear, Gaussian and hyperbolic cosine functions for $N_{pm} = 1.75\times10^{18}$ cm^{-3}, L_p = 25 nm and $N_{sub} = 4.2\times10^{17}$ cm^{-3}

Figure 9 shows that our model based on linear pocket profile exhibits better result of suppressing short channel effect in comparison with the other models for the pocket profiles based on Gaussian [12] and hyperbolic cosine functions [13]. This can be explained using the Fig. 10 where we show the effective carrier concentration variation with the channel length. Here we see that the effective carrier concentration increases very smoothly the linear pocket profile as the channel length shrinks. But

in case of hyperbolic cosine function, it does not start to increase until 200 nm. But we know that the SCE starts before 0.1 μm. Therefore, for hyperbolic cosine model, at first SCE starts and then again RSCE becomes stronger below 100 nm.

On the other hand, in case of Gaussian function, the effective carrier concentration increases more rapidly than that in the linear model. Therefore, in Fig. 9, we observe that RSCE is stronger. Thus in this case, threshold voltage is increasing until 40 nm. But in sub-100 nm regime our purpose is to suppress the SCE only by the RSCE by implanting the pockets. Besides, simulation time taken to calculate threshold voltage by using our model for pocket profile is less than that by using Gaussian function and hyperbolic cosine function.

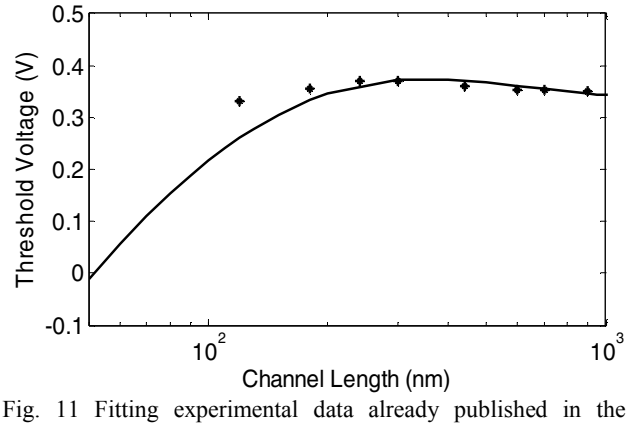


Fig. 11 Fitting experimental data already published in the literature with the simulated results of our proposed model

In Fig. 11, we tried to fit experimental data obtained from [4] with our simulated data for the device parameters given in [4], such as, substrate concentration, N_{sub} = 1.0×10^{17} cm^{-3}, peak pocket concentration, N_{pm} = 5.5×10^{17} cm^{-3}, pocket length, L_p = 60×10^{-7} cm either from source or drain side, oxide thickness, d=6×10^{-7} cm, junction depth, r_j = 80×10^{-7} cm, substrate bias, V_{BS} = 0.0 V and drain bias, V_{DS} = 0.05V. Flat band voltage obtained by simulation is V_{FB} = -0.9 V. From Fig. 11, it is clear that our simulated data almost agrees well with the experimental data. In the short channel device, our model shows better results. By changing the process conditions, it is possible to adjust the experimental results with the simulated data. Because we only adjusted the flat band voltage to fit the experimental data.

IV. Conclusion

A threshold voltage model for ultra thin oxide and sub-100 nm pocket implanted n-MOSFET has been developed incorporating the substrate and drain bias dependence. The well-known reverse short channel effect has been observed through the proposed model. The model is developed assuming two linear pocket profiles along the channel at the surface of the MOS device from the source and drain edges. The proposed model along with the other two models of threshold voltage has been simulated using the same values for the different parameters and compared for V_{th} vs. L_g characteristics. Experimental results already published in the literature have also been compared with our simulated results and fit very well. It is found that our model efficiently determines the threshold voltages of scaled n-MOSFETs having channel lengths in sub-100 nm regime and shows better performance than the other two models. Therefore, this model is very useful for circuit simulation.

References

[1] S. M. Sze, "Physics of Semiconductor Devices," 2nd Edition, John Wiley & Sons, New York, ch. 8, 1981.

[2] M. Miura-Mattausch, M. Suetake, H. J. Mattausch, S. Kumashiro, N. Shigyo, S. Oganaka and N. Nakayama, "Physical modeling of the reverse short channel effect for circuit simulation," IEEE Trans. on Electron Devices, vol. 48, pp. 2449-2452, Oct. 2001.

[3] K. Y. Lim and X. Zhou, "Modeling of Threshold Voltage with Non-uniform Substrate Doping," in Proc. of the IEEE Int'l Conf. on Semiconductor Electronics (ICSE'98), Malaysia, pp. 27-31, 1998.

[4] B. Yu, C. H. Wann, E. D. Nowak, K. Noda and C. Hu, "Short Channel Effect improved by lateral channel engineering in deep-submicrometer MOSFETs," IEEE Transactions on Electron Devices, vol. 44, pp. 627-633, Apr. 1997.

[5] B. Yu, H. Wang, O. Millic, Q. Xiang, W. Wang, J. X. An and M. R. Lin, "50 nm gate length CMOS transistor with super-halo: Design, process and reliability," IEDM Tech. Digest, pp. 653-656, 1999.

[6] K. M. Cao, W. Liu, X. Jin, K. Vasant, K. Green, J. Krick, T. Vrotsos and C. Hu, "Modeling of pocket implanted MOSFETs for anomalous analog behavior," IEDM Tech. Digest, pp. 171-174, 1999.

[7] Y. S. Pang and J. R. Brews, "Models for subthreshold and above subthreshold currents in 0.1 μm pocket n-MOSFETs for low voltage applications," IEEE Transactions on Electron Devices, vol. 49, pp. 832-839, May 2002.

[8] Y. Cheng, T. Sugii, K. Chen and C. Hu, "Modeling of small size MOSFETs with reverse short channel and narrow width effects for circuit simulation," Solid State Electronics, vol. 41, pp. 1227-1231, 1997.

[9] M. K. Khanna, M. C. Thomas, R. S. Gupta and S. Haldar, "An analytical model for anomalous threshold voltage behavior of short channel MOSFETs," Solid State Electronics, vol. 41, pp. 1386-1388, 1997.

[10] Y. Taur and E. J. Nowak, "CMOS devices below 0.1μm; How high will go?," IEDM Technical Digest, pp. 215-218, 1997.

[11] K. N. Ratnakumar and J. D. Meindl, "Short-Channel MOST Threshold Voltage Model," IEEE Journal of Solid State Circuits, vol. 17, pp. 937-948, Oct. 1982.

[12] X. Zhou, K. Y. Lim and D. Lim, "Physics-Based threshold voltage modeling with Reverse Short Channel Effect," Journal of Modeling and Simulation of Microsystems , Vol. 2, No. 1, pp. 51-56, 1999.

[13] X. Zhou, K. Y. Lim and D. Lim, "A general approach to compact threshold voltage formulation based on 2-D numerical simulation and experimental correlation for deep-submicron ULSI technology development," IEEE Trans. on Electron Devices, vol. 47, no. 1, pp. 214-221, Jan. 2000.

[14] M. H. Bhuyan, F. Ferdous and Q. D. M. Khosru, "A threshold voltage model for sub-100 nm pocket implanted NMOSFET," in Proc. of the 4th International Conference on Electrical and Computer Engineering (ICECE 2006), Dhaka, December 19-21, 2006, pp. 522-525.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

InN-based Dual Channel High Electron Mobility Transistor

Md. Tanvir Hasan[1], Md. Monibor Rahman[2], A.N.M. Shamsuzzaman[2], Md. Sherajul Islam[2] and Ashraful G. Bhuiyan[2]

[1]Department of Electronics & Telecommunication Engineering, Faculty of Science & Information Technology, Daffodil International University (DIU), Dhaka-1207, Bangladesh
[2] Department of Electrical & Electronic Engineering, Khulna University of Engineering and Technology (KUET), Khulna-920300, Bangladesh
E-mail: tan_vir_bd@yahoo.com

Abstract - This paper describes the theoretical design and predicts the performances using Monte Carlo simulation of InN-based dual channel high electron mobility transistor (HEMT). A high sheet carrier concentration and strong electron confinement at specific interfaces of the InN-based heterostructures are predicted as a consequence of piezoelectric and spontaneous polarization effects. The calculated sheet carrier concentration reaches as high as 1.64×10^{14} cm^{-2} for dual channel HEMT. The sheet carriers generated in InN-based double channel are found to be higher than the reported values for the conventional single channel HEMTs. The 2DEGs mobility is found to be 11.76×10^{4} cm^{2}V^{-1}sce^{-1} at sheet carrier concentration, n_s = 1.2×10^{13} cm^{-2} for 100 K. At room temperature, the drain current of the proposed InN-based dual channel HEMTs is 1.2 Amm^{-1} at gate voltage V_G = 2 V. The drain current capabilities are found to be substantially superior to the conventional single channel HEMTs.

I. Introduction

Indium nitride (InN) semiconductor has been regarded as a very interesting and highly promising material system for both optical and microwave applications. In recent years, there has been much interest in the development of high-performance high-electron mobility transistors (HEMTs) based on InN [1, 2]. We have studied the effects of polarization on carrier confinement and localization of two-dimensional electron gas (2DEG) in InN-based HEMTs [3]. A high sheet carrier concentration and strong electron confinement at specific interfaces of the InN-based heterostructures were predicted as a consequence of piezoelectric and spontaneous polarization effects. However, the maximum concentration of 2DEG is limited for conventional single channel HEMT. Ideally, one would like to have high 2DEG concentration to reduce source resistance and to increase both transconductance and current drive of the HEMT device.

To meet the future world demand device performances must be improved. Improving device performances requires a details understanding of how the mobility and the density of the 2DEG can be increased in the material system. Desirable designs of heterostructures should be investigated for superior transport properties. Channel carrier confinement is the key to achieving high performance of nitride-based heterostructures. The electron confinement can significantly be improved by increasing the number of channel. Dual channel HEMT device uses a heterojunction on both sides of the undoped conductive channel. The electron confinement in the channel is expected to be stronger in the double heterostructure than in the conventional single heterostructure due to the enhanced polarization-induced electric field [4]. Chen et al. have proposed and demonstrated an AlGaN/InGaN/GaN double-heterostructure transistor (DHFET) where the electron confinement is significantly improved due to enhanced potential barriers at the AlGaN/InGaN and InGaN/GaN heterointerfaces [5]. Current collapse free performance of the DHFET was demonstrated. Therefore, attention should be given in the design of InN-based dual channel HEMT to meet the future demand. However, there is very little theoretical works on the InN-based heterostructures. Most of the works carried out on conventional GaAs, InGaAs GaN-based heterostructures [5-8].

In this work, we have designed theoretically InN-based dual channel HEMT and evaluated its performances using Monte Carlo simulation. These include the theoretical analysis and calculation of polarization, polarization induced sheet charge and sheet carrier concentration at interfaces in InN/InGa(Al)N/InN heterostructures. A detailed analysis on carrier confinement and localization of 2DEGs in InN-based heterostructures has been made. The performances such as the 2DEGs mobility and I-V characteristics have been evaluated.

II. Device Structure

The proposed schematic device model is shown in Fig. 1. For dual channel HEMT structures, each undoped InN channel layer is sandwiched between two n-InGaN layers. Intentional stress will be produced at the interfaces due to the lattice mismatch between $In_xGa_{1-x}N$/InN and InN/ $In_xGa_{1-x}N$ heterostructures. Electrons will be defused to the lower energy InN layer where they are confined due to the energy barrier at the heterointerface. The technique of modulation doping is a perfect means of introducing

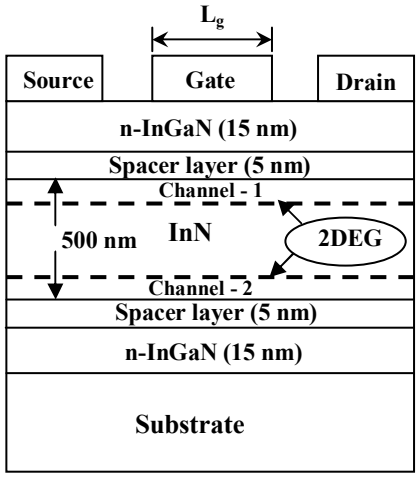


Fig. 1 Proposed schematic diagram of InN-based dual channel HEMT.

electrons into the InN layer without the adverse effects of donors. The region of the $In_xGa_{1-x}N$ depleted of electrons forms a positive space charge region, which is balanced by the accumulated electrons at the InN heterointerface. With increasing the spacer layer thickness enhances the electron mobility by reducing columbic scattering; however it reduces the 2DEG carrier density as well, which is not desirable because of the reduction in electron transfer. Therefore a compromise should be made between the donor density in n-$In_xGa_{1-x}N$, conduction band edge discontinuity (ΔE_c), which is controlled by the In content in the n-$In_xGa_{1-x}N$, and the thickness of the undoped InGaN spacer layer to maximize both saturation velocity and the electron density of 2DEG in the HEMT undoped InN channel.

III. Polarization and Carrier confinement

Polarization is the separation of positive and negative charges. For a material to exhibit an anisotropic property such as piezoelectricity, its crystal structure must have no centre of symmetry. When the structures are unstrained, there exists only the spontaneous polarization due to inherent properties of that material. In case of strained structures both spontaneous and piezoelectric polarization are present. The spontaneous polarization for GaN, InN and AlN was found to be negative [9]. In the absence of external electric fields, the total polarization of $In_xGa_{1-x}N$/InN HEMT is the sum of the spontaneous polarization (due to the cation and anion positions present in the lattice [10]) and the piezoelectric polarization (due to lattice mismatch between $In_xGa_{1-x}N$ and InN layers [10]). The piezoelectric polarization is negative for tensile and positive for compressive strained barriers, respectively. The values of spontaneous polarization are found to be - 0.066 C/m^2 and - 0.1C/m^2 for single and dual channel, respectively at a composition $x = 0$. Similarly for the other compositions $x = 0.2$, 0.4 the values are - 0.062 C/m^2 and - 0.06 C/m^2 for single channel and -0.092 C/m^2 and - 0.088 C/m^2, for dual channel, respectively, which is shown in Fig. 2 (a). The value of piezoelectric polarization of InN-based dual channel reaches to -0.263 C/m^2, while for single channel it is found to be -0.113 C/m^2, which is shown in Fig. 2 (b). For large lattice mismatch the value of piezoelectric polarization is more. The values of spontaneous and piezoelectric polarizations for dual channel are higher than conventional single channel HEMTs.

The presence of a large polarization field at the $In_xGa_{1-x}N$/InN heterointerface is responsible for higher values of sheet carrier concentration, n_s. Figure 3 shows the dependency of sheet carrier concentration with In composition of n-$In_xGa_{1-x}N$ barrier for both single and dual channel HEMTs. The values of sheet carrier concentration are found to be decreased from 7.06×10^{13} to 2.293×10^{13} cm^{-2} for the single channel, while 1.64×10^{14} to 1.69×10^{13} cm^{-2} for dual channel, with the increase of In composition from $x = 0$ to 1. Hence there will be increased current carrying capability in dual channel HEMT, which is suitable for high performance microwave devices. The calculated values of sheet carrier concentration are in good agreement with the theoretical and experimental values reported for the InN-based and conventional GaN-based heterostructures [2, 10, 11]. The sheet carriers generated in InN-based double channel are

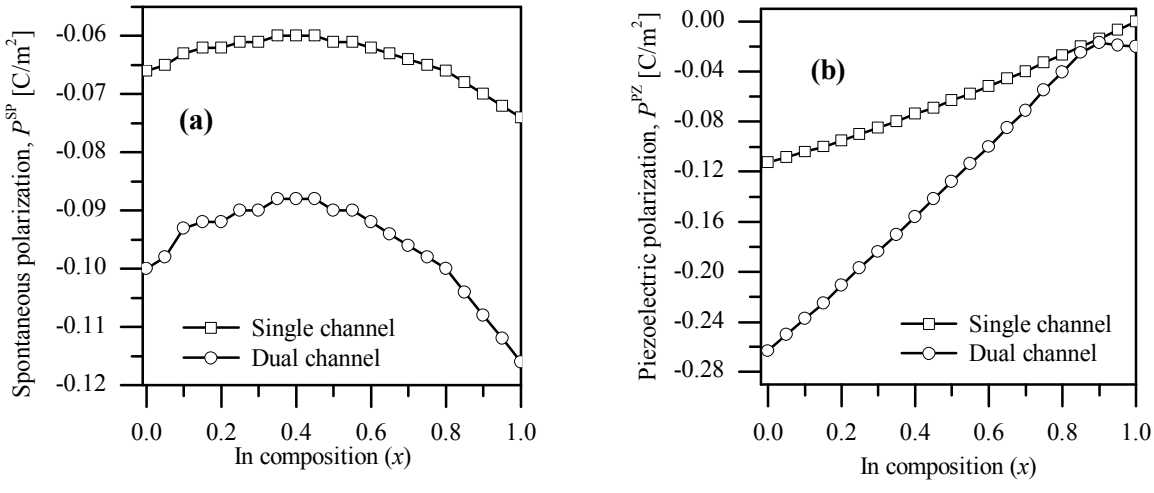


Fig. 2 Variation of spontaneous and piezoelectric polarization as a function of alloy composition(x) for single and Dual channel HEMTs.

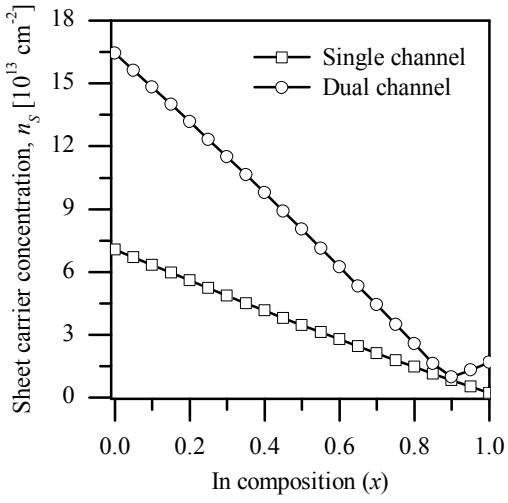


Fig. 3 The dependency of sheet carrier concentration with composition of single and dual channel HEMTs.

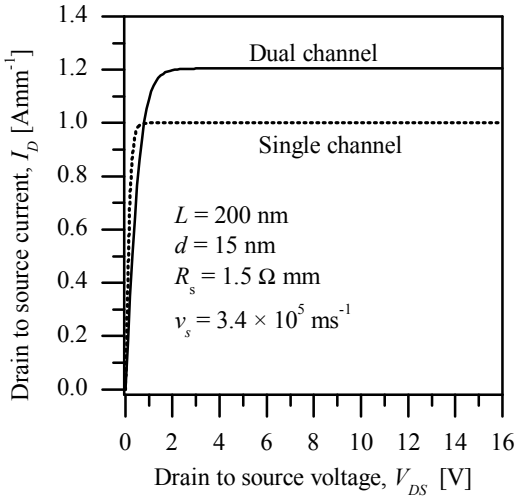


Fig. 5 Output I-V characteristics of the proposed InN-based single and dual channel of HEMTs at V_G = 2V.

found to be higher than the reported values for the conventional single channel HEMTs [6-8].

IV. Transport properties of 2DEG

The density of 2DEGs and mobility in heterostructures are the key parameters related to transport properties. There are several scattering mechanisms which play a role in limiting the 2DEGs mobility in InN-based HEMTs. To calculate the 2DEGs mobility in InN-based HEMTs, the different scattering mechanisms such as dislocations scattering due to the large lattice mismatch, impurity scattering by remote donors and due to interface charge, interface roughness in InGa(Al)N/InN heterointerfaces, alloy disorder scattering due to penetration of the 2DEG wavefunction into the barrier, and phonons scattering are consider using Monte Carlo simulation. Figure 4 shows the dependency of 2DEGs mobility for both single and dual channel HEMTs on sheet carrier concentration at 100 K. The 2DEGs mobility is found to be 8×10^4 cm^{2}V^{-1}sce^{-1} for single channel, while 11.76×10^4 cm^{2}V^{-1}sce^{-1} for dual channel at sheet carrier concentration $n_s = 1.2\times10^{13}$ cm^{-2} for 100 K. The 2DEGs mobility for dual channel is higher than conventional single channel HEMTs [7, 8].

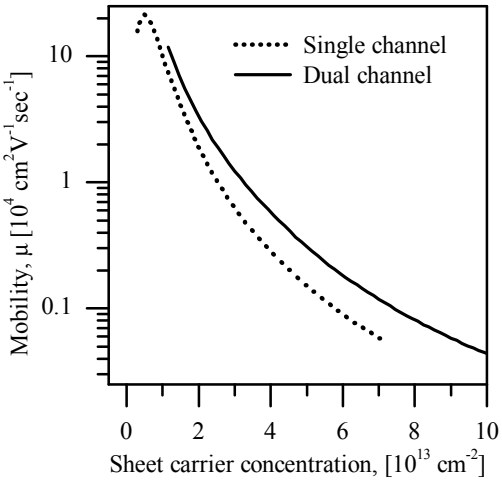


Fig. 4 The dependency of 2DEGs of mobility for both single and dual channel HEMTs on sheet carrier concentration.

Two-dimensional electron gas (2DEG) transport properties and device characteristics in InN-based HEMTs have investigated keeping high-power application in mind. A simple analytic model is developed for the output I-V characteristics of the proposed InN-based single and dual channel HEMTs. The effects relating to transistor self-heating, which can substantially reduce HEMT current as well as electron non-stationary dynamics, are not considered in our model. Figure 5 shows predicted output I-V characteristics of the proposed InN-based HEMTs with single and dual channel. The output I-V characteristics are in good agreement with the output I-V characteristics of the conventional HEMTs. At room temperature, the drain currents of the proposed InN-based HEMTs are 1 and 1.2 Amm^{-1} for single and dual channel, respectively at gate voltage V_G = 2 V. The drain current of dual channel HEMT is higher than the conventional single channel HEMT. The drain current capabilities are found to be substantially superior to the conventional GaN-based HEMTs capabilities.

IV. Conclusion

We have theoretically designed InN-based dual channel HEMT and evaluated its performance using Monte Carlo Simulation. The effect of piezoelectric and spontaneous polarization for sheet carrier confinement and formation of 2DEG have been investigated. The values of spontaneous polarization are found to be - 0.066 C/m^2 and - 0.1C/m^2 for single and dual channel, respectively at a composition x = 0. The value of piezoelectric polarization of InN-based dual channel HEMT reaches to - 0.263 C/m^2 and for single channel HEMTs this value is found to be - 0.113 C/m^2. The values of spontaneous and piezoelectric polarizations for dual channel are higher than conventional single channel HEMTs. The sheet carrier concentrations are found to be decreased from 7.06×10^{13} to 2.293×10^{12} cm^{-2} for the single channel, while 1.64×10^{14}

to 1.69×10^{13} cm^{-2} for dual channel, with the increase of In composition from $x = 0$ to 1. The sheet carriers generated in InN-based double channel are found to be higher than the reported values for the conventional single channel HEMTs. The 2DEGs mobility is found to be 8×10^{4} $cm^{2}V^{-1}sce^{-1}$ for single channel, while 11.76×10^{4} $cm^{2}V^{-1}sce^{-1}$ for dual channel at sheet carrier concentration $n_s = 1.2\times10^{13}$ cm^{-2} for 100 K. The 2DEGs mobility for dual channel is higher than conventional single channel HEMTs. At room temperature, the drain currents of the proposed InN-based HEMTs are 1 and 1.2 Amm^{-1} for single and dual channel, respectively at gate voltage $V_G = 2$ V. The drain current of dual channel HEMT is higher than the conventional single channel HEMTs. The drain current capabilities are found to be substantially superior to the conventional GaN-based HEMTs. The above calculated results indicate that the proposed InN-based dual channel HEMT is very promising for the fabrication of high performance high speed devices.

References

[1] Ashraful G. Bhuiyan, A. Hashimoto, A. Yamamoto, "Indium nitride (InN): a review on growth, characterization, and properties", J. Appl. Phys. Vol. 94, pp. 2779–808, September 2003.

[2] Y.C. Kong, Y.D. Zheng, C.H. Zhou, Y.Z. Deng, B. Shen, S.L. Gu, R. Zhang, P. Han, R.L. Jiang, Y. Shi, "A novel $In_xGa_{1-x}N$/InN heterostructure field-effect transistor with extremely high two-dimensional electron-gas sheet density", Solid-State Electronics, Vol. 49, pp.199–203, 2005.

[3] Md. Tanvir Hasan, Ashraful G. Bhuiyan, Akio Yamamoto, "Two dimensional electron gas in InN-based heterostructures: Effects of spontaneous and piezoelectric polarization", Solid-State Electronics, Vol. 52, No. 1, pp. 134-39, January 2008.

[4] R. Lee Ross, S. P. Svensson, and P. Lugli, "*Pseudomorphic HEMT Technology and applications*," NATO ASI Series, Kluwer Academic Publishers, 1996.

[5] C. Q. Chen, J. P. Zhang, V. Adivarahan, A. Koudymov, H. Fatima, G. Simin, J. Yang, and M. Asif Khan, "AlGaN/GaN/AlGaN double heterostructure for high-power III-N field-effect transistors", Appl. Phys. Lett., Vol. 82, No. 25, pp. 4593-95, June 2003.

[6] Jia-Chuan Lin, Yu-Chieh Chen, Wei-Chih Tsai, Po-Yu Yang, "Structure design criteria of dual-channel high mobility electron transistors", Solid-State Electronics, Vol. 51, no.1, pp. 64–68, January 2007.

[7] N. H. Sheng, C. P. Lee, R. T. Chen, D. L. Miller, and S. J. Lee, "Multiple-channel GaAs/AlGaAs high electron mobility transistors," IEEE Electron Device Letters, Vol. EDL-6, No. 6, pp. 307-310, June 1985.

[8] R. Gupta, M. E. Nokali, "A model for dual-channel high electron mobility transistors," Solid-State Electronics, vol. 38, No. 1, pp. 51-57, January 1995.

[9] F. Bernardini, V. Fiorentini, D. Vanderbilt, "Spontaneous polarization and piezoelectric constants of III–V nitrides", Phys. Rev. B, Vol. 56, No. 16, pp. R10024–7, 1997.

[10] O. Ambacher, J. Majewski, C. Miskys, A. Link, M. Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, W. Schaff, L. F. Eastman, "Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures", J. Phys: Condens Matter, Vol. 14, pp. 3399-434, March 2002.

[11] H. Lu, W. J. Schaff, L. F. Eastman, et al., "Surface charge accumulation of InN films by molecular-beam epitaxy", Appl. Phys. Lett., Vol. 82, pp. 1736–38, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

C-V Characteristics of n-channel Double Gate MOS Structures Incorporating the Effect of Interface States

A. Alam, S. Ahmed, M. K. Alam and Quazi D. M. Khosru

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
E-mail: ahsanulalam@eee.buet.ac.bd

Abstract - The Capacitance-Voltage (CV) characteristics of n-channel Double Gate (DG) MOS structures are investigated. An accurate and efficient fully coupled 1-D Schrödinger-Poisson self-consistent solver has been used. The numerical solver employs finite element method to calculate different electrostatics of n-channel DGMOS structures. The CV characteristics are modeled by taking the effect of interface trapped charges into account for varying densities of interface states. Both high frequency (HF) and low frequency (LF) operations are considered.

I. Introduction

As CMOS scaling is approaching its limit due to processing as well as fundamental considerations; double-gate (DG) MOSFET is becoming an intense subject of VLSI research. In theory, DG MOSFETs can be scaled to the shortest channel length possible for a given oxide thickness [1]. Among the advantages advocated for double-gate MOSFETs are: ideal 60mv/decade sub-threshold slope, volume inversion [2], setting of threshold voltage by the gate work function thus avoiding dopants and associated number fluctuation effects etc. There is still lack of numerical or experimental data that can reveal the quantum mechanical effects on CV characteristics of double gate MOSFETs. Recently, the effect of interface states on MOS device capacitance has become of great interest. The alteration of device C-V characteristics due to charges trapped in the interface states degrades the device reliability. So, in order to predict the CV characteristics efficiently, the effect of interface states should be taken into account. Present works on the effect of interface states [3], [4] cover only conventional single gate MOSFETs. So, a quantum mechanical analysis of double gate MOS capacitance incorporating the effect of interface states is in need.

In this work, coupled Schrödinger and Poisson equations have been solved self consistently [5] considering open boundary condition in order to calculate quantum mechanical charge distribution in MOS devices incorporating wave function penetration effect within the oxide layer. The solver developed for this purpose employs Finite element method using FEMLAB [6]. The CV characteristics are calculated using the electrostatics of the device revealed by the solver. The effect of interface states is then incorporated in the calculated C-V profile as alteration of device threshold voltage and as alteration of total charge.

II. Theory

A. Self-Consistent Schrödinger–Poisson's Solver

The solver uses classical PDEs in 'Multiphysics' mode to solve the Poisson's and Schrödinger's equations. Poisson's equation in coefficient form is given in FEMLAB as:

$$-\nabla.(c\nabla u) = f \tag{1}$$

The equation is equivalent to the equation for MOS electrostatic potential profile,

$$-\varepsilon_0\varepsilon\frac{d^2v(z)}{dz^2} = q[p(z)-n(z)+N_D-N_A] \tag{2}$$

Where $n(z)$, $p(z)$ are the electron, hole concentration and N_D, N_A are the ionized donor, acceptor concentration respectively. ε is the relative dielectric constant of the material and ε_0 is the permittivity of free space. The electron concentration $n(z)$ is obtained for n-MOS structure according to the following expression:

$$n(z) = \sum_{ij} N_{ij}\left|\psi_{ij}(z)\right|^2 \tag{3}$$

$$N_{ij} = \frac{n_{vi}m_{di}KT}{\pi\hbar^2}\ln[1+\exp(\frac{E_F-E_{ij}}{KT})] \tag{4}$$

Where N_{ij} is the carrier concentration in the j th subband of the i th valley, n_{vi} and m_{di} are the i th valley degeneracy and the i th density-of-states effective mass in Si. N_{inv} is the total inversion carrier concentration and E_F is the Fermi energy. E_{ij} and ψ_{ij}

are the eigenvalue and the eigenfunction of an electron in the j th energy level of the i th valley, which are obtained as a solution of the one dimensional Schrödinger equation.

Schrödinger equation in coefficient form is defined in FEMLAB as:

$$-\nabla.(c\nabla u)+au=\lambda u \tag{5}$$

The equation is equivalent to the one-dimensional effective mass Schrödinger equation,

$$[-\frac{\hbar^2}{2m*}\frac{d^2}{dz^2}+v(z)]\psi_{ij}(z)=E_{ij}\psi_{ij}(z) \tag{6}$$

At first we obtain the trial potential by solving (1) using the semi classical approximation [7] assuming zero mobile charge density i.e. $n(z)=0$ in (2) with appropriate boundary condition at each interface through linear PDE solver. A discontinuous electric field boundary condition can be easily set with Neumann boundary condition. Then the charge density profile $n(z)$ is determined from (3), (4) by solving (5) using eigen-value solver with Neumann boundary condition This charge density profile is added to the source term of Poisson's equation and then (1) and (5) are solved iteratively until the Fermi level & first eigen energies for each valley congregate to the given convergence criteria. After the numerical investigation is complete, the CV characteristics can be found from the device electrostatics.

B. Effect of interface states on CV profile

Interface states are electrically active defects with an energy distribution throughout the Si band gap, designated as D_{it} (cm^{-2} eV^{-1}), Q_{it} (C cm^{-2}), and N_{it} (cm^{-2}). They act as generation/recombination centres and contribute to threshold voltage shifts [5], given by

$$\Delta V_T=-\frac{\Delta Q_{it}(\phi_S)}{C_{ox}}, \tag{7}$$

Where Φ_s is the surface potential. The surface potential dependence of the occupancy of interface traps is illustrated in Fig. 1.

Interface traps at the SiO_2 / Si interface are acceptor-like in the upper half and donor-like in the lower half of the band gap [8]. This is in contrast to doping atoms, which are donors in the upper half and acceptors in the lower half of the band gap.

Hence as shown in Fig.1, the n channel has positive interface trapped charges at flatband, and negative interface trapped charges at inversion.

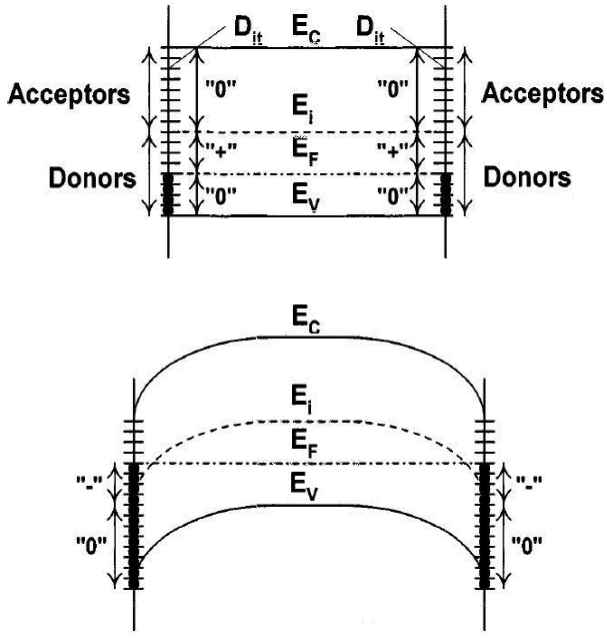


Fig.1. p substrate of DG nMOS at flatband and at inversion.

Hence as shown in Fig.1, the n channel has positive interface trapped charges at flatband, and negative interface trapped charges at inversion.

In HFCV measurement, HF AC is superimposed on a DC bias that is changed from accumulation to inversion. Interface trapped charges cannot respond to the HF AC signal. But they can respond to the DC bias. Therefore, the measured capacitance does not depend on the amount of interface traps present. But the threshold voltage is modified. In LFCV measurement, the interface traps respond to the LF AC signal along with the DC bias. Thus the measured capacitance also changes in addition to the alteration of V_G as capacitance is given by

$$C_g=\frac{\Delta Q_{total}}{\Delta V_G} \tag{8}$$

Where

$$Q_{total}=Q_{depletion}+Q_{inversion}+Q_{it} \tag{9}$$

III. Results and Discussion

In this section, we represent the outcome of our work. Fig.2 shows the CV characteristic of an n-channel Double Gate MOS including the quantum mechanical effects but neglecting the effect of interface states. The device under consideration is built on (100) Si, with an undoped substrate, oxide thickness, T_{ox}=1.5nm and silicon body thickness T_{body}=25nm.

The effects of interface states is modeled as a corresponding change in device threshold voltage and

also as a corresponding change in device carrier concentration and is incorporated in the C-V profile.

The effect of interface states is incorporated in the C-V characteristic shown in Fig. 2, considering both HF and LF operation. Fig. 3 and Fig. 4 portrays the HF and LF C-V characteristics respectively after incorporating the effect of interface trapped charges for varying densities of interface states.

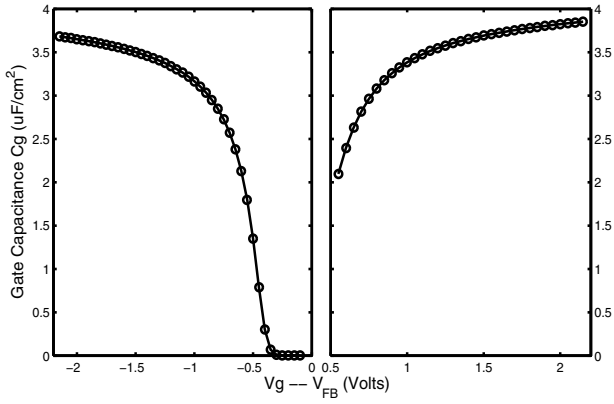


Fig.2. CV characteristic of an n-channel DGMOS (100 Si)

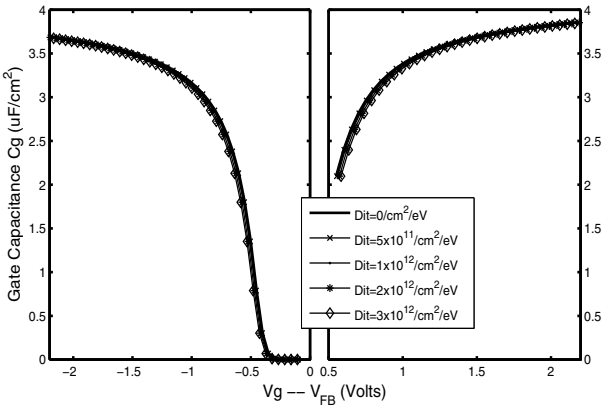


Fig.3. HFCV characteristic of an n-channel DGMOS (100 Si) under varying densities of interface states

It can be seen from Fig.3 and Fig.4 that at low frequency (LF), the alteration in the CV profile is much more pronounced than at high frequency (HF) operation. This is because the interface trapped charges respond to the DC bias only at HF operation whereas it responds to both the AC and DC signal during LF operation. In the case of HFCV, the curve shifts rightward as the effect of negative interface trapped charges on device threshold voltage corresponds to an increment in the corresponding gate voltage in the CV profile. In the LFCV profile, the curves shift rightward due to the same reason but in addition to that, the corresponding capacitance also increases since, the carrier density of the device increases due to the interface trapped charges.

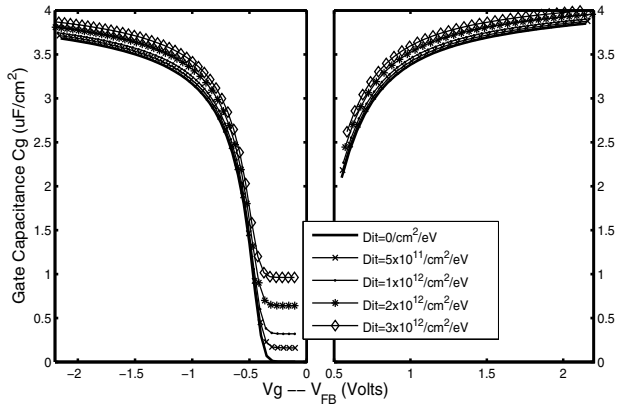


Fig.4. LFCV characteristic of an n-channel DGMOS (100 Si) under varying densities of interface states

Due to the alteration in threshold voltage, the device capacitance corresponding to applied voltage decreases which can be seen from Fig.3 at HF operation. But the increment in carrier concentration plays a greater role during LF operation and as a result the device capacitance increases in the LFCV characteristics as demonstrated in Fig.4.

Fig.5. shows the change in gate capacitance due to the effect of interface states for varying interface states densities at high frequency operation. It can be observed here that during strong inversion and accumulation, the capacitance isn't affected much by the presence of interface states. The alteration in device capacitance is most prominent during weak inversion and accumulation since the rate of change in carrier density with respect to voltage is comparatively much higher in this region.

Fig.6. shows the change in gate capacitance due to the effect of interface states for varying interface states densities at low frequency operation. It can be observed here that the presence of interface states affects the capacitance severely at low frequencies. But, the alteration in device capacitance is most prominent during depletion since the change in carrier density due to interface trapped charges compared to the already present carriers is the most in this region.

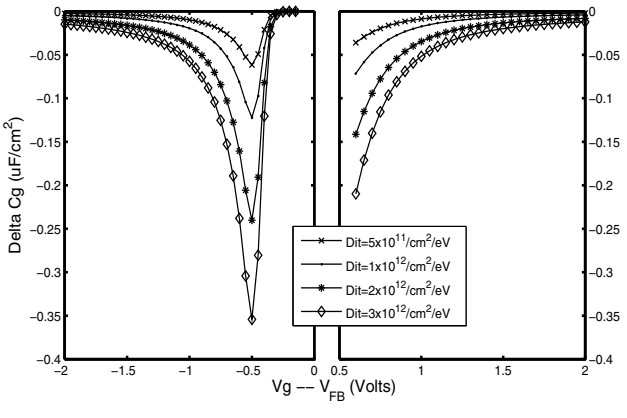


Fig.5. Change in HFCV due to effects of interface states.

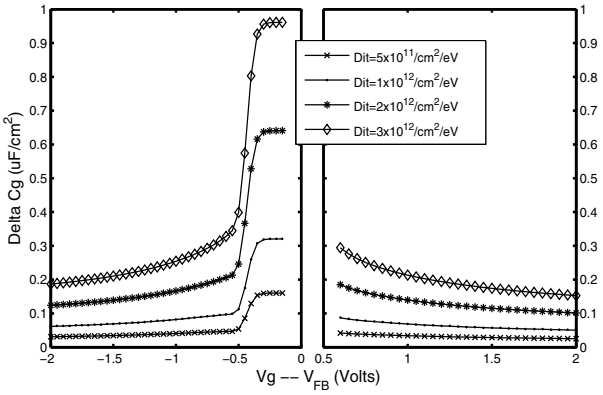


Fig.6. Change in LFCV due to effects of interface states

IV. Conclusion

An efficient technique for predicting the Capacitance-Voltage characteristic of n-channel double gate MOS structures has been demonstrated. The model can be used for modeling the CV characteristics of both symmetric and asymmetric, n-channel and p-channel DG MOS structures. The effect of interface states is found to be significant in both inversion and accumulation regions of DG nMOS devices. At low frequency (LF) operation the effect of interface states were found to be more prominent than at high frequency (HF) operation. Thus the operation of DGMOS devices in VLSI circuits at low frequency signals has to be modeled incorporating the effects of interface states.

References

[1] D. Frank, S. Laux, and M. Fischetti, "Monte Carlo simulation of a 30nm dual-gate MOSFET: How short can silicon go?," *IEDMTech Dig.,* p. 553, 1992.

[2] F. Balestra and S. Cristoloveanu *et al.,* "Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance," *IEEE Electron Device Letter.,* vol. EDL-8, p. 410, 1987.

[3] D. K. Schroder, J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing", *Journal of Applied Physics*, Vol 94, No. 1, July 2003.

[4] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost and S. Krishnan, "Impact of Negative Bias Temperature Instability on Digital Circuit Reliability" *IEEE 02CH37320. 40th Annual International Reliability Physics Symposium*, Dallas, Texas. 2002

[5] F. Stern, "Self-consistent results for n-type Si inversion layers," *Phys. Rev. B*, vol. 5, pp. 4891– 4899, 1972.

[6] M.K. Alam, A. Alam, S. Ahmed, M.G. Rabbani, Q.D.M. Khosru, **"***An Accurate and Fast Schrodinger-Poisson Solver using Finite Element Method***"**, Proceedings of the 18th IASTED International Conference on Modeling and Simulation ~MS 2007~, Montreal, Canada, pp.246-249, June, 2007.

[7] Y. Tsividis, *Operation and Modeling of MOS Transistor* (MacGraw-Hill, Chapter-2, 1999)

[8] P. V. Gray, D. M. Brown, "Electrical activity of interfacial paramagnetic defects in thermal (100) Si/SiO_2", *Appl. Phys. Lett*. 8, 31 (1966).

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Functional Device Design using Nonuniform Gate Voltage: Analytical Model of a Novel MOSFET

Suhad Shembil, *Member*, *IEEE*

19 Churchill Road, Morwell, Victoria 3840, Australia
suhad@ieee.org

Abstract - A new mode of MOSFET operation with distributed gate voltage is proposed and a one-dimensional (1D) analytical model presented. Operation in the strong inversion regime of the laterally nonuniform (LNU) long channel devices is analyzed at room temperatures using gradual channel approximation considering carrier transport by drift alone. Models with linearly graded gate voltage show that these four terminal devices are capable of performing multiplication operation in a single device, making them inherently superior in speed, linearity and power dissipation to multistage multiplier circuits currently used in communication and analog VLSI circuits.

I. Introduction

Device scaling received major attention of the researchers during the last three decades with a view to achieve faster and denser VLSI processor circuits. Transistor dimensions decreased steadily as technological enhancements were added. CMOS production technology crossed the 90 nm mark near the beginning of this century [1, 2]. However fundamental physical and process limitations are approached rapidly with continued scaling and advanced MOS structures, like ultrathin body (UTB), silicon-on-insulator (SOI), multiple gate (MuG) and Gate-All-Around (GAA) transistors were needed for more aggressive device scaling beyond 65 nm technology node (25 nm gate length). Significant performance advantages, gained through novel geometric shapes and in conjunction with other device specific innovations, pushed the scaling limits further [2, 3].

In contrast, distinctive gains can be achieved by adding new functionalities to devices [4]. Complex functional devices are likely to dominate the post multi-gate transistor technology era, with their full potential displayed in two and three-dimensional (2D and 3D) device architectures. Subnanometer dimensions, quantum phenomena, embedded elements, spatially localized effects, band-gap and strain engineering, lateral nonuniformities, etc. are likely to play dominant roles in the development of next generation devices [4].

We have taken a step forward in this direction and explored analytically a four-terminal MOS structure with laterally nonuniform gate voltage (LNUGV). The 1D device structure used for modeling is depicted in Fig. 1, which shows an ordinary MOSFET with the gate contacts modified for the application of distributed gate voltage.

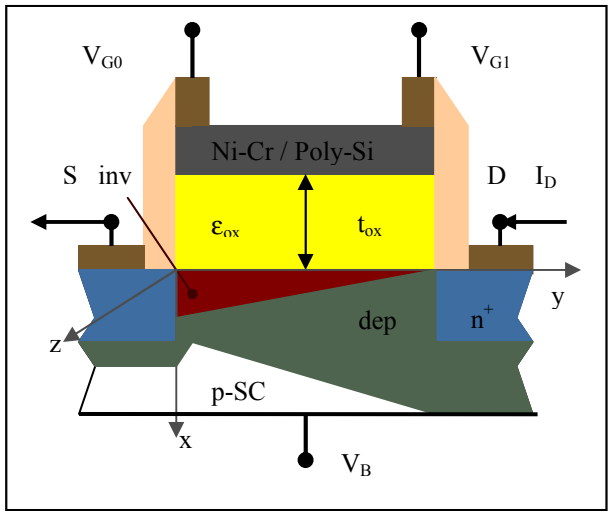


Fig. 1 Crossection of the LNUGV-MOSFET structure with depletion (dep) and inversion (inv) regions in the substrate.

II. Analytical Model

An enhancement-mode transistor on a p-type substrate is considered for modeling the LNUGV-MOSFET. A polysilicon [5, 10] or a high resistivity metal film [6] on a dielectric layer with two contacts parallel to source and drain are used to apply the linearly varying gate voltage $V_G(y)$ along the length of the channel. The device is oriented laterally along the y-axis and vertically along the x-axis. All non-uniformities and fringe effects in the z-direction are ignored in the 1D model. The p-surface is brought from flat-band and depletion conditions to inversion by positive gate voltages. In strong inversion, $V_G(y) > V_T(y)\ \forall\ y$ and $\rho_{s,n}(y)$, the mobile surface charge density is given by [7] –

$$\rho_{s,n}(y) = -C_{ox}\left[V_G(y) - V_T(y) - V(y) + V_S\right] \quad (1)$$

where $V(y)$ is the voltage dropped in the channel due to current flow, $V_T(y)$ is the threshold voltage at y, $V_G(y) = V_{G0} + \Delta V_G(y/L)$ is the gate voltage along y-axis, $\Delta V_G = V_{G1} - V_{G0}$ and $C_{ox} = t_{ox}/\varepsilon_{ox}$, the capacitance per unit area of oxide layer [7]. The surface drift current $I(y) = \int \boldsymbol{J(r)}.\boldsymbol{ds} = W\rho_{s,n}(y)v(y)$, where $\boldsymbol{ds} = \boldsymbol{a}_y dxdz$, and both $J_y(y)$ and the y-component of electric field are taken to be positive in y-direction. Constant $I(y)$ $(= -I_D)$, electron drift velocity $v(y) = \mu_{eff}.(dV/dy)$ and gradual channel approximation give the drift current in the strong inversion regime as [7]-

$$I_D = A\left[V_{G0} - V_T + \frac{\Delta V_G y}{L} - V(y) + V_S\right]\frac{dV}{dy} \quad (2)$$

$A \equiv C_{ox}\mu_{eff}W$ and $V_T = V_T(0)$ is the threshold voltage *at the source* in the above equation, neglecting bulk charge effect. Equation (2) is valid as long as the channel is not pinched off, i.e. $|\rho_{s,n}(y)| > 0$ or $V(L) = V_D < V_{D,sat}$.

We redefine the coefficients in equation (2) as $a \equiv \Delta V_G/L$, $b \equiv V_{G0} - V_T + V_S$ and $c \equiv I_D/A$. Equation (2) then takes the form $(V - ay - b)dV + cdy = 0$ which can be solved by substitution to give -

$$V(y) - ay - b_1 = K_1 \exp[\beta V(y)] \quad (3)$$

$\beta \equiv a/c$, $b_1 \equiv (b - c/a)$, K_1 an arbitrary constant and $a \neq 0$ in the equation above which is valid for both polarities of a. When $a = 0$, equation (3) becomes an exact DE and is integrated directly giving a quadratic dependence of I_D on $V(y)$ [7].

A. Non-linear Region: $V_{DS,sat} > V_{DS} \geq 0, \Delta V_G > 0$

Equation (3) can be solved for the specific case of $V(0) = V_S$ and $V(L) = V_D$ to get the I_D - V_{DS} relation in the non-linear region as -

$$I_D = G_0\left\{(V_{G0} - V_T) + \frac{V_{DS} - \Delta V_G}{\exp(\beta V_{DS}) - 1}\right\} \quad (4)$$

$V_{DS} \equiv V_D - V_S$, $G_0 \equiv A_0(\Delta V_G)$, $A_0 \equiv A/L = C_{ox}\mu_{eff}W/L$ and $\beta = A_0\Delta V_G/I_D > 0$. To facilitate further analysis, the above non-linear equation may be expressed parametrically as –

$$I_D = \frac{A_0(\Delta V_G)}{\xi - \ln(1+\xi)}\left(\xi(V_{G0} - V_T) - \Delta V_G\right)$$
$$V_{DS} = \frac{\ln(1+\xi)}{\xi - \ln(1+\xi)}\left(\xi(V_{G0} - V_T) - \Delta V_G\right) \quad (5)$$

It is required that $\xi > [\Delta V_G/(V_{G0} - V_T)] > 0$ to get positive values of I_D and V_{DS}.

B. Linear Region: $V_{DS} \to 0$

In the limit $V_{DS} \to 0$, we get from Equation (4) -

$$I_D = A_0(V_{G0} - V_T)V_{DS} \quad (6)$$

The equation applies for both polarities of V_{DS} and matches regular MOSFET equation in linear region [8].

C. Saturation Region: $V_{DS} \geq V_{DS,sat} > 0, \Delta V_G > 0$

At the onset of pinch off $V_D = V_{D,sat}$. One gets $V_{DS,sat} \equiv V_{D,sat} - V_S = V_{G1} - V_T$ by setting $\rho_{s,n}(L) = 0$ in equation (1). Substitution in equation (4) gives $I_{D,sat}$ as –

$$I_{D,sat} = I_S\left\{1 + \left(\exp\left(\frac{A_0\Delta V_G V_{DS,sat}}{I_{D,sat}}\right) - 1\right)^{-1}\right\} \quad (7)$$
$$I_S \equiv A_0\Delta V_G(V_{G0} - V_T)$$

D. Other Limiting Cases: (a) $\Delta V_G \to 0$

In this case, we have -

$$I_D \to A_0 V_{DS}\left\{(V_{G0} - V_T) - \frac{V_{DS}}{2}\right\}$$
$$I_{D,sat} \to \frac{1}{2}A_0(V_{G0} - V_T)^2 \quad (8)$$

Thus in the limiting case of $\Delta V_G \to 0$ equations of I_D and $I_{D,sat}$ of NUGV-MOSFET (equations (4) and (7) respectively) reduce to the familiar forms of a constant V_G MOSFET (equation (8), see [7]).

E. Other Limiting Cases: (b) $\Delta V_G >> (V_{G0} - V_T)$

In this case, we have from equation (7) -

$$I_{D,sat} \to I_S^+ = A_0(\Delta V_G)(V_{G0} - V_T)^+ \quad (9)$$

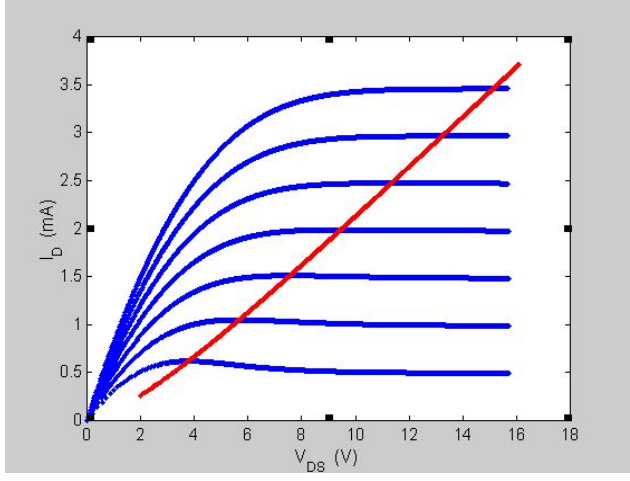


Fig. 2 Drain current-voltage characteristics (I_D vs. V_{DS}) in the nonlinear regime as predicted by the model. V_{G0} is held constant at $1.7V_T$ and ΔV_G varied as the line parameter. Corresponding saturation curve ($I_{D,sat}$ vs. $V_{DS,sat}$) is shown superimposed (red curve), indicating limits of validity of the nonlinear regime.

III. Results

An n-channel enhancement mode GaN-on-Sapphire MOSFET, with polysilicon gate and SiO_2 gate dielectric, is used to produce the simulation results. The following parameters for a long channel device are taken from [5] and [10] - width $W = 800$ μm, length $L = 20$ μm, $V_T = 2.7$ volts, effective surface mobility $\mu_{eff} = 100$ cm^2/volt-sec, oxide layer thickness $t_{ox} = 100$ nm and $\varepsilon_{ox} = 3.9\varepsilon_0$. Polysilicon gate materials have sheet resistance values around 2kΩ/□ [10]. Alternatively, high resistivity metal films may be used as gate metals (e.g. Nichrome with ρ ≥ 103μΩ-cm [6]).

I_D's predicted by the nonlinear model reach a maximum before starting to drop off (Fig. 2). Model is valid up to the saturation points. After reaching the maximum values, I_D's will remain constant unless predicted otherwise by physics-based saturation region model. Figure 3 shows $I_{D,sat}$ vs. ΔV_G with V_{G0} held constant at $1.7V_T$ (equation

(7)). Curvature of $I_{D,sat}$ vs. ΔV_G is checked against a straight line to show its high degree of linearity for larger values of ΔV_G.

IV. Application

In the limit $\Delta V_G >> (V_{G0} - V_T)$, we have from equation (9), $I_{D,sat} \approx A_0 \Delta V_G (V_{G0} - V_T)$. This property can be used to multiply two small-signals applied to the gate terminals with appropriate biases added, retrieving the desired components of the signals later through signal processing. Since the multiplication operation is performed by a single device without needing any additional components- higher speed and device density, better linearity and lower power dissipation, device noise power spectral density and distortion, can be achieved for same performance level by this device than multistage circuits currently used for such operations [9, 11]. The device will find extensive application in wireless communication and analog VLSI circuits such as multipliers, modulators, demodulators, low harmonic content mixers, phase detectors, etc.

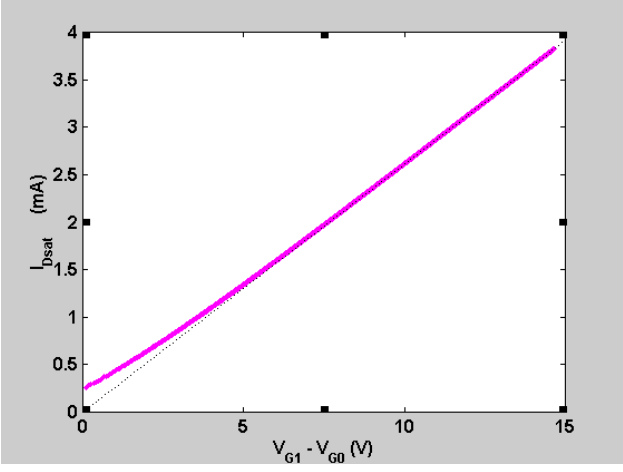


Fig. 3 Drain saturation current ($I_{D,sat}$) vs. potential difference between the two gates (ΔV_G), with $V_{G0} = 1.7V_T$, placed against a straight line, showing high degree of its linearity.

V. Conclusion

A new device concept using continuously varying gate voltage on an otherwise regular MOS transistor is investigated analytically and *I-V* characteristics are obtained for the strong inversion regime and enhancement-mode operation. Gradual channel approximation, transport by drift and long channel limitations are used in the model to derive the analytical expressions. Source-channel and channel-drain diffusion currents are not significant above threshold and justify use of drift model in inversion channel. In addition, the analytical models derived will aid in circuit designs greatly.

I_D-V_{DS} characteristics obtained are similar in appearance to standard MOSFET characteristics. $I_{D,sat}$ vs. ΔV_G shows excellent linearity in the ΔV_G x $(V_{G0} - V_T)$ characteristics of the devices. Superior performance as an analog multiplier is expected over contemporary multiplier circuits in terms of speed, noise, heat generation and device density as the operation is performed within a single device.

References

[1] D. J. Frank et al, "Optimizing CMOS technology for maximum performance," IBM J. Res. & Dev., vol. 50, no. 4/5, pp. 419-431, July/September 2006.

[2] L. Chang et al., "Extremely Scaled Silicon Nano-CMOS Devices," Proc. IEEE, vol. 91, no. 11, pp. 1860-1873, November 2003.

[3] H. -S. P. Wong, "Beyond the conventional transistor'" IBM J. Res. & Dev., vol. 46, no. 2/3, pp. 133-168, March/May 2002.

[4] T. Hiramoto, M. Saitoh, and G. Tsutsui, "Emerging nanoscale silicon devices taking advantage of nanostructure physics," IBM J. Res. & Dev., vol. 50, no. 4/5, pp. 411-418, July/September 2006.

[5] W. Huang, T. Khan and T. P. Chow, "Enhancement-mode n-channel GaN MOSFETs on p and n-GaN/Sapphire substrates," IEEE Electron Device Lett., vol. 27, no. 10, pp. 796-798, October 2006.

[6] S. Shembil, *SOS Cryogenic Bolometer: Analysis, Development and Self-Calibration*, Ph. D. thesis, Dept. Elec. & Comp. Eng., Univ. of Colorado, Boulder, Colorado, 1990.

[7] R. S. Muller, T. I. Kamins and M. Chan, *Device Electronics for Integrated Circuits*, 3rd ed., John Wiley and Sons, New York, NY, 2003, pp 390-396, 429-461.

[8] M. S. Lundstrom and J. Guo, *Nanoscale Transistors: Device Physics, Modelling and Simulation*, Springer-Verlag, New York, NY, 2006, pp 61-63.

[9] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., John Wiley and Sons, New York, NY, 1993.

[10] K. M. Matocha, T. P. Chow and R. J. Gutmann, "High-voltage normally off GaN MOSFETs on sapphire substrates," IEEE Trans. Electron Devices, vol. 52, no. 1, pp. 6-10, January 2005.

[11] C. Chen, and Z. Li, "A low-power CMOS analog multiplier," IEEE Trans. Circuits Syst. II, vol. 53, no. 2, pp.100-104, February 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Novel Four-Quadrant Analog Multiplier using Laterally Nonuniform Gate Voltage MOSFET

Suhad Shembil, *Member*, *IEEE*

19 Churchill Road, Morwell, Victoria 3840, Australia
suhad@ieee.org

Abstract - A four terminal MOSFET with gate voltage nonuniformly distributed along the channel (LNUGV) was analyzed recently by the author and was found to be capable of performing first-quadrant multiplication operation in a single device, making them inherently superior to multistage multiplier circuits currently used in communication and analog VLSI circuits. Fully active two-quadrant and four-quadrant analog multiplier circuits, using identical n-channel LNUGV-MOSFETs as core transconductors, operating in enhancement mode, saturation regime and within specific ranges, are presented in this paper.

I. Introduction

Multipliers and linear transconductors are essential circuit blocks for designing analog VLSI for signal processing and communication. Fully active VLSI implementation of multipliers find extensive use in modulators, mixers, frequency converters, adaptive filters, signal generators, integrated sensors, neural classifiers, data converters, field programmable arrays, etc. [4, 5].

Among many circuits used for multiplier implementation, Gilbert Cells in bipolar technology [6, 7] or its CMOS version [8] and four-quadrant analog multipliers [3-5] in MOS, CMOS or BiCMOS are popular configurations.

Four-quadrant multipliers designed for implementation on CMOS, NMOS or BiCMOS technology make use of the square-law *I-V* characteristics of the standard MOS transistors operated in saturation regimes [3 – 7]. Triode regimes are also used for designing multipliers for applications requiring low power consumption [5, 9].

In a recent paper [1], the author has analyzed a four-terminal version of MOSFET with distributed gate voltage, varied laterally along the channel, and found its output characteristics to exhibit first quadrant multiplication operation in the saturation regime. The device structure and a symbol used to represent it are shown in Fig. 1.

In this paper we analyze an NMOS circuit to implement a four-quadrant analog multiplier (4-QAM) using a LNUGV N-MOSFET.

II. LNUGV-MOSFET V→I Converter

The proposed laterally nonuniform gate voltage (NUGV) MOSFET modeled for operation in the inversion regime yielded the following drain current dependency on terminal voltages [1], -

$$I_D = G_0\left\{\left(V_{G1} - V_T\right) + \frac{V_{DS} - \Delta V_G}{\exp\left(\beta V_{DS}\right) - 1}\right\} \qquad (1)$$

in the triode region with

$V_{DS} \equiv V_D - V_S$
$\Delta V_G = V_{G2} - V_{G1} > 0$
$G_0 \equiv A_0 \cdot \Delta V_G$
$A_0 \equiv C_{ox} \cdot \mu_{eff} \cdot (W/L)$
C_{ox} = Gate oxide capacitance per unit area
μ_{eff} = effective inversion channel electron mobility
W = Width of the channel
L = Length of the channel and
$\beta = A_0 \Delta V_G / I_D > 0$.

In the saturation region, for $\Delta V_G > 0$, it has been shown that [1] -

$$I_{D,sat} = I_S\left\{1 + \left(\exp\left(\frac{A_0 \Delta V_G V_{DS,sat}}{I_{D,sat}}\right) - 1\right)^{-1}\right\}$$
$$I_S \equiv A_0 \Delta V_G \left(V_{G1} - V_T\right) \qquad (2)$$
$$V_{DS,sat} = V_{G2} - V_T = V_{G1} + \Delta V_G - V_T$$

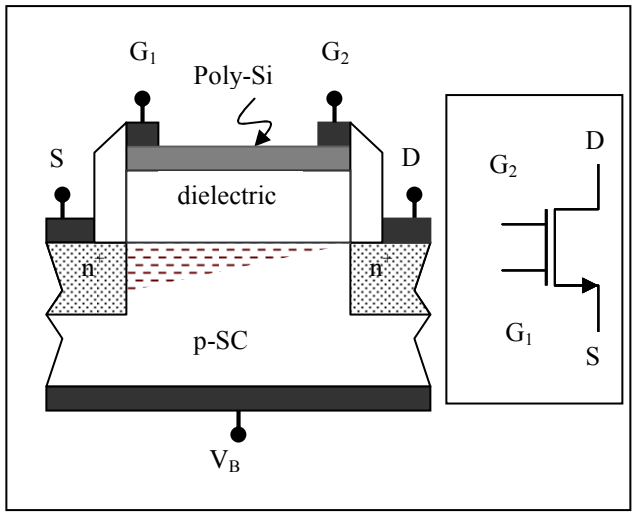


Fig. 1 Structure of the proposed LNUGV-MOSFET with its adopted symbol shown in the inset.

For the specific case of $\Delta V_G >> (V_{G1} - V_T)$, equation (2)

for $I_{D,sat}$ reduces to [1] -

$$I_{D,sat} \rightarrow I_S^+ = A_0\left(V_{G2} - V_{G1}\right)\left(V_{G1} - V_T\right)^+ \qquad (3)$$

A plot of $I_{D,sat}$ vs. ΔV_G shows that the difference gain of a nonuniform gate voltage (LNUGV) MOS transconductor is nearly linear for ranges $V_{G2} > 2.2V_{G1}$ ($\Delta V_G \geq 2V_T$) (see Fig. 4).

III. The Analog Multiplier Circuit

Thus, from equation (3), it is seen that the saturation current of a single device core represents one-quadrant multiplier operation containing two offset terms and one non-linear term –

$$I_{D,sat} \approx A_0\left[V_{G1}V_{G2} - k_V V_{G2} + k_V V_{G1} - V_{G1}^2\right] \qquad (4)$$

where $k_V \equiv V_T$. A circuit can be constructed using two identical core NUGV MOS transconductors to cancel the quadratic term as shown in Fig. 2.

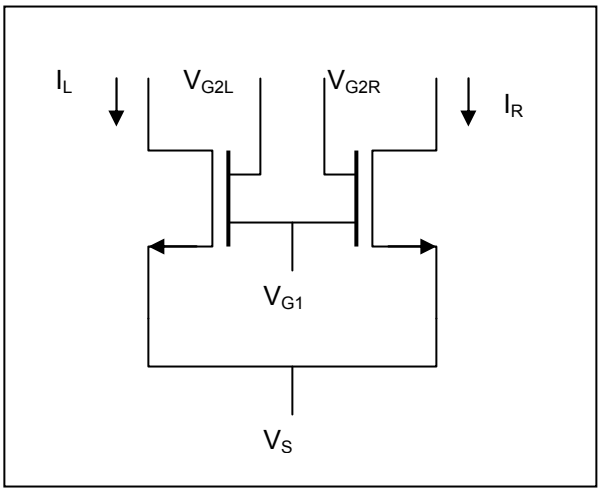


Fig. 2 Two-quadrant multiplier using LNUGV-MOSFETs.

The difference of the output currents of the resulting two-quadrant multiplier is given by –

$$I_R - I_L \approx A_0\left(V_{G2R} - V_{G2L}\right)\left(V_{G1} - k_V\right) \qquad (5)$$

In the above equation, $V_{G1} = V_{G1R} = V_{G1L}$. Next, the two-transistor multiplier circuit is duplicated and the output currents cross coupled for cancellation of the offset terms. The final circuit of the four-quadrant analog multiplier (4-QAM) is shown in Fig. 4. The output currents of the 4-QAM are given by-

$$\begin{aligned} I_A &= I_{XR} + I_{YL} \\ I_B &= I_{YR} + I_{XL} \end{aligned} \qquad (6)$$

The final output difference current of the 4-QAM is obtained, using equations (5) (adding subscripts X and Y to denote the specific 2-QAM sub-circuit) and (6), as –

$$I_A - I_B \approx A_0\left(V_R - V_L\right)\left(V_{XG1} - V_{YG1}\right) \qquad (7)$$

The corresponding differential transconductance of the 4-QAM circuit is obtained simply as [5] –

$$g_m = A_0\left(V_{XG1} - V_{YG1}\right) \qquad (8)$$

In the derivation of equation (7) above, $V_R \equiv V_{XG2R} = V_{YG2R}$ and $V_L \equiv V_{XG2L} = V_{YG2L}$ have been used. In other words, the left and right G2 gates of one 2-QAM ("X") have been tied with the corresponding G2 gates of the second 2-QAM ("Y") respectively.

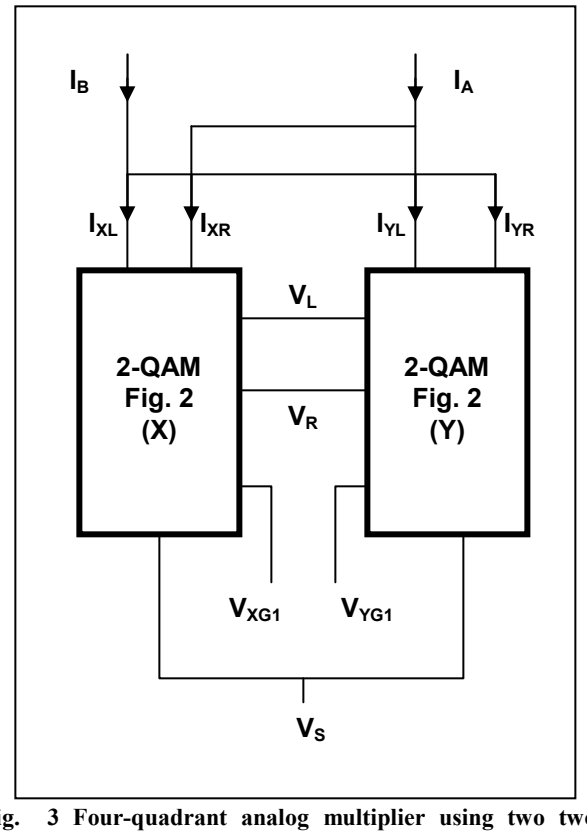


Fig. 3 Four-quadrant analog multiplier using two two-quadrant multipliers shown in Fig. 2.

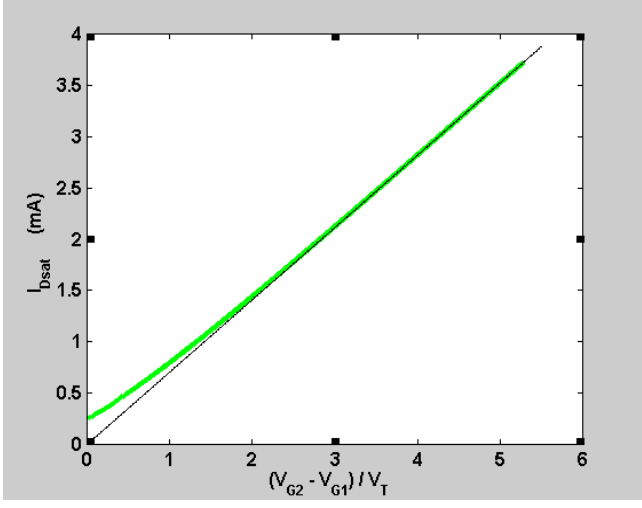


Fig. 4 Drain saturation current ($I_{D,sat}$) vs. potential difference between the two gates (ΔV_G), with $V_{GI} = 1.7V_T$, placed against a straight line, showing high degree of its linearity.

IV. Ranges and Results

An n-channel enhancement mode GaN-on-Sapphire MOSFET, with polysilicon gate and SiO_2 gate dielectric, is used to produce the simulation results. The following parameters for a long channel device are taken from [10] and [11] - width $W = 800$ μm, length $L = 20$ μm, $V_T = 2.7$ volts, effective surface mobility $\mu_{eff} = 100$ cm^2/volt-sec, oxide layer thickness $t_{ox} = 100$ nm and $\varepsilon_{ox} = 3.9\varepsilon_0$. Polysilicon gate materials have sheet resistance values around 2kΩ/□ [11].

Enhancement-mode operation requires minimum gate voltage at any point be larger than V_T. Therefore it is required that, $V_{XG1} > V_T$ and $V_{YG1} > V_T$. The corresponding minimum values of V_L and V_R should also be maintained at values greater than $2V_{XZ1}$ or $2V_{YG1}$. This can be done by adding equal dc bias voltages to both inputs so that the difference voltage can be controlled within a desired range. The requirement of having V_L (V_R) > $2V_{XG1}$ ($2V_{XG1}$) > $2V_T$ puts a limit on the maximum value of g_m (equation (8)) through the maximum values of gate voltages for any specific construction.

A differential current converter may be used to get a single-ended output current proportional to the difference in currents. Alternatively, a single-ended voltage output proportional to the difference of currents may be generated using current mirrors and a differential amplifier.

V. Conclusion

The LNUGV N-MOS in saturation regime implements a first-quadrant multiplier in a single device. Two such devices have been used to make two-quadrant multiplier and a four device core can implement a four-quadrant multiplier. The number of devices required to implement a 4-QAM using LNUGV-MOSFET is thus considerably reduced over the CMOS or BiCMOS design. This in turn helps to reduce interconnection wiring and nodes, device noise, interference, device density and heat dissipation. Fully active design is also compatible with current technology and is expected to exhibit superior performance over existing design configurations. Four-quadrant multipliers is also comparable to the only multigate FET 4-QAM designed and tested by Akarvardar, et al [2], where four G^4-FETs were used, same as the number of LNUGV transistors used in our case. Short channel, velocity saturation, quantum confinement and other second order effects (body effect, channel length modulation, etc.) need to be addressed for further refinement of the circuits or models. Triode operation may also be considered for developing circuits for low voltage and low power applications.

References

[1] S. Shembil, "Functional device design using nonuniform gate voltage: analytical model of a MOSFET," in Proc. ICECE 2008, Dhaka, Bangladesh, accepted, 20-22 December 2008.

[2] K. Akarvardar et. al., "A novel four-quadrant analog multiplier using SOI four-gate transistors (G^4-FETs)," in Proc. ESSCIRC 2005, Greenoble, France, pp. 499-502, Sep. 2005.

[3] K. Bult and H. Wallinga, "A CMOS four-quadrant analog multiplier," IEEE J. Solid-State Circuits, vol. SC-21, no. 3, pp. 430-435, Jun. 1986.

[4] A. Demosthenous and M. Panovic, "Low-voltage MOS linear transconductor/squarer and four-quadrant multiplier for analog VLSI," IEEE Trans. Circuits Syst. I, vol. 52, no. 9, pp.1721-1731, September 2005.

[5] M. Ismail and T. Fiez, *Analog VLSI: Signal and Information Processing*, McGraw-Hill, Inc., New York, NY, 1994, pp. 63-83.

[6] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., John Wiley and Sons, New York, NY, 1993.

[7] A. B. Grebene, *Bipolar and MOS Analog Integrated Circuit Design*, John Wiley and Sons, New York, NY, 1984.

[8] S. Qin and R. L. Geiger, "A ± 5-V CMOS analog multiplier," IEEE J. Solid-State Circuits, vol. SC-22, no. 6, pp. 1143-1146, Dec. 1987.

[9] C. Chen, and Z. Li, "A low-power CMOS analog multiplier," IEEE Trans. Circuits Syst. II, vol. 53, no. 2, pp.100-104, February 2006.

[10] W. Huang, T. Khan and T. P. Chow, "Enhancement-mode n-channel GaN MOSFETs on p and n-GaN/Sapphire substrates," IEEE Electron Device Lett., vol. 27, no. 10, pp. 796-798, October 2006.

[11] K. M. Matocha, T. P. Chow and R. J. Gutmann, "High-voltage normally off GaN MOSFETs on sapphire substrates," IEEE Trans. Electron Devices, vol. 52, no. 1, pp. 6-10, January 2005.

5th Internatinonal Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

New Approach for Throughput Analysis of IEEE 802.11 in AdHoc Networks

T.D.Senthilkumar
ECE Dept.
K.S.R College of Tech.
TamilNadu, India
tdsenthil_ece@yahoo.co.in

A. Krishnan
ECE Dept.
K.S.R. College of Tech.
TamilNadu, India
a_krishnan26@hotmail.com

P.Kumar
ECE Dept.
K.S.R College of Tech.
TamilNadu, India
kumar_ksrct@yahoo.co.in

***Abstract* – This paper provides an analysis of optimum offered load that maximizes the throughput of IEEE 802.11 Media Access Control (MAC) protocol for the multihop network. We extend Ping Chung Ng [1] analysis in presence of hidden terminal and capture effect, with some modifications in the derivation of throughput. The analytical throughput results are validated against simulation using Direct Sequence Spread Spectrum (DSSS) MAC layer. Simulation result shows that analytical results are valid under assumed conditions. This approach analyses the resulting tradeoffs between the performance of the network is hidden node limited and spatial reuse limited. In this paper, we provide a throughput analysis of the IEEE 802.11 based multihop network for various distances between the two successive nodes.**

***Text Index* – IEEE 802.11, MAC, DCF, DSSS, Multihop**

I. Introduction

The IEEE 802.11 standard specifies Wireless Local Area Networks (WLANs). It supports both basic access mechanism and four way hand shaking mechanism. IEEE 802.11 MAC specifies Distributed Coordination Function (DCF) and Point Coordination Function (PCF). PCF supports real time traffic which is not suitable for IEEE 802.11 based multihop networks. Multihop network opens the door for the researchers in the area of MAC protocol design. IEEE 802.11 based multihop network exhibits channel capture in which the connection with strong SNR (above the defined capture threshold value) can capture the channel. MAC layer throughput influences the network layer throughput. Packet dropping in the MAC layer causes the routing discovery in the network layer. The capacity of the multihop network is limited by the hidden nodes and the carrier sensing. The transmission range and the carrier sensing range are fixed for the radio channel but the interference range varies with respect to the physical layer technique and the distance between the source and the destination. In multihop network, some stations are source stations and some stations are relaying stations i.e. intermediate stations in the packet transmission. Saturated traffic in the multihop network severely degrades the performance of the network.

II. Related Work

In the multihop wireless networks, the collision probability is location dependent i.e. it depends on the distance between the source and the destination or next hop node. Reference [1] expressed the per hop throughput analysis for the string topology. This paper deals the source node which is the first node that experiences lighter contention than others but in the network all the nodes are experiencing equal contention. Most of the previous studies for the hidden node problem deals about the collision probability which is in terms of transmission probability. Here we are introducing air time [1] instead of transmission probability to investigate the throughput. The IEEE 802.11 MAC and PHY protocol [3] has achieved a worldwide acceptance for WLANs. Bianchi in [4] used Discrete Markov chain model for analyzing the DCF operation and calculate the saturation throughput. In particular, David Malone in [5] modeled DCF for non saturated case. In [6] IEEE 802.11 MAC protocol is modified to improve the throughput and extended its analysis to investigate the throughput degradation by false blocking due to the RTS and CTS. In this paper, we are not modifying the IEEE 802.11 MAC protocol. The collision probability for the multihop IEEE 802.11 is different from the single hop IEEE 802.11 [7]. Aruna [8] presented the effects of hidden and exposed terminal problem in multihop network. The rest of this paper is organised as follows. Section III analyses the parameters which influences the throughput degradation and estimates the offered traffic in a multihop networks. It presents the expression of throughput which is degraded by carrier sensing and hidden nodes. Section IV examines the numerical results and its discussions. Section V concludes the paper.

III. Throughput Analysis

The capacity of the network is limited by two major factors, namely 1).Hidden nodes 2).Carrier sensing mechanism.

Step1: Capacity Limited by Hidden Node Problem

To find the throughput, we are using the method proposed by [1], where air time is used instead of transmission probability. Consider a long time interval [0,Time], which includes the idle times, window backoff times, transmission times, collision times and busy times. The duration of time in which no node in the carrier sensing range transmits is the idle time. The duration of time for successful packet transmission and packet collision are transmission time and collision time respectively. During

the time interval, a steady state node 'i' uses the air times S_i that consists of the air times used by the successive packets of node 'i' and the times used up for retransmission in case of collisions. It does not include the countdown of the contention window since it is shared time by adjacent nodes also in the carrier sensing range of node 'i'.

Let 'p' be a collision probability for the transmission, 'T' is the traffic throughput and $x = |S_i| /$ Time is the fraction of time used by node 'i' in the interval [0,Time].

$$T = x \cdot (1-p) \cdot d \cdot \text{Trans\_rate} \quad (1)$$

Where, $d = \frac{\text{DATA}}{\text{DIFS PACKET ACK}}$

'd' is the proportion of time within 'x' that is used to transmit the data payload, Trans_rate is the data transmission rate.

PACKET = PHY_header + MAC_header + DATA

Consider node 4 in Fig.1. In this figure, the transmission of node 4 will not be collided with the transmission of nodes 2,3,5 and 6.

The assumptions in this analysis are,

1).The collision is mainly due to hidden nodes because the carrier sensing mechanism eliminates the collision due to the nodes in its carrier sensing range.

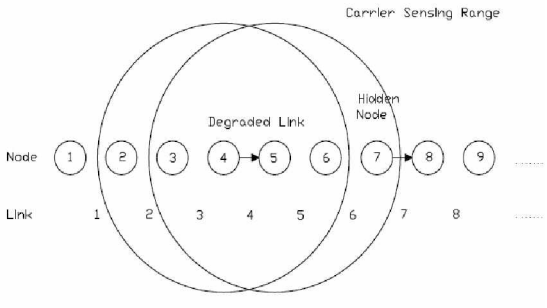


Fig.1. Node 7 as a hidden node for node 4

2).The collision due to at least two nodes in the same carrier sense range that reach zero contention window at the same time is negligible.

3). All the packets are in same size and same data rate is used for all packets.

4).Same and fixed transmission range and carrier sensing range for all nodes.

5).Spatial reuse cells access the channel for same time duration 'Time.'

From Fig.2, during the vulnerable period if the node 4 and node 7 transmits simultaneously it experiences a collision. We consider data flow in one direction. In this case there is no chance for DATA-ACK collision. Then the collision probability 'p' can be computed by finding the DATA-DATA collision probability (p_{HN}) and ACK-ACK (p'_{HN}) collision probability.

$$p = 1 - (1-p_{HN})(1-p'_{HN}) \quad (2)$$

In this paper, we are considering the signal capture. When both node 4 and node 7 are transmitting, signal capturing property allows the transmission from node 4 successfully, if it transmits before node 7. Suppose node 4 transmits first and the signal power at node 5 is P_4. Then node 7 transmits and the signal power at node 5 is P_7. If the signal power P_4 is greater than the signal power P_7 by an amount of threshold capture value, the transmission from node 4 will be successful. If node 7 transmits first, node 5 senses the signal from node 7 and the carrier sensing mechanism prevents node 5 to receive the signal from node 4. We assume that the distance between the successive nodes is constant such that all nodes experience the same situation. The MAC protocol parameters are given in the Table I.

Table I
Parameter Settings Used

DATA	1460 bytes
MAC header	28 bytes
PHY header	24 bytes
UDP header	20 bytes
ACK	14 bytes
Channel bit rate	11Mbps
PHY header bit rate	1Mbps
Slot time	20µsec
SIFS	10µsec
DIFS	50µsec
CW_{min}	32 slots
CW_{max}	1024 slots
Retransmission Limit	7

a . Analysis of collision probability by idden nodes for DA A DA A collisions

If the transmission of node 7 precedes node 4, the collision occurs in the transmission of node 4 at node 5. After receiving the physical header from node 7, node 5 will not receive the data from node 4 for the remaining time of packet duration. So, the vulnerable period offered by the hidden node to the transmission of node 4 is the duration of the MAC header and DATA.

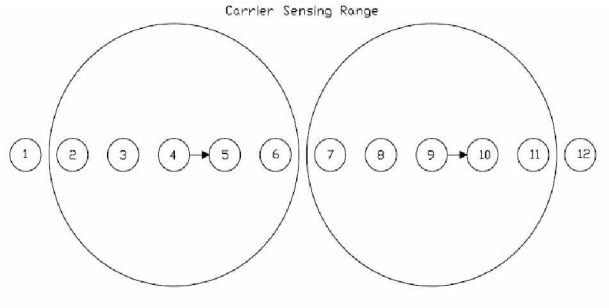


Fig.2. Spatial reuse cells in string multihop network

Consider a two geographically non-overlapping cell around node 4 and node 9. As far as spatial channel reuse is concerned, the time interval [0,Time] is shared by all nodes in each cell. From Fig.2, when node 6 or node 8 transmits node 7 can not transmit. This means S_6, S_7 and S_8 are non-overlapping airtimes. Now, nodes 6 and 8 use up 2x fraction of air time during the interval [0,Time].

As airtimes are common for all spatial channel reuse cells, airtimes of all the possible nodes which disallow the hidden node 7 to transmit should be considered in this analysis. When node 5 or node 9 transmits, node 7 can not transmit but S_5 and S_9 are overlapping airtimes. These nodes may or may not transmit at the same time. Node 5 or node 9 uses x fraction of air time during the interval [0,Time]. Then the remaining fraction of air time that node 4 and node 7 may collide is (1-3x). The collision probability offered by node 7 on node 4 is

$$p_{HN} = \frac{x}{1-3x} \cdot a \qquad (3)$$

Where,

$$a = \frac{MAC\_Header \quad DATA}{DIFS \quad PACKET \quad SIFS \quad ACK}$$

'a' is the fraction of airtime used for transmitting MAC header and data within x.

b . Analysis of collision probability by idden nodes for ACK ACK collisions

In Fig.3 node 1 and node 4 are outside the carrier sensing range of each other. When both node 1 and node 4 are transmitting simultaneously to node 2 and node 5 respectively, transmission of node 1 does not affect the transmission of node 4. Node 4 can sense ACK from node 2 to node 1. If ACK from node 5 reaches node 4 later than that of ACK from node2, a collision occurs.

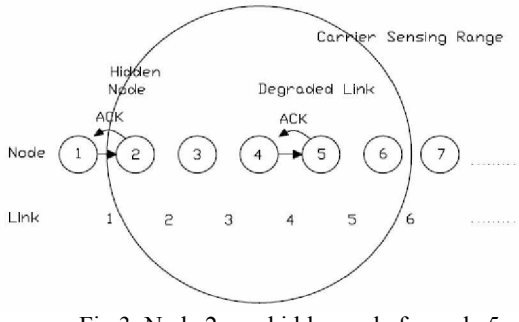


Fig.3. Node 2 as a hidden node for node 5

The vulnerable period offered by the node 1 to node 4 is duration of ACK. The ACK-ACK collision occurs if the transmission of the node 4 begins at time t<ACK later than the transmission of node 1. If t>ACK time, when node 2 is sending ACK, node 4 is in transmission and it is not aware of ACK from node 2. The fraction of airtime for the ACK-ACK collision at node 4 is (1-3x). The collision probability is

$$p'_{HN} = \frac{x}{1-3x} \cdot b \qquad (4)$$

Where,

$$b = \frac{ACK}{DIFS \quad PACKET \quad SIFS \quad ACK}$$

As per the specification given in Table I, the value of 'b' in the collision probability is equal to 0.00637. The 'b' value is very small so the chance for ACK-ACK collision is less. This can be ignored while calculating the collision probability.

Then,

$$p = p_{HN}$$

The expression for the throughput is simplified as,

$$T = x \cdot \left(1 - \frac{x.a}{1-3x}\right) \cdot d \cdot Trans\_rate \qquad (5)$$

Differentiating T with respect to x and equate to zero.

$$\frac{dT}{dx} = 0 \qquad (6)$$

The optimal value of x that maximizes the throughput is given by

$$x^* = \frac{3 \quad a - \sqrt{a^2 \quad 3a}}{3a \quad 9} \qquad (7)$$

Where x* is the optimum offered load which maximizes the throughput. Substitute this x* in (1) which yields the maximum throughput. Fig.4 shows that the throughput is maximum for the optimum offered load x*and the throughput is gradually decreased when the offered load is moving towards right or left from x*.

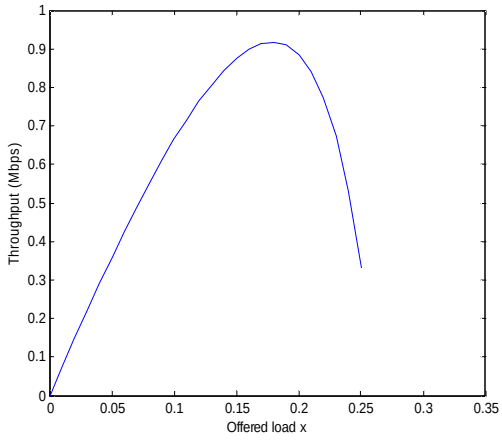


Fig.4. Offered load versus Throughput in a string multihop network

Step2: Capacity imited by Carrier Sensing Property

Carrier sensing property does not allow simultaneous transmission of nodes in the carrier sensing range of a node. This property limits the channel spatial reuse. Throughput T(x*) is the obtained maximum sustainable throughput in Step1. We have to verify whether carrier sensing supports the maximum throughput T(x*). If carrier sensing range is larger than the interference range, it avoids the collision. In this case carrier sensing property may disallow simultaneous transmissions that do not cause collisions. It unnecessarily limits the network capacity but avoids the collision due to simultaneous transmission. In this simulation, a node uses 1Mbps data rate for physical and MAC header and 11Mbps

data rate for DATA. It makes carrier sensing range larger than the interference range.

Consider node 4 as a local observer. Let C_i be the airtime used for counting down the contention window of node 'i'. During the time interval [0,Time], it observes the airtimes used by the nodes within its carrier sensing range. According to node 4, it observes C_4, S_2, S_3, S_4, S_5 and S_6. The airtime used for counting down the contention window of node 4 is shared time by other nodes in its carrier sensing range. The total airtimes used up by these nodes can not exceed 'Time.'

Thus,

$$|c_2 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6| \quad \text{Time}$$

Let

$$y = |c_2 \cup s_2 \cup s_3 \cup s_4 \cup s_5 \cup s_6| \ / \text{Time} \tag{8}$$

'y' is the fraction of airtime used up by the nodes in the carrier sensing range of node 4 during the interval [0,Time].

Now, $|c_4 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6|$ can be decomposed into

$$\begin{aligned} &|c_4| \ |s_2| \ |s_3| \ |s_4| \ |s_5| \ |s_6| - \\ &|c_4 \quad s_2| - |s_2 \quad s_3| - |s_2 \quad s_4| - \\ &\ldots \ |c_4 \quad s_2 \quad s_3| \ |s_2 \quad s_3 \quad s_4| \ \ldots. \end{aligned}$$

The intersection of airtime used for counting down of node 4 (C_4) and airtime used by other nodes (S_i) is null. Similarly the intersections of the airtimes used by any three and above nodes are null. Therefore the above expression is simplified into

$$\begin{aligned} &|c_4| \ |s_2| \ |s_3| \ |s_4| \ |s_5| \ |s_6| - |s_2 \quad s_5| \\ &- |s_3 \quad s_6| - |s_2 \quad s_6| \end{aligned} \tag{9}$$

A node selects a contention window size randomly between [0,CWmin-1] before the transmission. The average time for counting down the minimum contention window becomes (CWmin-1)σ/2 where, σ is the mini slot time. In this analysis hidden probability is zero so the contention window size is between [0,CWmin-1].

Let $z = |C_i| / \text{Time}$, be the fraction of time used for the counting down the contention window.

Thus,

$$z = x.c \tag{10}$$

Where,

$$c = \frac{(CWmin - 1) \ /2}{DIFS \quad PACKET \quad SIFS \quad ACK}$$

Then the expression for 'y' becomes

$$y = cx+5x- \frac{|s_2 \cap s_5|}{\text{Time}} - \frac{|s_3 \cap s_6|}{\text{Time}} - \frac{|s_2 \cap s_6|}{\text{Time}} \tag{11}$$

Node 2 and node 5 do not transmit when node 3 or node 4 transmits or node 4 is counting down. When node 6 transmits, node 2 can transmit but node 5 can not transmit. While the nodes in the string topology experience the data flow at the very first, the downstream nodes of the transmitting node will not have any packets. So, there is no overlapping time between the airtimes S_2 and S_6. The amount of airtime of node 2 that may overlap with that of node 5 is $|s_2| - |s_2 \quad s_6|$.

Then the intersection of S_2 and S_5 is

$$|s_2 \quad s_5| = \frac{|s_2| - |s_2 \quad s_6| |s_5|}{\text{Time} - |s_3| - |s_4| - |s_6| - |c_4|} \tag{12}$$

Similarly,

$$|s_3 \quad s_6| = \frac{|s_6| - |s_2 \quad s_6| |s_3|}{\text{Time} - |s_2| - |s_4| - |s_5| - |c_4|} \tag{13}$$

Here,

$$|s_2 \quad s_5| = |s_3 \quad s_6| \tag{14}$$

For $|s_2 \quad s_6|$, the amount of airtime of node 2 that may overlap with that of node 6 is $|s_2| - |s_2 \quad s_5|$ and the amount of airtime of node 6 that may overlap with node 2 is $|s_6| - |s_3 \quad s_6|$. Then,

$$|s_2 \quad s_6| = \frac{|s_2| - |s_2 \quad s_5| . |s_6| - |s_3 \quad s_6|}{\text{Time} - |s_3| - |s_4| - |s_5| - |c_4|} \tag{15}$$

When the node in the string topology experiences the data flow at the very first, the (12) is simplified into

$$|s_2 \quad s_5| = \frac{|s_2||s_5|}{\text{Time} - |s_3| - |s_4| - |s_6| - |c_4|} \tag{16}$$

$$|s_2 \quad s_5| = \frac{x^2}{1 - 2 \quad cx} . \text{Time} \tag{17}$$

Similarly,

$$|s_3 \quad s_6| = \frac{x^2}{1 - 2 \quad cx} . \text{Time} \tag{18}$$

Substitute (15) into (13)

$$|s_2 \quad s_6| = \frac{\left(x - \frac{x^2}{1 - 2 \quad cx}\right)^2}{1 - 3 \quad cx} . \text{Time} \tag{19}$$

Equation (12) encounters (19) when the string topology experiences continuous data flow and (12) is simplified as

$$\frac{|s_2 \quad s_5|}{\text{Time}} = \frac{x^2 1 - 2 \quad cx^2 - x^2 \quad 3 \quad cx^3}{1 - 2 \quad cx^2 1 - 3 \quad cx} \tag{20}$$

The expression for $|s_2 \quad s_6|$ can be obtained by substituting (20) into (15). In this step, the maximum throughput that can be supported by the network in absence of hidden terminals can be found.

By substituting x* in (11), we can identify whether the capacity of the network is limited by hidden terminal or spatial reuse. When spatial reuse is considered in step1, if y(x*)>1, the system is limited by the carrier sensing mechanism. Let x' be the x at which y(x) =1. This is the saturated case i.e. the nodes always have a packet to transmit. This will not be allowed if the system is hidden node limited. If T(x') is greater than the T(x*), the maximum sustainable throughput T(x*) can be supported by the network.

I . Results and Discussions

In section III, the expression of throughput for both hidden node limited case and carrier sensing limited case was derived. As per the parameters assumed in the Table I, the numerical results have to be verified with the simulation results. Fig.5 shows the simulation results which indicates that the sustainable throughput is 1.16Mbps for packet length of 1460 bytes. Our analytical results are approximately nearer to the simulation results.

Analytical results are given in Table II. The assumption in step1 [1] is, (1-2x) fraction of air time the node 7 may interfere the transmission from the node 4. The air time 'Time' is equal to sum of S_i for nodes 2,3,4,5 & 6, the time for counting down the contention window of node 4 and ideal time. From the x* value found in [1], the air times used by all nodes in the transmission range of node 4 is greater than the air time 'Time'. According to the spatial reuse concept, in the transmission range of node 4, if node 4 uses 'x' fraction of airtime then the remaining nodes in its carrier sensing range uses more than 'x' fraction of air time. Even spatial reuse was not considered, the x* value [1] given in the Table II is greater than 0.2. In our approach, we are also not considering the spatial reuse in step1 and the Table II shows that optimum offered load x* is less than 0.2.

Table II

Maximum Throughput in Both Steps for Packet Length of 1460 Bytes

Parameters	Analytical result	Analytical result in [1]
x* T(x*) Y(x*)	0.1787 0.91823 Mbps 0.8682	0.24445 1.2183 Mbps 0.95166
x' T(x') Y(x')	0.277 2.0839 Mbps 1	0.3110 2.3421 1

Table III compares the sustained throughput values for our analysis, analysis in [1] and simulation results. From Table III, our analysis result closely matches with the simulation result for small packet lengths. In step2, the optimal offered load x' was found without considering the hidden terminals.

Table III

Analytical and Simulation Throughput Results for Various Packet Lengths

Packet Length (in bytes)	Analytical result (Mbps)	Analytical result in [1] (Mbps)	Simulation result (Mbps)
300	0.469	0.587	0.501
500	0.622	0.752	0.677
1000	0.824	1.002	0.964
1460	0.918	1.218	1.160

By substituting x′ value and p=0 in (1), the maximum sustained throughput T(x') can be obtained for step2. The throughput T(x') is the maximum sustained throughput that can be supported by the network. If the throughput T(x') is less than the throughput T(x*) obtained in step 1, the capacity of the network is limited by the spatial reuse. When the number of nodes in the carrier sensing range is increased, it will decrease the throughput T(x') i.e. the capacity of the network is influenced by the spatial reuse. For y(x*) < 1, the capacity is limited by hidden node instead of carrier sensing. Then the sustained throughput T(x*) can be supported by the network. Fig.5 plots the sustainable throughput versus packet length for both analytical and simulated values.

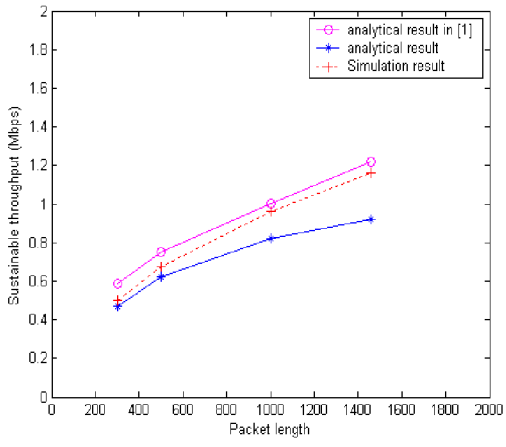


Fig.5. Packet length versus sustained throughput

Table IV shows the analytical and simulation throughput results for various distances between successive nodes in the string topology. The result in the table IV shows that if the distance is decreased, the performance of the network is further decreased by the spatial reuse concept in the string topology. But the throughput can be improved using the dynamic carrier sensing range for various distances. Fig.6 shows that our analytical results closely match with the simulation results while distance between successive nodes is decreased.

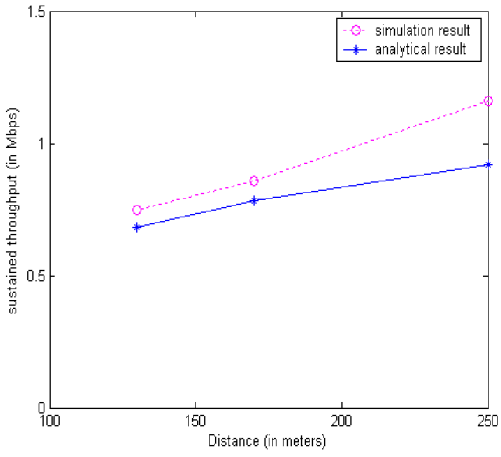


Fig.6. Sustained throughput versus the distance between the successive nodes

Table IV
Analytical and Simulation Results for Packet Length of 1460 Bytes

Distance (in meters)	Throughput (Analytical) in Mbps	Throughput (Simulation) in Mbps
250	0.9182	1.16
170	0.7812	0.8472
130	0.6824	0.7263

. Conclusions

This paper presents a throughput analysis for multihop network. This analysis is based on the work of [1] which allows one to identify whether the performance of the network is hidden terminal limited or spatial reuse limited. Our analytical results closely match with the simulation results for small packet sizes. When the packet size is increased, the throughput can be improved in absence of transmission errors. If there is more number of nodes in the carrier sensing range, the hidden probability is reduced and the capacity of the network is influenced by the spatial reuse. We conclude that the fixed carrier sensing range should be optimized to improve the network throughput performance for various distances between the successive nodes in the network. If the distance between successive nodes is decreased in the fixed carrier sensing range, the performance is influenced by the spatial reuse otherwise it is influenced by hidden terminals.

References

[1] P. C. Ng and S. C. Liew, "Throughput Analysis of IEEE802.11 Multi-Hop Ad- Hoc Networks", *IEEE ACM Trans. Networking,* vol.15, no.2, pp 309-322 April 2007.

[2] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[3] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1999.

[4] G.Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", *IEEE ournal on Selected Area in Communications,* vol.18, no.3, pp. 535-547, March 2000.

[5] David Malone, Ken Duffy, and Doug Leith "Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions", *IEEE ACM Trans. Networking,* vol.15, no.1, pp 159-172 February 2007.

[6] F. Alizadeh-Shabdiz, Suresh Subramaniam, "A Finite Load Analytical Model for the IEEE 802.11 Distributed Coordination Function MAC", WiOpt'03, March 2003.

[7] Saikat Ray and David Starobinski , "On False Blocking in RTS/CTS-Based Multihop Wireless Networks", *IEEE Trans. ehicular technology,* vol.56, no.2, pp 849-862 March 2007.

[8] Barowski, S.Biaz and P.Agrawal, "Towards the Performance Analysis of IEEE 802.11 in Multi-hop Ad-Hoc Networks", I*EEE Wireless Communications and Networking Conference*, Vol. 1, pp. 100-106 2005.

[9] Aruna Jayasuriya, et al, "Hidden vs. Exposed Terminal Problem in Ad hoc Networks", *Proc. of Australian Telecommunication Networks and Applications Conference,* Dec 2004.

[10] F. Ye, S. Yi and B. Sikdar, "Improving spatial reuse of IEEE 802.11 based ad hoc networks," in *IEEE lobal Telecommunications Conference LOBECOM '03* , San Francisco, CA, USA, December 1-5 2003.

[11] The network simulator - ns-2. [Online]. Available: http:/lwww.isi.edul nsnam/ns/

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Nonsaturation Throughput Analysis of IEEE 802.11 Distributed Coordination Function

T.D.Senthilkumar
ECE Dept.
K.S.R College of Tech.
TamilNadu, India
tdsenthil_ece@yahoo.co.in

A. Krishnan
ECE Dept.
K.S.R. College of Tech.
TamilNadu, India
a_krishnan26@hotmail.com

P.Kumar
ECE Dept.
K.S.R College of Tech.
TamilNadu, India
kumar_ksrct@yahoo.co.in

***Abstract* – This paper presents an analytical model for the performance study of IEEE 802.11 distributed coordination function (DCF). The maximum throughput can be achieved well below the saturated load. The analytical model for finite load must be studied to know the performance of the Medium Access Control (MAC) layer below the saturated load. We extend the Bianchi's saturation model for the nonsaturated environment. We employ discrete Markov chain to analyse the probability for attempting transmission at each node for an arbitrary time slot. In this paper, we explore the discrete Markov chain model with the post backoff. This model is validated against the simulation.**

***Text Index* – IEEE 802.11, MAC, DCF, nonsaturation**

I. Introduction

The IEEE 802.11 standard specifies Wireless Local Area Networks (WLANs). For recent years, WLANs are becoming more popular for wireless and mobile networks. IEEE 802.11 Medium Access Control (MAC) layer provides Distributed Coordination Function (DCF) and Point Coordination Function (PCF). DCF supports basic access mechanism and Request-To-Send/ Clear-To-Send (RTS/CTS) mechanism. DCF is an access scheme based on the contention principle using Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA). PCF mechanism is proposed for time bounded traffic.

A station uses Binary Exponential Backoff (BEB) algorithm in DCF for accessing the common channel. DCF and PCF have different performances. Most of the researchers are attracted towards DCF for their simplicity and flexibility. In DCF, contention window size is increased by using BEB algorithm. The station will decrease the contention window size by one whenever one idle slot time elapses. It will freeze this counter when the channel is sensed being busy. If the contention window size reaches zero, it can transmit the packet. If the transmitted packet collides, the station will assume that the channel is busy, then double the contention window and select a new backoff time for retransmission. Most of the previous analytical works are based on the discrete Markov chain model and it considers the saturation throughput. In real networks, packet may be queued at nodes buffer before being handled by the MAC protocol. In the saturation throughput analysis, the MAC service time is less than the packet arrival rate. In this paper, we present an extended Markov chain model to obtain the expression for the transmission probability and is expressed in terms of the stationary probabilities.

II. Related Work

Bianchi in [1] used Discrete Markov chain model for analyzing the DCF operation and calculate the saturation throughput. IEEE 802.11 DCF is generally in nonsaturated traffic condition. In particular, David Malone [2,3] modeled DCF for nonsaturated case and heterogeneous traffic. In this model, after successful transmission from the post backoff stage, state transition returns to the same post backoff stage regardless of the queue status. In our discrete Markov chain model, we consider the queue status after successful transmission from the post backoff stage. In [5], Bianchi model is extended to the case of finite loads. In this paper, the expression for the transmission probability is not derived and the state transition from post backoff stage to the other backoff stage is not clearly explained. There are some works on finite load models for IEEE 802.11 DCF by introducing a new idle state [6, 7]. These models are not considering the post backoff in the discrete Markov chain.

Bianchi's Markov chain model is used to analyze [8] the saturation throughput and the finite load condition is not considered. In our paper, we extend the previous works by looking at the important issues, namely nonsaturation and post backoff. Our assumptions are similar to those of assumptions in [2] except the state transition from post backoff stage to the stage zero in the Markov chain model. In the Markov chain model, the queue length is used as the third dimension [9]. In [9], an empty queue state is introduced to model the IEEE 802.11 MAC protocol and the post backoff stage is not used.

The rest of this paper is organised as follows. Section III analyses the nonsaturation throughput of the IEEE 802.11 DCF. It presents the expression of throughput of the finite load heterogeneous traffic. For heterogeneous traffic, different packet arrival rates are use in the network. Poisson process is used for the packet arrival in

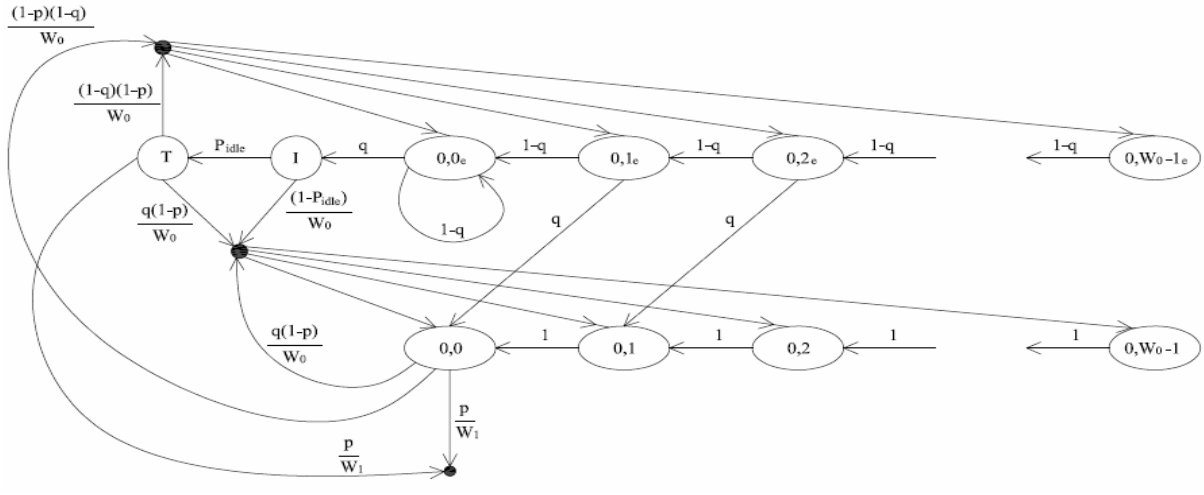


Fig. 1 Nonsaturated Markov chain model for IEEE 802.11

the network. Section IV examines the numerical results and its discussions. The analytical results are validated against the simulation results. Section V concludes the paper.

III. Throughput Analysis

Our analysis is based on the Bianchi model in [1] which gives the saturated throughput analysis for DCF accessing scheme at MAC layer. In real network, the traffic is heterogeneous traffic. Our analytical model is also heterogeneous traffic model. In the Markov model, we are employing post backoff when the transmission queue of the station is empty. From the per-station Markov model, the probability for a station attempting transmission can be computed.

A. Markov Model

Following the Markov model presented in [2], each station is modeled by pair of stochastic processes b(t) and s(t). Let b(t) be the stochastic process representing the backoff time counter. If a station finds a idle slot time, backoff counter decrements the counter one at the beginning of each idle slot time. When the channel is sensed being busy, a station stops the decrement and freeze the backoff counter.

Let s(t) be the stochastic process representing the backoff stage (0,1,……m) of the station. Let m be the maximum backoff stage. For mathematical convenience, a short notations (i,k) are used for the stochastic processes s(t) and b(t) respectively to represent the each state in the Markov model.

The backoff stage i starts at 0 at the first attempt to transmit and it increases by 1 every time in case of collision, upto maximum value m. Initially, a station selects a counter value uniformly from the contention window size $(0,W_0-1)$. The contention window size is doubled for every attempt, where $W_i=2^iW_0$, is the range of contention window size for i^{th} attempt. The station attempts to transmit when the backoff counter reaches zero. After successful transmission, the counter value is randomly chosen from the minimum contention window and initiates the new transmission.

At the maximum backoff stage, the packet is not discarded as per the Markov model in [1]. A new state $(0,k)_e$ is introduced for a station with empty queue after successfully transmitting a packet. In this case, a station need not wait and it can decrement its counter value. This is called post backoff. Per-station quantities in the Markov model are, q, the probability of at least one packet awaiting transmission at the start of a counter decrement; m, the maximum backoff stage; p, the probability of collision; b, the stationary probability of the state in Markov chain; τ, the probability that the station transmits in a slot. Markov chain's evolution is not real time, and so the estimation of throughput requires an estimate of the average state duration. The collision probability p is a constant value. The two dimension discrete time Markov chain is depicted in figure1. In this Markov chain, some transition probabilities [2] under our assumptions,

For $0 \le i \le m$ and $0 \le k \le W_i-1$

$$P[(i,k-1)/(i,k)] = 1$$
$$P[(0,k-1)_e/(0,k)_e] = 1-q$$
$$P[(0,k-1)/(0,k)_e] = q$$

If the backoff counter reaches zero, the station can transmit a packet. Whether it has a packet in a queue after a successful transmission, it selects a counter value from the window $(0,W_0-1)$ in stage 0 otherwise it selects a counter value from the window $(0,W_0-1)_e$ in the post backoff stage. The backoff stage is increased for each retransmission. This Markov model allows us for infinite retransmission attempts, which does not need extra per-station parameters.

In the Markov model of our analysis, we are introducing the state I for checking whether the channel is idle or not. In [2], if the backoff counter in the post backoff stage reaches zero and the channel is idle, a station attempts to

transmit a packet. After a successful transmission, regardless of the queue status, it returns to the post backoff stage again. In our Markov model, after a successful transmission from the post backoff stage, it returns to the post backoff stage again if the queue is empty, otherwise it returns to the stage 0.

$$P[(0,k)e/(i,0)] = \frac{(1-p)(1-q)}{W_0}.$$

$$P[(0,k)/(i,0)] = \frac{(1-p)q}{W_0}.$$

$$P[(i+1,k)/(i,0)] = \frac{p}{W_{i+1}}.$$

The final transitions in the post backoff stage is $(0,0)_e$. In this case, we remain in this state if the station's buffer remains empty. If the packet arrives, the state transition moves to the state I. If the medium is busy, the 802.11 MAC begins another stage-0 backoff, with a packet. Otherwise, it moves to the T(transmission) state. A station can transmit a packet in this state. There are two possibilities of transition in this state which are successful transmission and collision. Let P_{idle} be the probability that the medium is sensed idle during a slot.

$$P[(0,k)_e/T] = \frac{(1-p)(1-q)}{W_0}.$$

$$P[(0,k)/T] = \frac{(1-p)q}{W_0}.$$

$$P[(1,k)/T] = \frac{p}{W_1}.$$

$$P[(0,k)/I] = \frac{(1-p_{idle})}{W_0}.$$

All the stationary probabilities in the given Markov model can be expressed as a function of the value $b(0,0)_e$ and the conditional collision probability. The expression for the $b(0,0)_e$ can be obtained from this normalization equation.

$$\sum_{i=0}^{m}\sum_{k=0}^{W_i-1} b(i,k) + \sum_{k=0}^{W_0-1} b(0,k)_e + b(I) + b(T) = 1. \quad (1)$$

$b(i,k)$ and $b(0,k)_e$ are denoting the stationary probability of states in (i,k) and $(0,k)_e$.

We will obtain the closed-form solution for this Markov chain. This will relate the stationary probability b(1,0) with all stationary probabilities b(i,k) for $i \geq 1$ and $0 \leq k \leq W_i-1$. For $1< i<m$, we have $b(i,0) = p^{i-1}\, b(1,0)$. For i=m

$$b(m,0) = \frac{p^{m-1}}{1-p} b(1,0). \quad (2)$$

The collision must have occurred from state (0,0) or an arrival to state T followed by detection of an idle medium and then a collision to reach a Markov chain state $(1,W_1-1)$. So we have

$$b(1,0) = b(0,0)p + b(T)p. \quad (3)$$

$$\sum_{i\geq 1} b(i,0) = \frac{b(1,0)}{(1-p)}. \quad (4)$$

Using (3) and (4) we can express

$$\sum_{i\geq 1} b(i,0) = \frac{p\, b(0,0) + pb(T)}{(1-p)}. \quad (5)$$

Transitions from the state I into the state T occurs if the medium was found to be idle. Transitions from the state T and (i,0) into $(0,W_0-1)$ state occurs if there is no collision and the transmission queue is not empty. Transition into $(0,W_0-1)$ also occurs from the state I if medium is sensed busy.

$$b(0,W_0-1) = b(I)\,\frac{q(1-p)}{W_0}\,P_{idle} + b(I)\,\frac{1-P_{idle}}{W_0} + \sum_{i\geq 0} b(i,0)\,\frac{q(1-p)}{W_0}. \quad (6)$$

Transition into $(0,W_0-1)_e$ can be expressed as

$$b(0,W_0-1)_e = b(T)\frac{(1-p)(1-q)}{W_0} + \frac{(1-p)(1-q)}{W_0}\sum_{i\geq 0} b(i,0). \quad (7)$$

The recursive equation for $b(0,k)_e$ is expressed as

$$b(0,k)_e = (1-q)\, b(0,k+1)_e + b(0,W_0-1)_e. \quad (8)$$

Using (8), the relationship between b(0,0) and $b(0,0)_e$ can be obtained. With $b(0,k)_e$ on the left hand side is replaced by $q\, b(0,0)_e$ if k=0.

$$b(0,0)_e = b(0,0)\frac{1-q}{q}\,\frac{(1-(1-q)^{w_0})}{qW_0 - P_{idle}(1-q)(1-(1-q)^{w_0})}. \quad (9)$$

The above equations help us to find the second sum in (1)

$$\sum_{k=0}^{W_0-1} b(0,k)_e = b(0,0)_e\,\frac{qW_0}{(1-(1-q)^{w_0})}. \quad (10)$$

Similarly the summation of (0,k) chain can be expressed as

$$\sum_{k=0}^{W_0-1} b(0,k) = b(0,0)_e\,\frac{W_0+1}{2}\left(\frac{q^2}{1-q}\,\frac{qW_0}{(1-(1-q)^{w_0})} + q(1-P_{idle})\right) + \left(\frac{qW_0(qW_0+q-2)}{2(1-(1-q)^{w_0})} + 1 - q\right) b(0,0)_e \quad (11)$$

Using (9),we derived the expression for b(1,0) and obtained as

$$b(1,0) = b(0,0)_e\,\frac{pq^2}{1-q}\left(\frac{W_0}{(1-(1-q)^{w_0})}\right). \quad (12)$$

All the stationary probabilities can be expressed in terms of $b(0,0)_e$ that can be derived from (1). The first summation term in (1) can be obtained by using the expression for b(1,0) in (12). We now express the transmission probability τ that a station transmits in a randomly chosen slot time in terms of $b(0,0)_e$.

$$\frac{1}{b(0,0)_e} = 1 + qP_{idle} + \frac{q^2W_0(W_0+1)}{2(1-(1-q)^{w_0})} + \frac{W_0+1}{2}\left(\frac{q^2}{1-q}\,\frac{qW_0}{(1-(1-q)^{w_0})} + q(1-P_{idle})\right) + \frac{pq^2}{2(1-p)(1-q)}\left(\frac{W_0}{(1-(1-q)^{w_0})}\right)\left(2W_0\,\frac{1-p-p(2p)^{m-1}}{1-2p} + 1\right) \quad (13)$$

A station attempts transmission if it is in the state (i,0) for all i or if it is in the T state. Thus,

$$\tau = b(T) + \sum_{i\geq 0} b(i,0) . \quad (14)$$

The above equation is reduced to

$$\tau = b(0,0)_e \left[\frac{q^2 W_0}{(1-p)(1-q)(1-(1-q)^{W_0})} \right]. \quad (15)$$

Now, the transmission probability τ is expressed in terms of p, q, P_{idle}, W_0 and m. By letting $q \to 1$, the model reduces to that of Bianchi model [1]. In this case, the station is in saturation. With the station parameters for each station, the transmission probability and the collision probability can be computed.

B. Finite Load Throughput Model

We consider n identical nodes where every node can sense other nodes' transmission. A collision with a neighboring node occurs only due to any of the (n-1) stations also transmits in the same time slot. In our analysis, we consider the Poisson process for packet arrival in the network. In the Poisson process, the packet arrival rate is λ packets/sec. The packet arrival rate is greater than the MAC service time in the saturation condition. In this network, different traffic loads are used. The stations having same packet arrival rate are grouped together.

Suppose there are k different groups within the network. Let there be n_i stations in group i with arrival rate λ_i. The total number of stations from each group is $\sum_{i=1}^{k} n_i = n$. For k=1, all the stations are in the network are having same arrival rates i.e., n_i=n. The aggregate mean packet arrival rate can be expressed as sum of arrival rates of all the groups. A collision occurs if any of the station in any group is transmitting in the same time slot. Each station in group i transmits with probability τ_i. The collision probability is expressed as

$$p_i = 1 - (1-\tau_i)^{n_i - 1} \prod_{\substack{j=1 \\ j\neq i}}^{k} (1-\tau_j)^{n_j} . \quad (16)$$

The probability of at least one transmission in a time slot is expressed as

$$p_{tr} = 1 - \prod_{j=1}^{k} (1-\tau_j)^{n_j} . \quad (17)$$

Let P_{si} be the probability that the transmission from group i is successful in the same group and P_{sj} be the probability that the transmission from group i is successful in all other groups in the network.

$$P_{si} = n_i \tau_i (1-\tau_i)^{n_i - 1} . \quad (18)$$

$$P_{sj} = \prod_{\substack{j=1 \\ j\neq i}}^{k} (1-\tau_j)^{n_j} \quad (19)$$

Let P_s be the successful probability of packet transmission from any group given that at least one transmission in the network.

$$P_s = \sum_{i=1}^{k} \frac{P_{si}.P_{sj}}{P_{tr}} . \quad (20)$$

The finite load throughput of the DCF can be derived as

$$S = \frac{P_{tr}.P_s.PACKET}{E[slot]} . \quad (21)$$

Where, E[slot] is the average slot time and PACKET is the packet duration. Let σ be the idle slot duration, T_c be the collision time and T_s be the successful transmission time.

$$E[slot] = (1-P_{tr})\sigma + P_{tr}.P_s.T_s + P_{tr}.(1-P_s).T_c$$

This model accurately predicts the finite load throughput of IEEE 802.11 DCF.

C. Modeling offered load and estimation of probability q

In our analysis, we consider Poisson process for packet arrival. The aggregate mean packet arrival rate is denoted by λ and is measured in packets/sec. The time between two packet arrivals is defined as inter-arrival time. To compute the transmission probability τ, we need a probability q [2] which is the probability of having at least one packet to be transmitted in the buffer. The probability q can be expressed as

$$q_i = 1 - e^{-\lambda_i E[slot]} . \quad (22)$$

MAC layer receives a packet from upper layer during the average slot time which can be used to calculate the probability q. The probability for k packet arrivals during the time interval T in the Poisson process is

$$p(k,T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} . \quad (23)$$

The probability for at least one packet in a queue can be obtained as

$$q_i = 1 - p(0, E[slot]) = 1 - e^{-\lambda_i E[slot]} . \quad (24)$$

Where, $p(0, E[slot])$ is the probability for zero packet arrival during the expected time per slot.

D. Cannel Idle Probabilities

It is the probability that the channel is found to be idle at the time a packet arrived in the state $(0,0)_e$ state. Simply it is the probability that the next slot is empty given that our station is not transmitting. The idle probability is expressed as

$$P_{idlei} = (1-\tau_i)^{n_i - 1} \prod_{\substack{j=1 \\ j\neq i}}^{k} (1-\tau_j)^{n_j} = 1 - p_i . \quad (25)$$

Putting together (15), (16), (24) and (25) the nonlinear equations can be solved and the values of τ, p, q and P_{idle} can be obtained.

IV. Model Validation

In our analysis, heterogeneous network considers three different groups that follow different packet arrival rates. Each station in the network has one of the three arrival rates. To validate this model, we have compared the results with that of obtained simulation results. Table 1 lists the values of station parameters used in the theoretical analysis and the simulation. This section

focuses on simulation results for validating the theoretical models and derivations presented in the previous section.

Table 1 Parameter settings used.

Packet payload	364 µsec 500 bytes @ 11Mbps
ACK	304 µsec 192 bits+14 bytes
Slot time	20µsec
SIFS	10µsec
DIFS	50µsec
δ	2 µsec
CW_{min}	32 slots
CW_{max}	1024 slots
Retransmission Limit	5
T_s	944 µsec Header+Payload+SIFS+δ+ACK+δ+DIFS
T_c	944 µsec Header+Payload+SIFS+δ+ACK Timeout

Fig. 2 shows the throughput prediction for a station in each group, with n_1=6, n_2=12 and n_3=24. The predicted and simulated throughput are plotted against normalized arrival rate for a station in each group. Fig. 3 shows the collision probability as a function of packet arrival rate for three different groups. The packet arrival rate of the second group is 1/2 of that of the first group and third group is 1/4 of that of the first group. The predicted throughput closely matches with the simulated throughput.

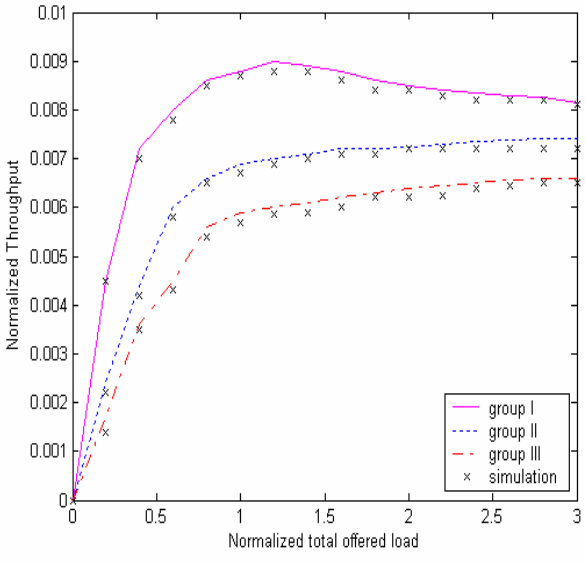


Fig. 2 Normalized per-station throughput versus the normalized offered load.

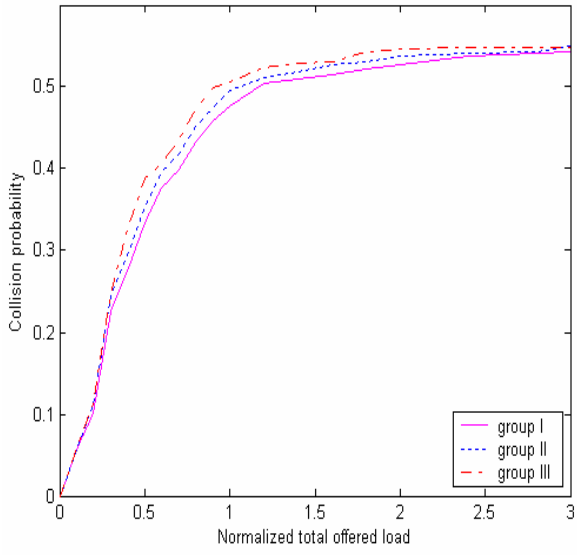


Fig. 3 Collision probability versus the normalized offered load.

When the packet arrival rate increases, the transmission probability τ also increases upto the saturation load and it remains constant beyond the saturation load. Upon comparing the curves in fig. 2 and fig. 3, the collision probability and the throughput remains constant beyond the certain offered load. The collision probabilities of a station in each group are very close because a station in group I sees 5 other group I stations, 12 group II stations and 24 group III stations.

V. Conclusions

In this paper, we have presented an accurate analytical model to evaluate the performance of IEEE 802.11 DCF under nonsaturated and heterogeneous conditions. Simulation and analysis results show that our analytical model can accurately predict the throughput performance of IEEE 802.11 DCF for different packet arrival rates. The predicted result shows that the collision probability of a group which has more number of stations is slightly greater than that of other groups in the heterogeneous network.

References

[1]. G.Bianchi, " Performance Analysis of the IEEE 802.11 Distributed Coordination Function", *IEEE Journal on Selected Area in Communications,* vol.18, no.3, pp. 535-547, March 2000.

[2]. David Malone, Ken Duffy, and Doug Leith ," Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions", *IEEE/ACM Trans. Networking,* vol.15, no.1, pp 159-172 February 2007.

[3]. K. Duffy, D. Malone, D. J. Leith, "Modeling the 802.11 Distributed Coordination Function in Non-Saturated Conditions," *IEEE Communications Letters,* vol. 9, No. 8, August 2005, pp. 715-717.

[4]. Wireless LAN Medium Access Control (MAC) and Physical Layer Specification, IEEE Std. 802.11, 1999.

[5]. F. Alizadeh- Shabdiz, Suresh Subramaniam, "A Finite Load Analytical Model for the IEEE 802.11 Distributed Coordination Function MAC", WiOpt'03, March 2003.

[6]. Y.S. Liaw, A. Dadej, and A.Jayasuriya, " Performance analysis of IEEE 802.11 DCF under limited load", *In Proc. of Asia-Pacific Conference on Communications*, Vol.1, pp.759 - 763, 03-05 Oct. 2005.

[7]. A. N. Zaki and M. T. EI-Hadidi, " Throughput analysis of IEEE 802.11 DCF under finite load traffic," in *Proc. of IEEE-EURASIP ISCCSP,*, Hammamet, Tunisia, Mar. 2004, pp. 535–538.

[8]. P. Chatzimisios and V. Vitsas, " Throughput and delay analysis of IEEE 802.11 protocol, " in IEEE International Workshop on Network Appliances, (IWNA), Liverpool, U.K, oct 2002.

[9]. Y.Barowski , S.Biaz, P.Agrawal, " Towards the Performance Analysis of IEEE 802.11 in Multi-hop Ad-Hoc Networks", I*EEE Wireless Communications and Networking Conference*,Vol. 1, 2005, pp. 100-106.

[10].The network simulator - ns-2. [Online].Available: http:/lwww.isi.edul nsnam/ns/

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Efficient Group Key Agreement Protocol for Ad-Hoc Networks

Rony Hasinur Rahman [1], and M. Lutfar Rahman [2]

Department of Computer Science and Engineering, University of Dhaka [1, 2]
Dhaka – 1000, Bangladesh [1, 2]
E-mail: ronycsdu@yahoo.com [1], lrahman@univdhaka.edu [2]

Abstract - Ad-hoc networks offer communication over a shared wireless channel without any pre-existing infrastructure. Forming security association among a group of nodes in ad-hoc networks is more challenging than in conventional networks due to the lack of central authority. With that view in mind, group key management plays an important building block of any secure group communication. The main contribution of this paper is a low complexity key agreement scheme that is suitable for fully self-organized ad-hoc networks. The protocol is also password authenticated, making it resilient against active attacks. Unlike other existing key agreement protocols, ours make no assumption about the structure of the underlying wireless network, making it suitable for "truly ad-hoc" networks. Finally, we will analyze our protocol to show the computation and communication burden on individual nodes for key establishment.

I. Introduction

Suppose that a group of people at a conference has come together in a room for an ad-hoc meeting. They would like to set up a secure wireless network session with their laptop computers for the duration of the meeting so that no one outside the room can eavesdrop and learn about the contents of the meeting. The people physically present in the room know and trust one another. However, they do not have any a priori means of digitally identifying each other, such as public key certificate authority. An attacker can monitor all traffic on the wireless communication channel and may also attempt to impersonate as a valid member of the group. There is no secure communication channel to connect the computers. The problem is: how can the group set up a secure session among their computers under these circumstances? The network in the scenario described above is an example of an Ad-hoc network in which entities construct a communication network with little or no infrastructural support.

To ensure security, encryption can be used to protect messages exchanged among group members. A vital element of any encryption technique is the *cryptographic key* (also called *group key* in ad-hoc networks). In ad-hoc networks, secure distribution of the group key to all valid participants is a very big issue. In this paper, we have proposed an efficient *group key distribution* (most commonly known as *group key agreement*) protocol which is based on multi-party Diffie-Hellman group key exchange and which is also password-authenticated. The rest of this paper is organized as follows. In the next section, a review of related works is given. In section III, we present the details of the various stages of the proposed protocol. Finally, we discuss some performance issues in section IV and conclude in section V.

II. Related Works

Key agreement in ad-hoc networks is divided into three main classes:

- **Centralized group key agreement protocols (CGKAP)**: A single entity called the Key Distribution Center (KDC) is employed for controlling the whole group.
- **Decentralized group key agreement protocols (DeGKAP)**: The management of a large group is divided among subgroup managers, trying to minimize the problem of concentrating the work in a single place.
- **Distributed group key agreement protocols (DiGKAP)**: There is no explicit KDC, and all the members participate in the generation of the group key and each member contributes to a portion of the key.

A. CGKAP

With only one managing entity, the central server is a single point of failure. The group privacy is dependent on the successful functioning of the single group controller; when the controller is not working, the group becomes vulnerable. Furthermore, the group may become too large to be managed by a single party, thus raising the issue of scalability. The group key agreement protocol used in a centralized system seeks to minimize the requirements of both group members and KDC in order to augment the scalability of the group management. Some popular centralized protocols are: Group Key Management Protocol (GKMP) [1], Logical Key Hierarchy (LKH) [2], One-way Function Tree (OFT) [3], Efficient Large-Group Key (ELK) Protocol [4] etc.

B. DeGKAP

In the decentralized subgroup approach, the large group is split into small subgroups. Different controllers are used

to manage each subgroup, minimizing the problem of concentrating the work on a single place. In this approach, more entities are allowed to fail before the whole group is affected. Scalable Multicast Key Distribution [5], Kronos [6], Intra-Domain Group Key Management (IGKMP) [7], Hydra [8] are some of the popular protocols that follow the decentralized architecture.

C. DiGKAP

The distributed key agreement approach is characterized by having no group controller. The group key can be either generated in a contributory fashion, where all members contribute their own share to computation of the group key, or generated by one member. In the latter case, although it is fault-tolerant, it may not be safe to leave any member to generate new keys since key generation requires secure mechanisms, such as random number generators, that may not be available to all members. Moreover, in most contributory protocols (apart from tree-based approaches), processing time and communication requirements increase linearly in term of the number of members. Additionally, contributory protocols require each user to be aware of the group membership list to make sure that the protocols are robust. Some popular protocols in this category are Burmester and Desmedt (BD) Protocol [9], Group Diffie-Hellman Key Exchange (G-DH) [10], Octopus Protocol [11], Conference Key Agreement (CKA) [12], Diffie-Hellman Logical Key Hierarchy (DH-LKH) [13], Password Authenticated Multi-Party Diffie-Hellman Key Exchange (PAMPDHKE) Protocol [14]. Our proposed protocol falls in this category. We use the following attributes to evaluate the efficiency of DiGKAP:

- *Number of rounds:* The protocol should try to minimize the number of iterations among the members to reduce processing and communication requirements.
- *Number of messages:* The overhead introduced by every message exchanged between members produces unbearable delays as the group grows. Therefore, the protocol should require a minimum number of messages.
- *Number of Exponentiations:* Since exponentiations impose more overhead than additions/multiplications, the number of exponentiations performed by a node should be kept to as low as possible.

III. The Proposed Protocol

The basic idea of the protocol is to securely construct and distribute a secret session key, *K*, among a group of nodes/users who want to communicate among themselves in a secure manner. The group is formed in an ad-hoc fashion and hence no pre-assumption can be made about the overall physical structure of the group.

The proposed protocol starts by constructing a spanning tree on-the-fly involving all the valid nodes in the scenario. It is assumed, like all other protocols, that each node is uniquely addressed and knows all its neighbors (i.e. the protocol runs on top of the network layer and it assumes that a valid route among the nodes have already been constructed by some underlying routing protocol). It is also assumed that each valid member of the scenario shares a password (also called a weak secret) *P*. After that the tree is traversed from bottom-to-top where each node *i*, sends to its parent, its Diffie-Hellman contribution α^{g_i}, where α is a generator of the multiplicative group Z_p^* (i.e. the set {1,2,…p – 1}) and g_i is node *i*'s secret. In this way every contribution ultimately reaches the root of the tree. Then the root creates separate messages for each of its children where each message contains sufficient information so that the child can compute the secret session key *K*. This process continues in a top-to-bottom fashion from every internal node to all its children. In the end, all the valid nodes in the tree contain sufficient information to construct the session key *K*.

A. Construction of the Spanning Tree

Our key agreement protocol functions in an arbitrary rooted tree structure. For this purpose, a spanning tree over the graph has to be constructed first. This can be done in several ways. Below we will describe one possible protocol for constructing a spanning tree where the node initiating the protocol becomes the root. The tree is indexed using universal addresses. In the initial state it is assumed that the nodes know their neighbors. The initiator sends a message to each of its neighbors. It thereby becomes the root of the tree and the neighbors become its children. After receiving a message, a node acknowledges it and sends a similar message to all its neighbors, except to the parent. The nodes that acknowledge a message from a node become its children in the tree. If a node gets more than one of these messages, it acknowledges and processes only the message that it receives first. Consequent messages are ignored. This continues until every node has received this kind of a message. A leaf is a node that does not receive acknowledgements from any of its neighbors. The initial network is shown in Fig. 1 and the spanning tree constructed by applying the above protocol is shown in Fig. 2. It should be noted that the structure of the final spanning tree might be different based upon the order in which messages are received by each individual node.

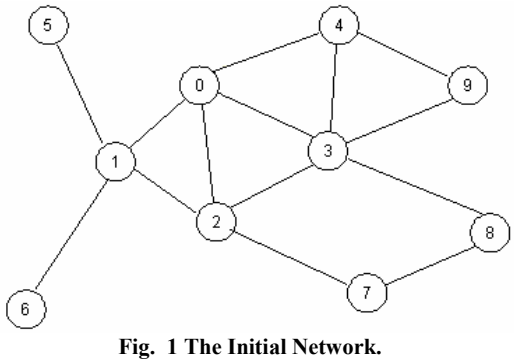


Fig. 1 The Initial Network.

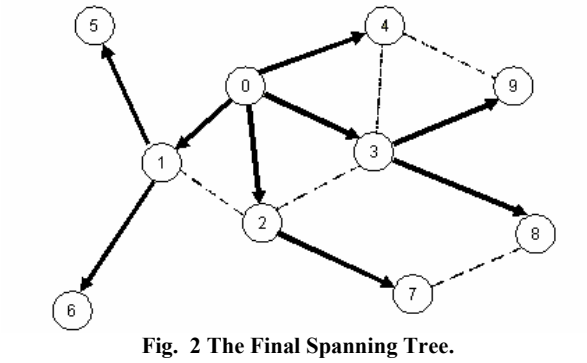


Fig. 2 The Final Spanning Tree.

B. Phase I of the Protocol

Let Z_p^* (i.e. the set {1,2,…p – 1}) be a finite multiplicative group where p is a prime and let α be the generator of the group. A participant/node, *i*, is assumed to pick his/her secret exponent g_i randomly where $1 \le g_i \le p-1$. The steps of phase I are described below:

- Every internal node gets the contributions from all its children.
- Each node generates its own contribution and multiplies all its descendants' contributions with its own.
- If node *i* **is not** the root, then it executes this step. Node *i* sends the product obtained from step (2) to its parent along with its own contribution and all the other contributions (of *i*'s descendant nodes) that was forwarded to *i* from its immediate children.
- If node *i* **is** the root, then it executes this step. The product obtained from step (2) represents the final group session key *K*. Phase I stops here.

Formally,

- In phase I, each node *x* sends a message M = {M_1, M_2, M_3} (having 3 parts) to its parent *y*, where M_1 = the product of *x*'s contribution and all contributions from the descendants of *x*, M_2 = *x*'s own contribution and M_3 = all contributions from the descendants of *x*.
- If *x* is a leaf node with contribution α^{g_x} and *y* is its parent, then, *x* sends to *y* a message containing the quantity α^{g_x}, i.e.

  $x \to y$: $\{\alpha^{g_x}, \alpha^{g_x}, -\}$.
- If *x* is an internal non-root node with contribution α^{g_x}, *y* is its parent and *a*, *b*, *c* etc. are its children (with contributions α^{g_a}, α^{g_b}, α^{g_c} etc. respectively) and if *x* receives messages M_a, M_b, M_c etc. from *a*, *b*, *c* etc. respectively, then, *x* sends to *y* a message containing the quantity $\{\alpha^{g_x}.A.B.C.D, \alpha^{g_x}, M'\}$, i.e.

  $x \to y$: $\{\alpha^{g_x}.A.B.C.D, \alpha^{g_x}, M'\}$, where,

  $M' = \{\{M_2 \text{ part of } M_a\} \cup \{M_3 \text{ part of } M_a\} \cup \{M_2 \text{ part of } M_b\} \cup \{M_3 \text{ part of } M_b\} \cup \{M_2 \text{ part of } M_c\} \cup \{M_3 \text{ part of } M_c\} \cup \{M_2 \text{ part of etc.}\} \cup \{M_3 \text{ part of etc.}\}\}$

  $A = M_1 \text{ part of } M_a; B = M_1 \text{ part of } M_b; C = M_1 \text{ part of } M_c;$
  $D = M_1 \text{ part of etc.}$
- If *x* is the root node with contribution α^{g_x} and *a*, *b*, *c* etc. are its children (with contributions α^{g_a}, α^{g_b}, α^{g_c} etc. respectively) and if *x* receives messages M_a, M_b, M_c etc. from *a*, *b*, *c* etc. respectively, then, *x* computes the final session key *K*, $K = \alpha^{g_x}.A.B.C.D$, where,

  $A = M_1 \text{ part of } M_a; B = M_1 \text{ part of } M_b; C = M_1 \text{ part of } M_c;$
  $D = M_1 \text{ part of etc.}$
- The quantity $\alpha^{g_x}.A.B.C.D$ indicates the product of α^{g_x}, *A*, *B*, *C* and *D*.

C. Phase II of the Protocol

Before the beginning of phase II, first, the root takes the union of all the M_2 parts and all the M_3 parts of all the messages it has received from its immediate children. Then it raises each quantity of this newly formed set by its own secret exponent. The steps of phase II are described below:

- Every internal node *x* sends to its child *i* sufficient information needed by *i* to construct the session key *K*. The node *x* also sends to *i* a quantity encrypted by *K* for authentication purpose and forwards sufficient information so that descendants of *i* may successfully construct the session key *K*.
- When every leaf node gets messages from its parent, phase II stops. Every *valid* node now has the session key *K* and has been authenticated.

Formally,

- In phase II, each internal node *x* sends a message M_i^* = {$P(\overline{M_1})$, $\overline{M_2}$, $\overline{M_3}$, $\overline{M_4}$} (having 4 parts) to each of its child *i*, where $\overline{M_1}$ = *i*'s contribution raised to the power of root's secret exponent, $\overline{M_2}$ = all contributions from all other nodes except the root and the descendants of *i*, $\overline{M_3}$ = all contributions of the descendants of *i* raised to power of the root's secret exponents and $\overline{M_4}$ = *K*(n) = a quantity encrypted with the session key *K* needed for authentication.

- If x is an internal node with contribution α^{g_x}, y is its parent (may be null if x is the root) and a, b, c etc. are its children (with contributions α^{g_a}, α^{g_b}, α^{g_c} etc. respectively) and if x receives messages M_a, M_b, M_c etc. from a, b, c etc. respectively and creates the message M in phase I and receives M_x^* from y, then, x sends to its child $i \in \{a,b,c,etc.\}$ a message M_i^* containing the quantity $\{P(\alpha^{g_i g_{root}}), G_i, H_i, K(n_x)\}$, i.e.

  $x \rightarrow y$: $\{P(\alpha^{g_i g_{root}}), G_i, H_i, K(n_x)\}$ for $i \in \{a,b,c,etc\}$

  where,

  $\alpha^{g_i g_{root}}$ is obtained from $\overline{M_3}$ part of M_x^* or is already present in x (if x = root),

  $G_i = \{\{\overline{M_2} \text{ part of } M_x^*\} \cup \{\alpha^{g_x}\} \cup \{M_1 \text{ part of } M_j\}\}$

  $where\ j \in \{a,b,c,etc.\}\ \&\ j \neq i\ [\{\alpha^{g_x}\} = \phi\ if\ x = root]$,

  $H_i = \{k^{g_{root}}\}\ where\ \ k \in \{M_3\ part\ of\ M_i\}$,

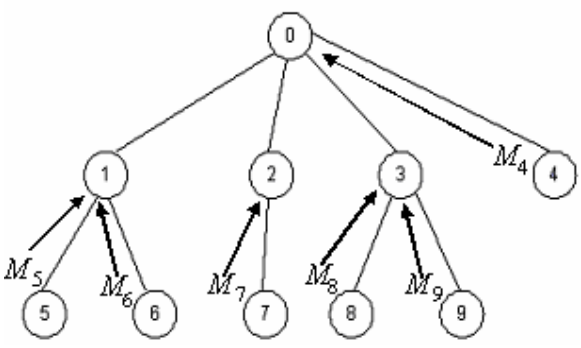
  $K(n_x) = ID\ of\ x\ encrypted\ \ by\ session\ key\ K$.

- If a non-root node x receives the message M_x^* and also creates the message M in phase I, then, it calculates the key K as follows: (a) It first decrypts $\overline{M_1}$ with the weak password P to retrieve $L = \alpha^{g_x g_{root}}$, (b) Then it retrieves $\alpha^{g_{root}}$ by performing $L^{\frac{1}{g_x}}$ and (c) $K = \alpha^{g_{root}}.s. < M_1\ part of\ M >, \forall s \in \{\overline{M_2}\ part of\ M_x^*\}$.

- So now all the nodes have the key $K = \alpha^{\sum g_i}$, where i is a node of the network and g_i is its secret exponent.

- After that, each node decrypts $\overline{M_4}$ with K and verifies whether the quantity is the identity of its parent. This step authenticates the parent to all of its children.

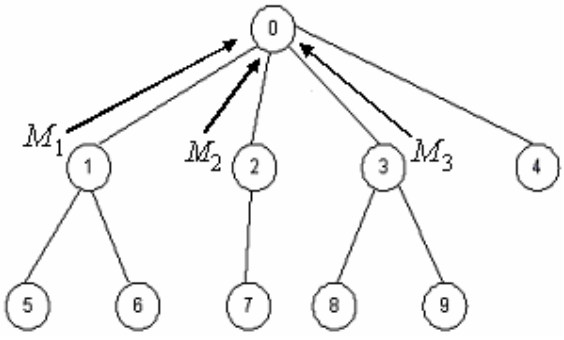
Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7 demonstrate the aforementioned phases I and phase II of the proposed protocol.



$M_5 = \{\alpha^{g_5}, \alpha^{g_5}, -\}; M_6 = \{\alpha^{g_6}, \alpha^{g_6}, -\}; M_7 = \{\alpha^{g_7}, \alpha^{g_7}, -\}$

$M_8 = \{\alpha^{g_8}, \alpha^{g_8}, -\}; M_9 = \{\alpha^{g_9}, \alpha^{g_9}, -\}; M_4 = \{\alpha^{g_4}, \alpha^{g_4}, -\}$

Fig. 3 Phase I, Round 1.

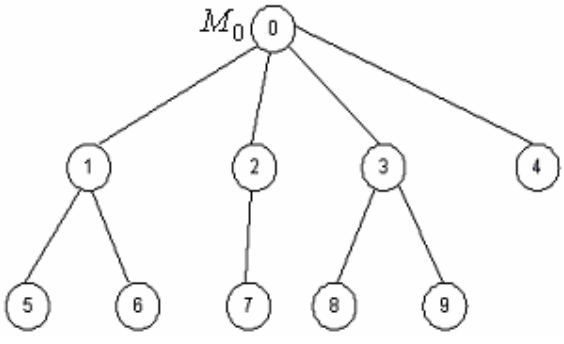


$M_1 = \{\alpha^{g_1+g_5+g_6}, \alpha^{g_1}, \{\alpha^{g_5}, \alpha^{g_6}\}\}$

$M_2 = \{\alpha^{g_2+g_7}, \alpha^{g_2}, \{\alpha^{g_7}\}\}$

$M_3 = \{\alpha^{g_3+g_8+g_9}, \alpha^{g_3}, \{\alpha^{g_8}, \alpha^{g_9}\}\}$

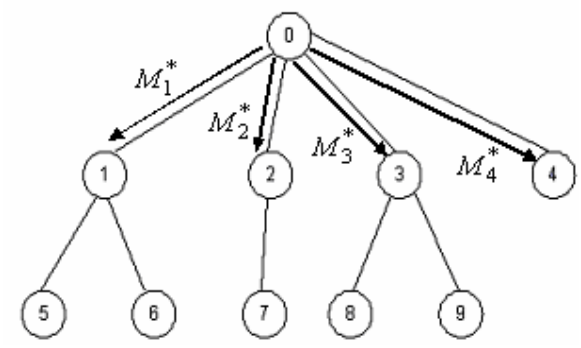
Fig. 4 Phase I, Round 2.



$M_0 = \{\alpha^{g_0+g_1+g_5+g_6+g_2+g_7+g_3+g_8+g_9+g_4}, \alpha^{g_0},$

$\{\{\alpha^{g_1}, \alpha^{g_5}, \alpha^{g_6}\}, \{\alpha^{g_2}, \alpha^{g_7}\}, \{\alpha^{g_3}, \alpha^{g_8}, \alpha^{g_9}\}, \alpha^{g_4}\}\}$

Fig. 5 End of Phase I and Start of Phase II.



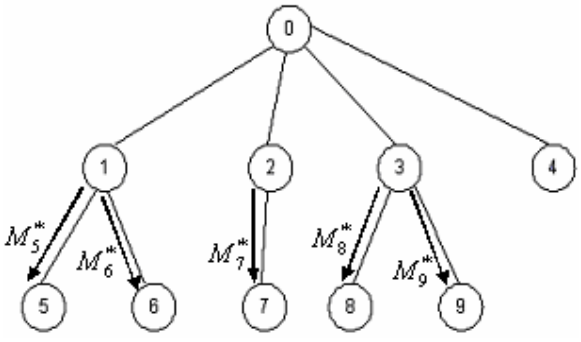
$$M_1^* = \{P(\alpha^{g_0 g_1}), \{\alpha^{g_2+g_7}, \alpha^{g_3+g_8+g_9}, \alpha^{g_4}\}, \{\alpha^{g_0 g_5}, \alpha^{g_0 g_6}\}, K(ID_0)\}$$

$$M_2^* = \{P(\alpha^{g_0 g_2}), \{\alpha^{g_1+g_5+g_6}, \alpha^{g_3+g_8+g_9}, \alpha^{g_4}\}, \{\alpha^{g_0 g_7}\}, K(ID_0)\}$$

$$M_3^* = \{P(\alpha^{g_0 g_3}), \{\alpha^{g_1+g_5+g_6}, \alpha^{g_2+g_7}, \alpha^{g_4}\}, \{\alpha^{g_0 g_8}, \alpha^{g_0 g_9}\}, K(ID_0)\}$$

$$M_4^* = \{P(\alpha^{g_0 g_4}), \{\alpha^{g_1+g_5+g_6}, \alpha^{g_2+g_7}, \alpha^{g_3+g_8+g_9}\}, -, K(ID_0)\}$$

Fig. 6 Phase II, Round 1.



$$M_5^* = \{P(\alpha^{g_0 g_5}), \{\alpha^{g_2+g_7}, \alpha^{g_3+g_8+g_9}, \alpha^{g_4}, \alpha^{g_1}, \alpha^{g_6}\}, -, K(ID_1)\}$$

$$M_6^* = \{P(\alpha^{g_0 g_6}), \{\alpha^{g_2+g_7}, \alpha^{g_3+g_8+g_9}, \alpha^{g_4}, \alpha^{g_1}, \alpha^{g_5}\}, -, K(ID_1)\}$$

$$M_7^* = \{P(\alpha^{g_0 g_7}), \{\alpha^{g_1+g_5+g_6}, \alpha^{g_3+g_8+g_9}, \alpha^{g_4}, \alpha^{g_2}\}, -, K(ID_2)\}$$

$$M_8^* = \{P(\alpha^{g_0 g_8}), \{\alpha^{g_1+g_5+g_6}, \alpha^{g_2+g_7}, \alpha^{g_4}, \alpha^{g_3}, \alpha^{g_9}\}, -, K(ID_3)\}$$

$$M_9^* = \{P(\alpha^{g_0 g_9}), \{\alpha^{g_1+g_5+g_6}, \alpha^{g_2+g_7}, \alpha^{g_4}, \alpha^{g_3}, \alpha^{g_8}\}, -, K(ID_3)\}$$

Fig. 7 Phase II, Round 2.

IV. Performance Evaluation

A. Security

In our protocol, it is assumed that a weak secret/password P is shared among the valid users/nodes. This P helps in the authentication process and prevents *man-in-the-middle* attack. This assumption is not at all inappropriate. The ad-hoc scenarios that were mentioned in the beginning of this paper (people coming together in a conference or a military troop deployed in an hostile environment) indicate that the people involved in those scenarios trust each other. So it is possible for them to decide on a simple password because they will definitely come in contact with each other before forming the actual ad-hoc network. That password may be written down on piece of paper and circulated to all the trusted parties. The recipients can then enter the password in his/her computer and use it as the weak shared secret for the protocol described in this paper. It may be noted that this password will never be used to encrypt data traffic. It will merely help to authenticate the nodes. Then the paper may be destroyed to remove all physical existence of the password.

It is very obvious from the example given in the previous section, that, every valid node has necessary and sufficient information to construct the session secret key K, which will be the group key for that session. Now it remains to show that a passive adversary as well as an active adversary will never be able to construct K from the messages that travel through the wireless network. First of all, it is very clear that the secret exponent g_x for some node x, is never exposed to the network. To construct K, an adversary needs the contribution α^{g_x} of each valid node x. Each contribution α^{g_x} of each valid non-root node x, is passed through the network in plaintext. But one can see, very obviously, from phase II of the protocol that the contribution of the root, $\alpha^{g_{root}}$, is never sent into the network by itself, i.e. it is always sent in the form $\alpha^{g_{root} g_x}$, where g_x is the secret exponent of a valid non-root node x. Without knowing g_x, no one (not even another valid node y) can obtain $\alpha^{g_{root}}$ from $\alpha^{g_{root} g_x}$. Since g_x is never exposed to the network by x, only x can extract $\alpha^{g_{root}}$ from $\alpha^{g_{root} g_x}$. The only way that an adversary (active or passive) can get a hold of g_x from x, is to hack into node x and compromise it. And if a node is compromised by an adversary, then any session key, past, present or future, can always be obtained by that adversary. This is true for all existing security protocols (both in wired and wireless networks). Dealing with such events is beyond the scope of this paper. Moreover, the problem of calculating $\alpha^{g_{root}}$ by using $\alpha^{g_{root} g_x}$ and $\alpha^{g_{root}}$ is a Decision Diffie-Hellman Problem (DDHP) which is intractable. In a nutshell, unless a valid node is compromised, an adversary (passive or active) will never be able to construct any session key K by observing/obtaining the messages of phase I and phase II of the protocol.

The second line of defense is the weak shared secret P. P is only used to encrypt the first part of a message in phase II. This is used to prevent active adversaries from carrying out man-in-the-middle attack. The adversary does not know P. So if it tries to mislead a valid node x by sending dummy or meaningless messages or by impersonating as a valid node, it will always fail because it does not have the capability to encrypt the first part of a message in phase II. So when a valid node gets such a message in phase II, and if it can not decrypt the first part of that message, it will immediately know that the message has come from an adversary. Then the valid node can take appropriate actions. So, although there may be a delay, ultimately the active adversary will definitely be caught by a valid node in some round of phase II. One thing needs to be mentioned that our protocol is not free from DoS (Denial of Service) attacks. An active adversary can disrupt or delay the protocol by sending huge amount of junk messages for a period of time.

B. Efficiency

As the topology of the spanning tree is arbitrary, the number of children is not limited and exact figures are not always possible. We estimate the figures by assuming that the tree is a balanced perfect k-ary tree having total of *n* nodes. In phase I, every node except the root sends a message. This makes the number of messages in phase I *n-1*. In phase II, every node except the root receives a message. So the total number of messages in the protocol is *2n-1*. Broadcast/multicast is a serious bottleneck for wireless networks. Unlike many other protocols, ours does not need broadcast/multicast capability. Our protocol uses Diffie-Hellman key shares, and hence it is contributory. The protocol needs no explicit leader election technique. Anyone wishing to form a secure group can start constructing the spanning tree and thereby can become the root of tree. And the root implicitly becomes the leader of the group. If 2 or more nodes start to construct the tree simultaneously, only the messages originating from the root node with the lowest ID (highest priority) will be considered by all other nodes. Messages originating from other root nodes will be discarded. In both phases of the protocol, the number of rounds/iterations needed is $O(\log_k n)$. In phase I, each member generates its own contribution and so the total number of exponentiations in this phase is *n*. At the start of phase II, the root raises each member's contribution to the power of its own exponent, i.e. performs *n* exponentiations. In the remaining rounds of phase II, each non-root node performs one exponentiation to retrieve the root's contribution. So the total number of exponentiations in the protocol is $3n-1 = O(n)$. So by amortized analysis, the number of exponentiations performed (on an average) by a single node is $O(n) / n = O(1)$. Finally, during the construction of the spanning tree, each node can send at most *n* request messages and at most *1* reply message. So the total number of messages during this stage is $O(n^2)$. This quantity may seem high but it is a price one must pay if one wants this protocol (or any other protocol) to work under any circumstances. But unfortunately none of the other existing protocols take this issue under consideration. More or less all of the existing protocols pre-assume some sort of infrastructure among the nodes and, hence, are not suited for *"truly ad-hoc"* (fully infrastructure less) environment. On the other hand, our protocol has the capability to work under and adapt to any *"truly ad-hoc"* environment. Table 1 compares the proposed protocol with some popular ones in the same category. Here, *n* is the number of nodes in the network.

Table 1 Comparative Study.

Protocol	No. of rounds	No. of messages	
		Multicast	Unicast
BD	3	$2n$	0
G-DH	n	n	n - 1
DH-LKH	$\log_2 n$	$\log_2 n$	0
PAMPDHKE	n + 2	2	$3n$ - 3
Our Protocol	$2\log_k n$	0	$2n$ - 1

Although BD has the best performance with respect to number of rounds, it requires a large number of multicast messages. Ours is the only protocol having zero multicast messages while keeping number of rounds tight up to a logarithmic factor. Unlike BD and DH-LKH, where the number of unicast messages is zero, ours has a linear upper bound. Still this is better because both BD and DH-LKH have non-zero multicast messages where each multicast message can be considered equivalent to maximum of *n* unicast messages.

V. Conclusion

In this paper we have proposed a new group key agreement protocol suitable for ad-hoc networks of arbitrary topology. Several lines of future work are possible. Formal security analysis is a missing step. Moreover, effects of changes in the physical topology of the group have to be addressed. We conclude that, the issue of group key agreement has to be addressed more extensively to achieve security in ad-hoc networks.

References

[1] H. Harney and C. Muckenhirn, "Group Key Management Protocol (GKMP) Specification", RFC 2093, 1997.

[2] D. Wallner, E. Harder, and R. Agee, "Key Management for Multicast: Issues and Architectures", RFC 2627, 1999.

[3] D. A. McGrew and A. T. Sherman, "Key establishment in large dynamic groups using one-way function trees", Tech. Rep. No. 0755 (May), TIS Labs at Network Associates, Inc., Glenwood, Md, 1998.

[4] A. Perrig, D. Song, and J. D. Tygar. "ELK, a new protocol for Efficient Large-group Key distribution". IEEE Security and Privacy Symposium, May 2001.

[5] A. Ballardie, "Scalable Multicast Key Distribution", RFC 1949, 1996.

[6] S. Setia, S. Koussih, S. Jajodia, and E. Harder, "Kronos: A scalable group re-keying approach for secure multicast", IEEE Symposium on Security and Privacy, May 2000.

[7] B. DeCleene, L. Dondeti, S. Griffin, T. Hardjono, D. Kiwior, J. Kurose, D. Towsley, S. Vasudevan, and C. Zhang, "Secure group communications for wireless networks", MILCOM, June 2001.

[8] S. Rafaeli and D. Hutchison, "Hydra: a decentralized group key management", in Proc. 11th IEEE International WETICE: Enterprise Security Workshop, June 2002.

[9] M. Burmester and Y. Desmedt, "A secure and efficient conference key distribution system", EUROCRYP'94, LNCS (950): 275–286, 1994.

[10] M. Steiner, G. Tsudik, and M. Waidner, "Diffie-Hellman key distribution extended to group communication", in Proc. 3rd ACM Conference on Computer and Communications Security, pp. 31–37, March 1996.

[11] C. Becker and U. Wille, "Communication complexity of group key distribution", in Proc. 5th ACM Conference on Computer and Communications Security, November 1998.

[12] C. Boyd, "On key agreement and conference key agreement", in Proc. Information Security and Privacy: Australasian Conference, LNCS (1270): 294–302, 1997.

[13] Y. Kim, A. Perrig, and G. Tsudik, "Simple and fault-tolerant Key Agreement for Dynamic Collaborative groups", in Proc. 7th ACM Conference on Computer and Communications Security, November 2000.

[14] N. Asokan and P. Ginzboorg, "Key Agreement in ad hoc networks", Elsevier Journal of Computer Communications. Vol. 23, pp. 1627 – 1637, 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Analysis of Relay Selection Methods for IEEE802.16j

Najmeh Forouzandeh mehr , Hossein Khoshbin

School of Electrical Engineering
Ferdowsi University of Mashhad, Iran
E-mail:na_fo44@stu-mail.um.ac.ir

Abstract- The IEEE 802.16e standard serves as the backhaul of broadband wireless access in the emerging fourth-generation mobile networks. Deploying relay stations as defined in IEEE 802.16j enhances the coverage area and throughput of the IEEE 802.16e. In this paper, we study the problem of relay selection for a wireless cooperative network in context of IEEE 802.16j emerging standard. We propose two selection methods: Best relay selection method, which selects optimal relay station for each subchannel and Nearest relay selection method, which also is performed in pre-subchannel mode. We provide analytical approximations of these schemes performance in terms of outage probability .Both theoretical and simulation results show that the best relay selection method can achieve significant gain compared to the nearest relay selection method.

***Index Terms*—Relay selection, cooperative communications, IEEE 802.16j, outage probability**

I. Introduction

The WiMAX is a broadband wireless technology that supports fixed, nomadic, portable and mobile access. To support portability and mobility, IEEE 802.16e amendment has been defined to this standard. IEEE 802.16e has adopted SOFDMA (Scalable OFDMA) which allows for a variable number of carriers, in addition to the previously-defined OFDM and OFDMA modes [1].

The emerging IEEE 802.16j Mobile Multi-hop Relay (MMR) is currently being developed for increasing the coverage area and throughput of the IEEE 802.16e standard via the deployment of fixed or nomadic relay terminals. It also takes advantages of the less complexity and lower cost of relay stations (RSs) [2].

Cooperative transmission can be seen as a "virtual" MIMO system, where multiple transmit antennas are in fact implemented distributed by antennas both at the source and the relay terminal. Cooperation among users in a wireless network can substantially improves the robustness to fading due to the diversity gain which, in turn, can translate into an increase in the achieved transmission range, the achievable data rate, or reliability [3]. Therefore, the cooperative transmission techniques are very promising for IEEE 802.16j network [4]. The performance of various cooperative diversity schemes to choose the best relay among a collection of available relays in a two-hop cellular network has been analyzed for instance, relay selection protocol based on geographic positions of terminals [5] and on the instantaneous values of the source-destination and source-relay channels gain [6].

In this paper, we propose relay selection methods in the context of IEEE 802.16j standard using opportunistic relaying in per sub channel basis. In particular, each sub channel chooses the best relay independently. Performance of the proposed relaying scheme will be evaluated in term of outage behaviour and compared to some other schemes like Nearest relay selection. It is shown that the proposed relay selection has much better performance that other considered schemes. Simulation results validate the analysis very well.

This paper is organized as follows. In Section II, we present the system model and the cooperative relaying Protocol. In Section III, we describe the details of our proposed relay selection and other relay selection schemes. In Section IV, we analyze these methods and provide a lower bound to outage probability of these schemes. . In Section V, we provide simulation results and compare them with the analytical approximation. Section VI provides our conclusion and final remarks.

II. System Model

We consider an IEEE 802.16j based relay network with two hops; network in each hop is similar to IEEE802.16e based cellular network. A single cell with multiple users is adopted so that the BS is at the center of cell and users are uniformly located within the cell. We assumed users are low mobility, each mobile station is equipped with single antenna and its transmission is constrained to half- duplex mode. It is also assumed that the best placement for relay stations is pre-determined and the relaying scenario is done in downlink where base station (BS) transmits information to mobile stations (MSs).

For direct transmission, the received signal at subcarrier n of destination d from source user s can be modelled as:

$$y_{s,d}(n) = \sqrt{PD_{s,d}^{-v}} H_{s,d}(n) x_s(n) + z_{s,d}(n), \qquad (1)$$

where P is the transmitted power in each subcarrier. $x_s(n)$ is the source transmitted signal , $y_{s,d}(n)$ is the received signal at destination and $z_{s,d}(n)$ is an additive white Gaussian noise with zero mean and variance N_0 . We can express the channel coefficient $H_{s,d}(n)$, the frequency response of the channel evaluated at the n^{th} subcarrier as $H_{s,d}(n) = \sum_{l=0}^{L-1} \alpha_{s,d}(l) \exp\left(-\frac{j2\pi n \tau_{s,d}(l)}{N}\right)$, L is the number of independent delay paths, $\tau_{s,d}(l)$ is the delay of l^{th} path and N is number Of subcarriers. For the $S \rightarrow R$, $R \rightarrow D$ and $S \rightarrow D$ links , the wireless NLOS channel model presented in [7],[8] is used with a path-loss exponent of 3. $\alpha_{s,d}(l)$, L and $\tau_{s,d}(l)$ are defined from this channel model. The observed SNR at destination node d obtained by transmission from source node s can be written as:

$$SNR_{s,d}^{(n)} = SNRD_{s,d}^{-v} \left|H_{s,d}(n)\right|^2 \qquad (2)$$

For Rayleigh fading, $\left|H_{s,d}(n)\right|^2$ has an exponential distribution with parameter $\sigma_{s,d}^{-2}$, where $\sigma_{s,d}^2$ is the variance of channel gains. Throughout this paper, we characterize performance of the system in terms of outage probability. Outage probability is the probability that the instantaneous achievable rate of the channel is less than the target rate R in other words, $P_{out} = P(I < R)$, where I is the mutual information and for an OFDM symbol and it is equal to $I = \frac{1}{N}\sum_{n=0}^{N-1} I^{(n)}$. $I^{(n)}$ is the random element indicating the mutual information for the n^{th} OFDM subcarrier .For direct transmission and with independent and identically distributed Zero-mean circularly symmetric complex Gaussian inputs, it can be shown as $I^{(n)} = \log\left(1 + SNR_{s,d}^{(n)}\right)$. A two-phase transmission as shown in Fig.1, is adopted as cooperative relaying format. In the first phase, the base station transmit data packet for the pre-selected relay, in the second phase both BS and RS transmit simultaneously to the MS. We considered cooperative transmit diversity relaying method based on the use of transmit diversity using STBCs [4]. In this method, each signal source playing the role of different transmit antenna in the conventional STBC. The STBC encoding is performed at the RSs, so the channel utilization is more efficient because the MMR-BS needs to transmit the packet only once. BS and RS use space- time coding in the form of Alamouti scheme [9]. After the two phases end, the MS space-time decodes.

III. Relay Selection Methods

In this section we describe the details of our proposed relay selection methods. Particularly, Best relay selection and Nearest relay selection method. In both of these schemes, one or two relays can be selected in the second phase. If one relay is selected, as shown in fig.1 (a) half of space-time coding process is done at the BS and another half is done as shown in fig.1 (b) at the selected RS. In the case of two relay selection, space-time coding is not performed at the BS instead half of coding process is done at one of selected RSs and another half is performed at the other RS.

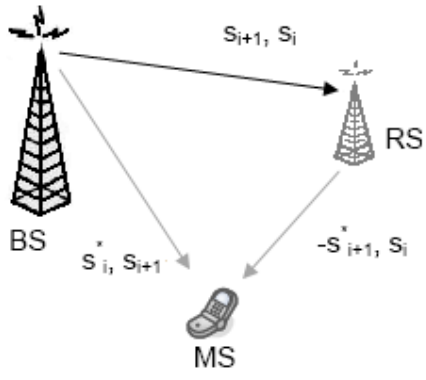


Fig.1 (a) Partial encoding in BS and RS when just one RS is selected

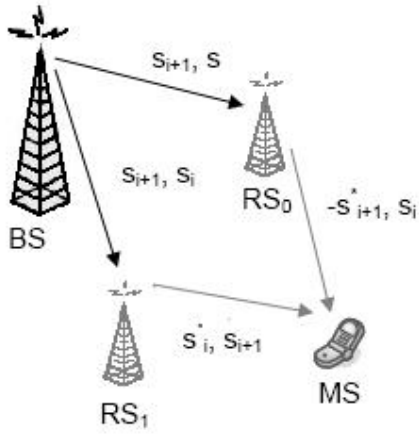


Fig.1 (b) Partial encoding in BS and RSs when two relays are selected

A. Best Relay Selection Method

The proposed relay selection scenario is performed in a per-subchannel manner. In practice the allocation in the frequency domain is not addressed at the level of subcarriers .Typically, subchannels which are the smallest granular units in the allocation are created by grouping subcarriers in an OFDM symbol in various ways. The formation of these subchannels from subcarriers is important concept in OFDMA systems .the formation can be classified in two types; one is mapping Contiguous group of subcarriers into a subchannel called as adjacent subcarrier mode (ASM) and the other is the permutation based grouping called as diversity subcarrier mode (DSM) [10]. Each sub-channel can be modeled as a flat fading channel with approximately equal SNR level. In 802.16e both ASM and DSM based subchannelization have been defined. The BS choose to use the best RS in the first phase .Such decision is done for each subchannel so that, the best relay is chosen prior to BS transmission. It's needed for each RS to know the received SNR at RS and MS ($SNR_{s,r}^{(n)}$, $SNR_{r,d}^{(n)}$) for relaying the nth subchannel .

The best relay for transmission of each subchannel is selected to maximize the minimum of the received SNR at RS and MS for all available RSs.

B. Nearest Relay Selection Method

In the first phase of this relaying scheme, each RS is assigned to the mobile station (MS) with the shortest distance to it then the selected RS listens to BS's transition .This scheme requires infrastructures for knowledge or estimation of distances.

C. Two Best Relay Selection Method

This relaying method is based on the best relay selection is explained above, but two best relays are chosen in the first phase. That is to say, after minimums of received SNR at RS and MS pair for all RSs , min ($SNR_{s,r_k}^{(n)}$, $SNR_{r_k,d}^{(n)}$) , k=1,⋯,K , are sorted in decreasing order , simply relay stations which are assigned to first and second values are chosen. Then, the BS transmits its data to them . Moreover, in the second phase these two RSs transmit for the MS. Like the previous cases relaying in done in per subchannel manner.

D. Two Nearest Relay Selection Method

This relaying scheme is similar to Nearest relaying selection .In the first phase the BS transmit to two nearest RSs and in the second phase these two RSs relay to the MS .

IV. Outage Analysis

In this section, we first analyze the outage probability of the proposed cooperative protocol. For the best relay selection scenario, the mutual information between BS and MS can be shown to be

$$I^{(n)} = \frac{1}{2}\log\left(1 + SNR_{s,d}^{(n)} + SNR_{best}^{(n)}\right) \qquad (3)$$

$$SNR_{best}^{n,k} = \arg\max_{k} \min\left(SNR_{s,r_k}^{(n)}, SNR_{r_k,d}^{(n)}\right)$$
$$SNR^{*} = \min\left(SNR_{s,r_k}^{(n)}, SNR_{r_k,d}^{(n)}\right)$$

Where r_k is k^{th} relay station, $k = 1,\cdots,K$ the outage probability is bounded as follows:

$$\begin{aligned} P(I^{(n)} < R) &= P(SNR_{s,d}^{(n)} + SNR_{best}^{n,k} < 2^{2R} - 1) \\ &= P(SNR_{s,d}^{(n)} + \max(SNR^{*}) < 2^{2R} - 1) \\ &> P(\max\left(SNR_{s,d}^{(n)} + SNR^{*}\right) < 2^{2R} - 1) \\ &= \prod_{k} P(SNR_{s,d}^{(n)} + SNR^{*} < 2^{2R} - 1) \end{aligned} \qquad (4)$$

Since SNR^{*} is the minimum of two independent exponential random variables, it is exponential random variable with parameter $\sigma_{*}^{-2} = \left(\sigma_{s,r_k}^{-2} + \sigma_{r_k,d}^{-2}\right)$. So $(SNR_{s,d}^{(n)} + SNR^{*})$ is the sum of two independent exponential random variables with distinct parameters and its c.d.f is obtained from theorem 2 in [6] .considering above notifications, the lower bound for outage probability yields:

$$P\left(I^{(n)} < R\right) > \prod_{k}\left(1 - \left\{\begin{array}{l}\left(1 - \frac{\sigma_{*}^{-2}}{\sigma_{s,d}^{-2}}\right)^{-1} e^{-(2^{2R}-1)/\sigma_{s,d}^{-2}} + \\ \left(1 - \frac{\sigma_{s,d}^{-2}}{\sigma_{*}^{-2}}\right)^{-1} e^{-(2^{2R}-1)/\sigma_{*}^{-2}}\end{array}\right\}\right) \qquad (5)$$

The mutual information between BS and MS for nearest relay selection can be written as:

$$I^{(n)} = \frac{1}{2}\log\left(1 + SNR_{s,d}^{(n)} + SNR_{r_{nearest},d}^{(n)}\right) \qquad (6)$$

And the outage probability is given by:

$$P(I^{(n)} < R) = P(SNR_{s,d}^{(n)} + SNR_{best}^{(n)} < 2^{2R} - 1)$$

$$= 1 - \left\{\begin{array}{l}\left(1 - \frac{\sigma_{r_{nearest},d}^{-2}}{\sigma_{s,d}^{-2}}\right)^{-1} e^{-(2^{2R}-1)/\sigma_{s,d}^{-2}} + \\ \left(1 - \frac{\sigma_{s,d}^{-2}}{\sigma_{r_{nearest},d}^{-2}}\right)^{-1} e^{-(2^{2R}-1)/\sigma_{r_{nearest},d}^{-2}}\end{array}\right\}$$

V. Simulation Results

Each OFDMA symbol has 1024 subcarriers which are partitioned into subchannels containing 24 data subcarriers. System bandwidth and carrier frequency are 10MHz and 2.5 GHz, respectively. The cell is equally divided into 3 sectors. In each sector, 4 RSs are placed, with 0.6 x cell radius distance to BS and central angle of $2\pi/15$, as illustrated in Figure 2. The cell radius is fixed at 2km. The heights of the MSs, BS and RS are 1.5 m,50 m and 30 m and path–loss at each link is calculated accordingly [7].

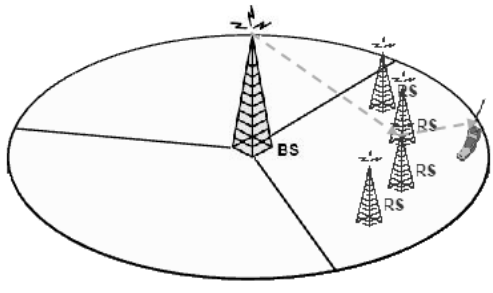


Fig.2 An example of RS assignment in one sector of the cell

Fig.3 shows the outage probability versus the average channel SNR and compares the performance of the proposed relay selection schemes (Best relay selection,

Two best relay selection, Nearest relay selection, Two nearest relay selection and the lower bound). We consider rate R=1 b/s/Hz. It can be seen that the best relay selection scheme outperforms the nearest selection scheme significantly. Besides , in both schemes a diversity gain is observed in the selection of two relays compare to the selection of just one relay whether the best one or the nearest one .

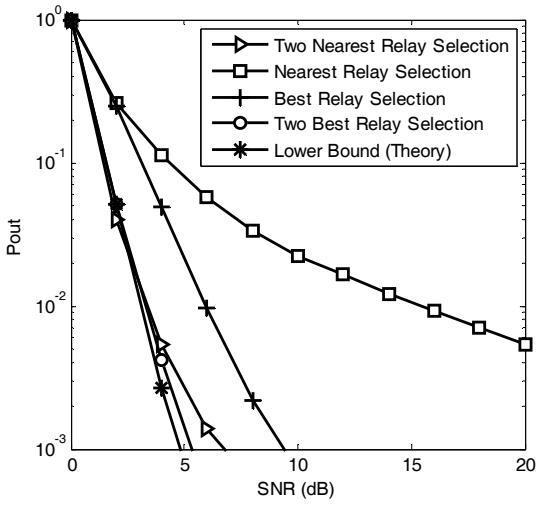


Fig3. Outage performance of Best and Nearest relay selection methods each for two cases: one and two relay selection along with lower bound (theory) for Best relay selection

VI. Conclusion

In this paper, we examined the problem of relay selection for cooperative communications in context of IEEE 802.16j standard . We proposed two relay selection methods: Best relay selection method and Nearest relay selection method, both of these methods are done in a per subchannel mode. We analyzed the outage performance of proposed Relaying schemes and showed that Best relay selection method outperforms Nearest relay selection method .in both cases when two relay selected instead of just one relay the performance is significantly improved. Future work may include the analysis of cooperative relaying with imperfect time and frequency synchronization and synchronization issues.

References

[1] "IEEE Standard for Local and Metropolitan Area Networks − Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment2: 2Mobile Operation in Licensed Bands," *IEEE Computer Society and the IEEE Microwave Theory and Techniques Society*, February 2006.

[2] "IEEE 802.16j Mobile Multihop Relay Project Authorization Request (PAR)," *Official IEEE 802.16j Website: http://standards.ieee. org/board/nes/projects/802-16j.pdf, March 2006.*

[3] J. Nicholas Laneman and Gregory W. Wornell, "Cooperative Diversityin Wireless Networks: Efficient Protocols and Outage Behavior", IEEE Trans. Inform. Theory, Vol. 50, No. 12 Dec. 2004 pp: 3062-3080.

[4] J. Chui , K.H.Lee , X.Li , A. Boariu , " Cooperative Relaying in Downlink for IEEE 802.16j ," *IEEE 802 . 16 mmr* , Jan 2007

[5] M. Zorzi and R. R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: Multihop performance," *IEEE Trans. Mobile Comput.*, vol. 2, no. 4, pp. 337–348, Oct.–Dec. 2003.

[6] A. Bletsas, H. Shin, M.Z. Win, "Cooperative Communications with Outage-Optimal Opportunistic Relaying", *IEEE Trans. on Wireless Communications* , Vol. 6, No. 9, pp. 3450-3460, September 2007.

[7] IEEE 802.16 Broadband Wireless Access Working Group, "Channel Models for Fixed Wireless Applications," IEEE, Tech. Rep., July 2001.

[8] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An Empirically Based Path Loss Model for Wireless Channels in Suburban Environments," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1205–1211, July 1999.

[9] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[10] S. Srikanth, V. Kumaran, C. Manikandan, "Orthogonal Frequency Division Multiple Access: is it the multiple access system of the future?," *AU-KBC Research Center, Anna University, India ,2006*

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Comparative Analysis of Feed-forward Neural Network & Recurrent Neural Network to Detect Intrusion.

Nipa Chowdhury, Mohammod Abul kashem

Dept of CSE, Dhaka University of Engineering & Technology
Gazipur: 1700
E-mail: {nipa, drkashemll}@duet.ac.bd

Abstract: As computer networks are grows exponentially security in computer system has become a foremost issue. Monitoring atypical activity can be one way to detect any violation that impedes computer systems security. Existing methods like statistical models [12] for intrusion detection not perform well whereas Neural network has been proved as an efficient method for intrusion detection [10]. In this paper Feed-forward and Recurrent Neural network is trained by Back propagation training algorithm and using normal data. Performances of these Neural Networks are compared against both normal data and intrusive data.

Keywords

Neural network, Recurrent neural network,
Feed-forward neural network, Back propagation training Algorithm, Elman Recurrent nerwork.

I. Introduction

An intrusion can be defined as any unauthorized attempt to access, manipulate, modify, or destroy information, or to render a system unreliable or unusable [15]. Security is always an important issue for Computer system and intruders are always trying to break it. As Computer networks size is vast and connectivity is increasing incessantly, the immense area is open for intruders and more & more systems are subject to attack. They try to break confidentiality, integrity of applications program and operating systems. It is not possible to build a completely secure system as novel attacks are created endlessly. Hence intrusion detection is extremely needed.

There are two basic approaches to intrusion detection - Misuse intrusion detection and Anomaly intrusion detection [8,12,16].In misuse intrusion detection, known patterns of intrusion (intrusion signatures) are used to try to identify intrusions when they happen [1,7,12]. It identify known attacks very well but False negative rate is high which means that it falsely identified novel attack as normal and it is difficult to keep them updated as the catalog of attacks grows[3].

In Anomaly intrusion detection, [9, 10, 12, 17] it is assumed that the nature of the intrusion is unknown, but that the intrusion will result in behavior different from that normally seen in the system [1].Anomaly detection techniques directly address the problem of detecting novel attacks against systems. This is possible because anomaly detection techniques do not scan for specific patterns, but compare current activities against statistical models of past behavior. Any activity sufficiently deviant from the model will be flagged as anomalous, and hence considered as a possible attack.

Drawback of anomaly detection is its inability to identify the specific type of attack that is occurring. However, probably the most significant disadvantage of anomaly detection approaches is high rates of false alarm. Because any significant deviation from the baseline can be flagged as an intrusion, non-intrusive behavior that falls outside the normal range will also be labeled as an intrusion - resulting in a false positive. In this paper our concern is only anomaly Detection [1].

The focus of research in intrusion detection has recently shifted from user-based and connection-based intrusion detection to process-based intrusion detection. Process-based monitoring intrusion detection tools analyze the behavior of executing processes for possible intrusive activity [2]. When a program is misused its behavior will differ from its normal usage. Therefore, if the behavior of a program can be collected, then the behavioral features can be used for intrusion detection. One of such behavior that is differing from misuse condition is System call. By tracing sequence of System call database is built [1].

Some statistical method [18, 19] can be used to classify the system call sequence as normal or intrusive. For example equality matching approach [1].Drawback of this approach is (1) large tables of program behavior must be built for each program, and (2) the equality matching approach does not have the ability to recognize behavior that is similar, but not identical to past behavior. The problem is that behavior that is normal, yet slightly different from past recorded behavior, will be recorded as anomalous. As a result, the false positive rate could be artificially elevated.

In this context, Neural network can be used as classifier because of its generalization capability which means that it can identify behavior that are present on training and also behavior that are not present in training but similar to those of training. This characteristic exactly matches with intrusion. In computer system or internet intrusion can occurred that are previously seen, similar to previous attacks or can be any novel attack. So Neural network can be used and it is trained by system call that is traced under normal condition.

In this paper, two NN architectures are used. One is Feed-forward and other one is Recurrent network. Both Networks are trained using Back propagation [20] training algorithm and the trained network is then used to detect possibly intrusive behavior by identifying significant anomalies.

II. Building Database

When intrusion is occurred in a process or application it creates some 'print' or 'signature' which is different from normal condition [2]. These signature or print can be System PID, System call, timing information, instruction sequences between system calls, and interactions with other processes and so on [1]. We have chosen system call sequence to define as these signatures or print. When a process is executed it needs resource, operating system parameter. So it made system call to request operating system to perform some operation. Sequence of system call in a process under normal condition must be different to that of the same process if intrusion is occurred.

The idea used to build the normal databases is extremely simple. At first system call of Linux process is traced. A separate database of normal behavior is build for each process of interest. A trace may consist following data:

```
execve("/usr/sbin/httpd", ["httpd"], [/* 37 vars */]) = 0
uname({sys="Linux", node="localhost.localdomain", ...}) = 0
brk (0)   = 0x8095f64
old_mmap(NULL,4096,PROT_READ|PROT_WRITE,MAP_PRIVATE|MAP_ANONYMOUS,-1,0)=0x40016000
open("/etc/ld.so.preload", O_RDONLY)    = -1 ENOENT (No such file or directory)
open("/etc/ld.so.cache", O_RDONLY)     = 3
fstat64(3, {st_mode=S_IFREG|0644,st_size=115094, ...}) = 0
old_mmap(NULL,115094,PROT_READ, MAP_PRIVATE, 3, 0) = 0x40017000
close(3) .                              .
```

Traced data include system call, no of times the system call executed & also some system parameter [1]. So some filtering is needed to remove other system parameter and resultant dataset contains only system call.

After filtering the above trace contains only:
execve, uname, brk, oldmmap, open, open, fstat64, old_mmap, close.

Length of system call sequence is depending on execution time and its length can be less than 100 or can be 8000 or more than that. To handle that huge amount of system call, sliding window [2] is used to build database. For example if we slide a window of size 3 for the above trace, then we get some small sequence like (execve, uname, brk), (uname, brk, old_mmap), (brk, old_mmap, open) and so on. Complete database is formed like Table1.

Table 1 shows complete database of system Call by sliding window

Sequence #	System call sequence after Sliding (Window size 3)
1	execve,uname,brk
2	uname,brk,old_mmap
3	brk,old_mmap,open
4	old_mmap,open,open
5	open,open,fstat64
6	open,fstat64,old_mmap
7	fstat64,old_mmap,close

For all process these type of separate database in need to built [2].

We use datasets from Computer immune systems [2, 13].Traces of system call at normal execution of process login, ps, named. inetd, stide is exists there. They also traced System call when intrusion is occurred for the above process. We have used the traces collected when Trojan attack is occurred in 'login' and 'ps', Buffer overflow in named, Denial service attack in inetd, stide process in this experiment. Trojan code[13] for login and ps that allow an intruder to login through a "back door" and hide their activities from system administrators whereas buffer overflow allowing a remote user to gain root access through a specially-formulated DNS query. Intrusion that ran against the 'inetd' program is a denial-of-service attack that ties up network connection resources. As the attack progresses, more of the system calls requesting resources return abnormally and are re-issued. The intrusion against the 'stide' program is a Denial-of-Service attack that affects any running program requesting memory. All of the above intrusive traces, Normal traces are defined in Computer Immune system [13].

We used only traces that collected in Linux 4.2 environment. Trace consists of System PID and number that represent system calls according to Linux 4.2 platform. To train NN this system call number is used.

III. Creation of Neural network

Success of intrusion detection technique lies to classify normal data, intrusive data, unseen normal and unseen/new intrusive data. It required that Intrusion detection tools are being able to generalize from previously observed behavior (normal or malicious) to recognize similar future behavior [6]. Anomaly detection tools that must be able to recognize future normal behavior that is not identical to past observed behavior in order to reduce false positive rates. To address this shortcoming, we utilize a simple neural network that can generalize from past observed behavior to recognize similar future behavior [6].

A neural network [4, 5] is composed of simple processing units, or nodes, and connections between them. The connection between any two units has some weight, which is used to determine how much one unit will affect the other. A subset of the units of the network acts as input nodes, and another subset acts as output nodes. By assigning a value, or activation, to each input node, and allowing the activations to propagate through the network, a neural network performs a functional mapping from one set of values (assigned to the input nodes) to another set of values (retrieved from the output nodes). The mapping itself is stored in the weights of the network [12].

Two types of Neural network architectures are used in this paper. One is Feed-forward Network- other one is Recurrent Neural networks

Feed-forward NN [4, 5] (Figure 1) allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward NNs tend to be straight forward networks that associate inputs with outputs [5]. The Feed forward Network used in this paper is given in Figure 1.

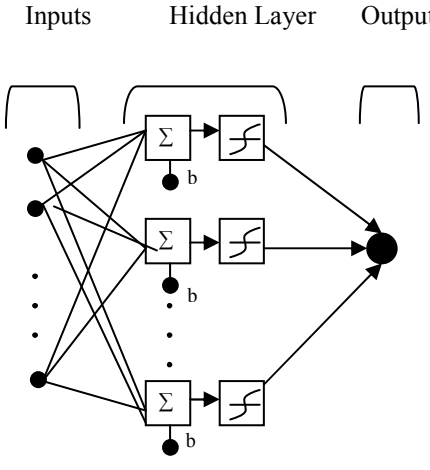


Figure 1: Feed-forward network.

On the other hand, in Recurrent network there is Input layer, Hidden layer, Output layer and feedback loop [4, 6]. The loop receives input from a node of hidden layer and sends its output to corresponding node and depends only on the activations of the hidden node from previous input. So this loop creates some delay and can store information about previous inputs. Recurrent network that used in this experiment is given in Figure 2.

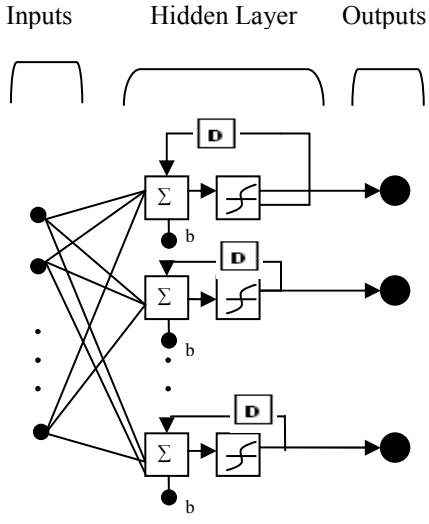


Figure 2: Recurrent neural network.

In Feed-forward networks only input, output mapping is happened not worked with previous input. But Next coming system call sequences set must be in a collection for a particular system call sequence and it can't be anything. For example two sequential sequence (execve, uname ,brk) and (fstat,old_mmap,close) both are valid sequence but it meaningless because you need to open a file before make any attempt to close. So some sequence like (open, fstat, old_mmap) is needed. So sequences are related to previous sequence and for intrusion detection this information is need to store and Recurrent network with delay loop used in this purpose.

IV. Design & Implementation

Matlab 7.0 is used to create neural network. Sliding window of size 3 is used in whole experiment means that no of input neuron is 3. 10 hidden neurons are used and Back propagation is used as training algorithm. 'tansig' is used as transfer function, 'purelin.' is used as network training function and 'trainlm' is used for weight/bias learning function.

In Feed-forward network architecture no of output neuron is 1. For example system call sequence (execve, uname, brk) target output is set to 1 means that it is normal. Training continues until one of the following two conditions is met. One is no of epoch reached to 200 or Goal is reached to 1.0000e-5.At this end testing dataset that contains both intrusive, novel attack, normal behavior is applied to the network and result is recorded in terms of mismatch. Mismatched occurred when neural network produce a value for a sequence which is far behind the value of the sequences that used for training. If no of mismatch of a process is greater than no of matches,

the process is identified as intrusive and result is recorded.

We used Elman Recurrent network as a type of Recurrent network to implement in Matlab. In this network all neural network parameters are same except that no of output neuron is same as no of input neuron. For input sequence (execve, uname, brk) Target output is set to the next sliding sequence i.e. (uname, brk, old_mmap) and for (uname, brk, old_mmap) target sequence is now (brk,old_mmap,open). Network is trained to predict the next incoming sequence. Network is trained by Back propagation and weights are modified in each pass so that Neural network can map to the next target sequence. After successful training testing dataset are applied to the network and performance is recorded in same way as for Feed-forward network.

V. Experimental Results

In this experiment network is trained by normal behavior of the process for example login(2 trace),ps(2 trace), inetd(1 trace), stide(4 trace).Performance of NN is tested against Normal trace of the above process, new process and intrusive attack[13]. We used intrusive attack when Trojan attack is occurred in login, ps, Buffer overflow in named, Denial of service attack in inetd, stide process.

Table 2 show the Success rate, False positive rate, False alarm rate when Feed- forward network is used.

Training Dataset	**Success (%)**	**False + ve (%)**	**False –ve (%)**
Stide (trace 1)	58	0	42
Stide (trace 3)	58	0	42
Stide (trace 4)	58	0	42
Stide (trace5)	64	0	36
Login(trace 1)	58	0	4
Login(trace 2)	58	0	42
Ps(trace 1)	58	0	42
Ps (trace 2)	58	0	42
Inetd	58	0	42

Table 3 shows the Success rate, False positive rate, False alarm rate when Neural network is trained by Recurrent neural network.

Training Dataset	**Success (%)**	**False + ve (%)**	**False –ve (%)**
Stide(trace 1)	70	30	
Stide(trace 3)	65	30	5
stide(trace 4)	70	30	
Stide(trace5)	70	30	
Login(trace 1)	42	42	16
Login(trace 2)	42	47	11
Ps(trace 1)	70	16	14
Ps (trace 2)	76	16	08
Inetd	42	42	16

From the result of Table 2 and 3 we see that success ratio is high in Recurrent neural network in comparison to Feed-forward network. Recurrent network can identified Denial of service attack in stide process, Trojan attack in ps process with success rate 70% while for other process success rate is near about 42%. But In Feed-forward network detection rate is poor in compare to Recurrent network. Although Recurrent network introduces some False positive but False negative rate is lower than false negative rate of Feed-forward network.

This is because in Feed-forward network output is depending only on current state i.e. present input sequence. This characteristics is appropriate for pattern recognition but not in the case where input states are correlated. But in out particular case-Length of system is not fixed it depend on execution time, operating platform but sequence are related. That is a valid system call sequence can be defined from previous input sequence. So NN needs to store previous sequence to predict next sequence and Recurrent network performs better that Feed- forward network in this context.

VI. Conclusion

In this paper we used Anomaly detection method to detect intrusions. Hence intrusions create a print when it occurred, we used one of such print i.e. system call. By using sliding window concept [2] we build database for normal behavior for each process. Feed-forward and Recurrent neural network is trained by using the data base and performance is tested against normal data, new normal data, novel attack and results shows that Recurrent network performs better because it exactly matches the characteristics with system call sequence. System call sequences are related to previous sequence and Recurrent network can store previous sequence and output is the combined result of both current input & previous input. In future we intend to expand our work to minimize false negative rate in Recurrent neural network.

REFERENCES

[1] Stephanie Forrest, Steven A. Hofmeyr, Anil Somayaji, Thomas A Longstaff: A Sense of Self for Unix Processes, Proceedings of the 1996 IEEE Symposium on Security and Privacy, 1996.

[2] Zhen Liu, Susan M. Bridges, and Rayford B. Vaughn- Classification of Anomalous Traces of Privileged and Parallel Programs by Neural Networks.

[3] Matthew Stillerman, Carla Marceau,Maureen Stillman, Intrusion Detection for Distributed Applications.

[4] Simon Haykin, Neural Networks, A comprehensive Foundation.

[5] Christos Stergiou and Dimitrios Siganos , Neural networks.

[6] Learning Program Behavior Profiles for Intrusion Detection- Anup K. Ghosh, Aaron Schwartzbard & Michael Schatz.

[7] S. Kumar and E.H. Spafford: A pattern matching model for Misuse intrusion detection. Proc. of the 17th National Computer Security Conference, pp. 11-21, October 1994.

[8] S. Kumar and E.H. Spafford: A Software Architecture to Support Misuse Intrusion Detection, Proc. 18th National Information Systems Security Conference, pp.194-204,1995.

[9] A.P.Kosoresow and S.A.Hofmeyr, "Intrusion Detection via System Call Traces", IEEE Software, Septemeber/October 1997, pp. 35-42

[10] Anomaly Detection Using Self/Nonself Discrimination for the Linux Kernel-Lars Olsson

[11]Intrusion Detection using Sequences of System Calls- Steven A. Hofmeyr, Stephanie Forrest, Anil Somayaji.

[12]Intrusion Detection Techniques and Approaches-Theuns Verwoerd and Ray Hunt.

[13].Datasets of Computer immune systems. http://www.cs.unm.edu/~immsec/systemcalls.htm

[14] MatLab help

[15] Jones A.K., Sielken R.S.: Computer system intrusion detection:a Survey.09.02.2000, http://www.cs.virginia.edu/~jones/IDS-research/Documents/jones-sielken-survey-v11.pdf

[16] James Cannady: Artificial Neural Networks for Misuse Detection.

[17] Zheng Zhang,Jun Li,C.N.Manikopoulos,Jay Jorgenson and Jose Ucles: Neural Networks in Statistical Anomaly Intrusion Detection.

[18] Intrusion Detection with Neural Networks: Jake Ryan, Meng-Jang Lin, Risto Miikkulainen

[19] HIDE: a Hierarchical Network Intrusion Detection System Using Statistical Preprocessing and Neural Network Classification: Zheng Zhang, Jun Li, C.N. Manikopoulos, Jay Jorgenson, Jose Ucles.

[20] A Performance Comparison of Different Back Propagation Neural Networks Methods in Computer Network Intrusion Detection: Vu N.P. Dao, Rao Vemuri

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Optimization of *k*-Fold Multicast Wireless Network Using *M/M/n/n+q* Traffic Model

Asfara R. Towfiq[1], N. A. Siddiky[2], Md. Imdadul Islam[3], and M. R. Amin[4]

[1,2,4]Department of Electronics and Communications Engineering, East West University, Mohakhali, Dhaka 1212, Bangladesh
[3]Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh
E-mail: ramin@ewubd.edu, imdad@juniv.edu

Abstract - An analytical model has been developed to determine the suitable value of the fold *k* of a *k*-fold multicast network with different traffic loads under Poisson traffic with finite queue. We have derived stationary probability distribution for the network states and then derived expressions for the throughput, blocking probability and the probability of delayed service of the network. It has been found in this study that the network throughput increases very fast as we increase the fold number. However, beyond a threshold or optimum value of *k* the throughput profile becomes flat. It has also been found that as the system parameter *k* is increased, the blocking probability decreases rapidly, but after the optimum value of *k*, found in the throughput variation, the blocking probability remains constant.

I. Introduction

Multicast communications [1-11] involve transmitting information from a single source to multiple destinations. This is an important requirement for high-performance communications networks. Multicast communication is one of the most important collective communication operations and is highly demanded in broad-band integrated services network (BISDN) and in communication-intensive applications in parallel and distributed computing systems, such as distributed database updates and cache coherence protocols. It is projected that multicast will also be increasingly used to support various other interactive applications such as multimedia, teleconferencing, web servers and electronic commerce on the Internet. Many of these applications require predictable communications performance, such as guaranteed multicast latency and bandwidth, called quality of service (QoS) in addition to multicast capability. The QoS guarantees and the non-uniform nature of multicast traffic make the problem of the analysis of multicast communication very challenging.

A *k*-fold multicast network was recently proposed [10] as a cost-effective solution for providing better QoS functions in supporting real-world multicast applications. A *k*-fold multicast assignment is defined as a mapping from a subset of network source nodes to a subset of network destination nodes, with up to *k*-fold overlapping allowed among the destinations of different sources. In other words, any destination node can be involved in multicast connections, from up to *k* different sources.

It is to be mentioned here that to provide a quantitative basis for the network designers, determining an optimum value of the system parameter *k* (the fold number) is essential. For a big multicasting network, for example, country-wide e-commerce system, the number of source and destination nodes are large enough to provide constant offered traffic which resembles to traffic model of unlimited users (Erlang's model), whereas [12] uses a limited users case (Engset model) for a small network. Keeping this view in mind, we have, in this paper, developed an analytical model to determine the suitable value of *k* under different traffic loads with unlimited users for a *k*-fold multicast network under Poisson traffic with finite buffer. We have derived stationary distribution for the network states and then derived expressions for the network throughput, blocking probability and probability of delayed service of the network.

We organize the paper as follows. Section II describes the mathematical model for the derivation of the stationary probability distribution of the network states. Expressions for the network throughput, blocking probability and the probability of delayed service have also been obtained in this section. Section III describes the results of the investigation by showing the variation of the above mentioned three important quantities with respect to the fold number for different offered traffic. Finally, Section IV concludes the paper.

II. Traffic Model

Here, we derive stationary distribution of the *k*-fold network, from which we can obtain network throughput and the blocking probability. We assume the Markovian *M/M/n/n+q* model, where *n* is the total number of servers in the system and *q* is the number of waiting positions. The corresponding Markov chain is shown in Fig. 1. The average arrival rate is denoted by λ and the average service rate is denoted by μ.

Let us consider that there are *j* multicast connection requests, and let $p_{\deg}(j,m)$ be the probability that a

destination node is the destination of exactly m of the multicast connection requests; or we can say that a destination node is of degree m under these j multicast connection requests. The probability that any multicast connection request chooses this destination node is θ and is independent of other multicast connections. Thus, we have

$$p_{\deg}(j,m) = \binom{j}{m}\theta^m(1-\theta)^{j-m}, \quad m \in \{0,1,\ldots,j\} \tag{1}$$

which is a binomial random variable. We assume that each destination node has the same distribution given by equation (1). Furthermore, we assume that choosing of a destination node by a multicast connection is independent of other destination nodes. Thus, in addition to having the same distributions, the degrees of the destination nodes are also independent of each other. That is, they are a group of independent, identically distributed (i.i.d.) random variables.

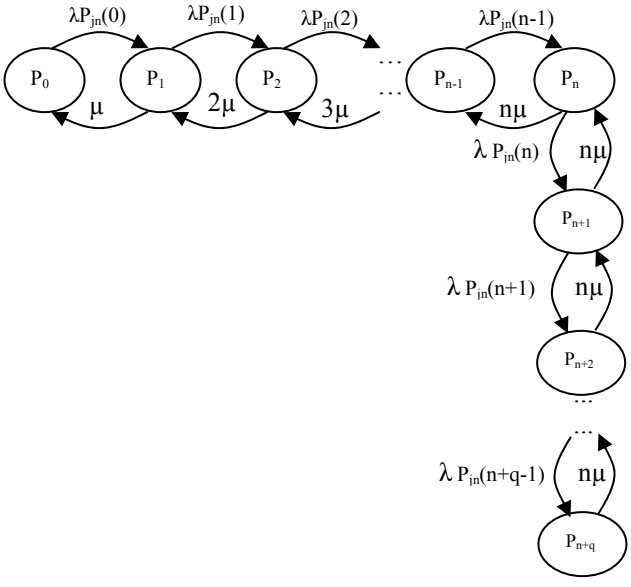


Fig. 1 Markov Chain for *k*-fold network.

Let $P_{\mathrm{mc}}(j)$ be the probability that j multicast connection requests are mutually compatible (*m.c*) in a *k*-fold multicast network. We note that a set of multicast connection requests are *m.c*. when none of the destination nodes has a degree more than k when realized simultaneously in the network. From equation (1), it is obvious that the probability of a destination node having a degree less than or equal to k is $\sum_{m=0}^{k} p_{\deg}(j,m)$ for $j > k$, and 1 for $j \le k$, because when $j \le k$, no destination node can have a degree more than *k*. Since the degrees of destination nodes are independent of each other, we have [12]

$$P_{\mathrm{mc}}(j) = \begin{cases} \left(\sum_{m=0}^{k} p_{\deg}(j,m)\right)^n, & j > k \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

Now, let us consider that a new multicast connection request arrives when j multicast connections are already in the network. If this new connection can be realized along with those ongoing connections, we say that it can *join* the ongoing connections. Let $P_{\mathrm{jn}}(j)$ be the probability that a new multicast connection can join j ongoing connections. It can be shown that [12]

$$P_{\mathrm{jn}}(j) = \frac{P_{\mathrm{mc}}(j+1)}{P_{\mathrm{mc}}(j)}. \tag{3}$$

By solving the Markov chain of Fig. 1, the stationary states are found to have the probabilities

$$P_r = \frac{\rho^r}{r!}\prod_{l=0}^{r-1} P_{\mathrm{jn}}(l)P_0, \quad 0 \le r \le n \tag{4}$$

and

$$P_{n+s} = \left(\frac{\rho}{n}\right)^s \frac{\rho^n}{n!}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l)P_0, \quad 1 \le s \le q \tag{5}$$

where $\rho = \lambda/\mu$ is the offered traffic. The constant P_0 is obtained from the normalization of the probabilities and is given by

$$P_0 = \left(\sum_{r=0}^{n} A_r + \sum_{s=1}^{q} B_s\right)^{-1}, \tag{6}$$

where

$$A_r = \frac{\rho^r}{r!}\prod_{l=0}^{r-1} P_{\mathrm{jn}}(l), \quad B_s = \left(\frac{\rho}{n}\right)^s \frac{\rho^n}{n!}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l). \tag{7}$$

In our present infinite source case, the arrival rate of the successful calls can be written as $\sum_{j=0}^{n+q} \lambda P_{\mathrm{jn}}(j)$. Thus, the total number of successful connection requests carried by the network during a long time period $[0,T]$ is obtained by summing over j:

$$\begin{aligned} N_{\mathrm{succ}} &= T\lambda \sum_{r=0}^{n+q} P_r P_{\mathrm{jn}}(r) \\ &= T\lambda \sum_{r=0}^{n} P_r P_{\mathrm{jn}}(r) + T\lambda \sum_{s=1}^{q} P_{n+s} P_{\mathrm{jn}}(n+s). \end{aligned} \tag{8}$$

Therefore, the network throughput can be written as

$$T_H = \frac{N_{\mathrm{succ}}}{T}. \tag{9}$$

Using equations (4) and (5) in equation (8), and the result so obtained when substituted in equation (9), the expression for the throughput, finally takes the following form

$$\begin{aligned} T_H = {} & \lambda \sum_{r=0}^{n}\prod_{l=0}^{r-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(r)\frac{\rho^r}{r!}P_0 \\ & + \lambda \sum_{s=1}^{q}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(n+s)\left(\frac{\rho}{n}\right)^s \frac{\rho^n}{n!}P_0. \end{aligned} \tag{10}$$

We can also calculate the expression for the carried traffic. In our present case of infinite source, the expression for the carried traffic turns out to be

$$\overline{X} = \sum_{r=0}^{n+q} r P_r P_{\mathrm{jn}}(r)\,, \tag{11}$$

which after simplification and using the expressions for P_{r} from equations (4) and (5), takes the following form

$$\begin{aligned}\overline{X} &= \sum_{r=0}^{n}\prod_{l=0}^{r-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(r)\frac{r\rho^r}{r!}P_0 \\ &+ \sum_{s=1}^{q}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(n+s)\left(\frac{\rho}{n}\right)^s \frac{(n+s)\rho^n}{n!}P_0 .\end{aligned} \tag{12}$$

The lost traffic is then

$$\begin{aligned}\overline{L} = \rho - \overline{X} &= \rho - \sum_{r=0}^{n}\prod_{l=0}^{r-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(r)\frac{r\rho^r}{r!}P_0 \\ &- \sum_{s=1}^{q}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(n+s)\left(\frac{\rho}{n}\right)^s \frac{(n+s)\rho^n}{n!}P_0 .\end{aligned} \tag{13}$$

Therefore, the blocking probability is

$$\begin{aligned}P_B = \frac{\overline{L}}{\rho} &= 1 - \sum_{r=0}^{n}\prod_{l=0}^{r-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(r)\frac{r\rho^{r-1}}{r!}P_0 \\ &- \sum_{s=1}^{q}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l)P_{\mathrm{jn}}(n+s)\left(\frac{\rho}{n}\right)^s \frac{(n+s)\rho^{n-1}}{n!}P_0 .\end{aligned} \tag{14}$$

Similarly, we can calculate the probability of delayed service as

$$P_D = \sum_{s=1}^{q} P_{n+s} = \sum_{s=1}^{q}\prod_{l=0}^{n+s-1} P_{\mathrm{jn}}(l)\left(\frac{\rho}{n}\right)^s \frac{\rho^n}{n!}P_0 . \tag{15}$$

The throughput T_H, the blocking probability P_B and the probability of delayed service P_D from equations (10), (14) and (15) are plotted against the fold-number *k* in Figs. 2, 3, and 4 respectively for different traffic loads in the next section.

III. Results and Discussions

For numerical appreciation of our results, we have plotted in Figs. (2), (3) and (4), the throughput, blocking probability and the probability of delayed service as functions of the fold number *k*. We have also plotted in Fig. 5 the probability of delayed service with respect to the fold number taking queue-length as a parameter.

It is seen from Fig. 2 that if the fold number *k* of the network is increased, network throughput increases very fast for the lower values of the system parameter *k*, in our study up to $k \approx 8$; beyond this value of *k*, the network throughput is almost constant with respect to the system parameter *k* for particular offered traffic. We also observe that as the offered traffic is increased, the throughput also increases.

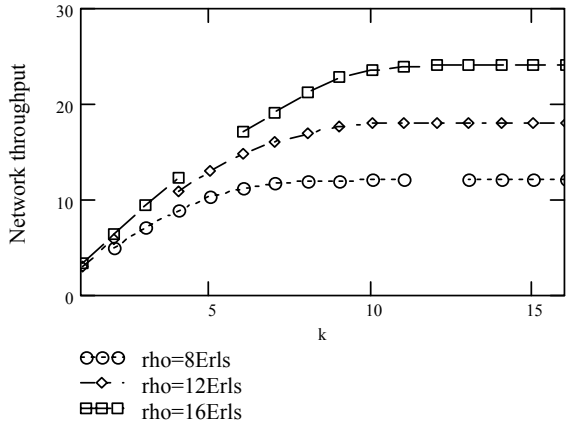


Fig. 2 Network throughput as a function of the fold number under different offered traffic (n=14, q=3, θ =0.31).

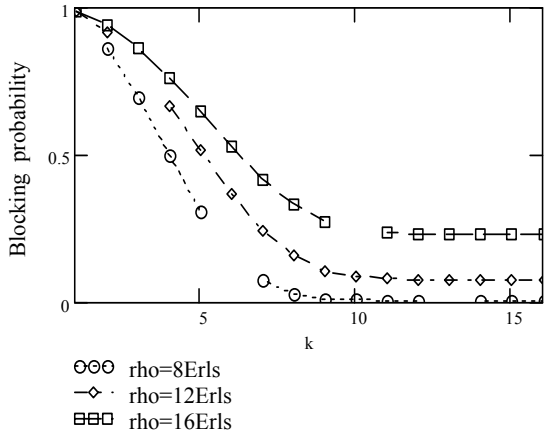


Fig. 3 Blocking probability as a function of the fold number under different offered traffic (n=14, q=3, θ =0.31).

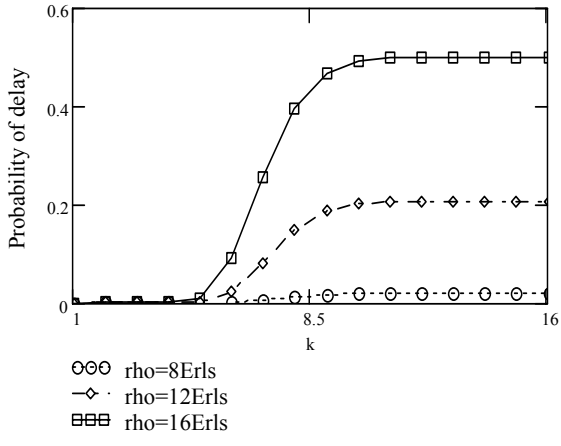


Fig. 4 Probability of delay as a function of the fold number under different offered traffic (n=14, q=3, θ =0.31).

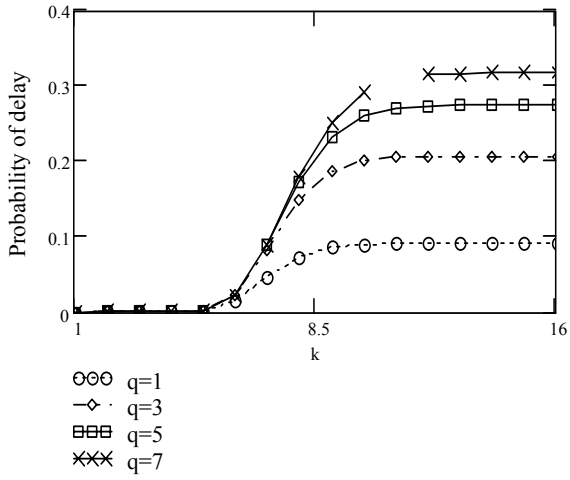


Fig. 5 Probability of delay as a function of the fold number under different buffer length (n=14, ρ =12, θ =0.31).

Fig. 3 shows the variation of the blocking probability with respect to the fold number *k*. It is seen from this figure that as the system parameter *k* increases, the blocking probability decreases rapidly for the lower values of the fold number. However, after the optimum value of *k* as found in the previous case of the variation of the network throughput, shown in Fig. 2, ($k \approx 8$), the blocking probability remains constant for particular value of the offered traffic.

Fig. 4 shows the variation of the probability of delayed service with respect to the fold number *k*. It is observed that the probability of delay is almost negligible for lower values of the fold number *k*, whereas, it suddenly starts increasing as the fold number approaches the optimum value $k \approx 8$. However, after a certain value of k, the probability of delay becomes constant.

Fig. 5 shows the variation of the probability of delay with the fold number *k* for different buffer length *q*. It is seen that the behavior of the figure is almost similar to Fig. 4, that the probability of delay is almost negligible for lower values of the fold number *k*, whereas, it suddenly starts increasing as the fold number approaches the optimum value of the system parameter $k \approx 8$. We see from the figure that as we increase the buffer length, the probability of delay also increases. However, the rate of increment of the probability of delay is reduced as we increase *q*.

IV. Conclusions

To provide a quantitative basis for the network designers of a multicast wireless network, determining an optimum value of the system parameter *k* (the fold number) is essential. This optimum value of the fold of the network should provide maximum throughput with minimum blocking probability. We have studied this problem with assuming unlimited users and finite queue length in a Markovian process. It has been found in this study that the network throughput increases very fast with the increase of the fold number. However, beyond a threshold or optimum value of *k* the throughput profile becomes flat. Furthermore, it has been found that as the system parameter *k* is increased, the blocking probability decreases very rapidly. It has been found that as the fold number *k* proceeds toward the optimum value as obtained from the throughput variation, the blocking probability also remains constant. We have thus obtained an optimum value of the system parameter *k* for the wireless multicast network.

References

[1] F. K. Hwang and A. Jajszczyk, "On nonblocking multiconnection networks", *IEEE Trans. Commun.*, vol. COM-34, pp. 1038-1041, Sep. 1986.

[2] Y. Yang and G. M. Masson, "Nonblocking broadcast switching networks", *IEEE Trans. Comput.*,, vol. 40, pp. 1005-1015, Sep. 1991.

[3] P. Feldman, J. Friedman, and N. Pippenger, "Wide-sense nonblocking networks", *SIAM J. Discr. Math.*, vol. 1, no. 2, pp. 158-173, May 1988.

[4] C. Lee and A. Y. Oruc, "Design of efficient and easily routable generalized connectors", *IEEE Trans. Commun.*, vol. 43, pp. 646-650, Feb.-Apr. 1995.

[5] Y. Yang and J. Wang, "A new self-routing multicast network", *IEEE Trans. Parallel Distrib. Syst.*, vol. 10, pp. 1299-1316, Dec. 1999.

[6] N. McKeown, A. Mekkittijul, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch", *IEEE Trans. Commun.* , vol. 47, pp. 1260-1267, Oct. 1999.

[7] M. Andrews, S. Khanna, and K. Kumaran, "Integrated scheduling of unicast and multicast traffic in an input-queued switch", in *Proc. IEEE INFOCOM*, 1999, pp. 1144-1151.

[8] Y. Yang and J. Wang, "On blocking probability of multicast networks", *IEEE Trans. Commun.*, vol. 46, pp. 957-968, Jul. 1998.

[9] Y. Yang, "The performance of multicast banyan networks", *J. Parralel Distrib. Comput.*, vol. 60, no. 8, pp. 909-923, 2000.

[10]Y. Yang and J. Wang, "Nonblocking k-fold multicast networks", *IEEE Trans. Parallel Distrib. Syst.*, vol. 14, pp. 131-141, Feb. 2003.

[11]Y. Yang, J. Wang, and C. Qiao, "Nonblocking WDM multicast switching networks", *IEEE Trans. Parallel Distrib. Syst.*, vol. 11, pp. 1274-1287, Dec. 2000.

[12]Zhenghao Zhang and Y. Yang, "Performance analysis of k-fold multicast networks", *IEEE Trans. Commun.*, vol. 53, pp. 308-314, Feb. 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Analysis of an Optical Burst Switching (OBS) Network

Md. Shamim Reza and Satya Prasad Majumder

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
E-mail: shamim_reza@eee.buet.ac.bd

Abstract—This paper presents a burst loss rate (BLR) scheme suitable for slotted optical burst switched networks. An analytical model for burst loss rate is presented. The effect of several design parameters on the above performance measure is examined with the aid of simulations. The model results are found to be satisfactory agreement with the expected results.

Index Terms – Optical burst switching (OBS), burst loss rate (BLR).

I. INTRODUCTION

In order to utilize efficiently the amount of raw bandwidth in wavelength division multiplexing (WDM) networks, all-optical transport method must be developed to avoid electronic buffering while handling bursty traffic [1]. Several technologies such as optical circuit switching (OCS), optical packet switching (OPS) have been proposed for the transfer of data over dense wavelength-division multiplexing. However Optical circuit switching is not primarily bandwidth efficient for IP traffic [2]. By enabling time division multiplexing in the optical domain, Optical packet switching networks benefit from statistical multiplexing; this ensures good utilization of network resources [3]. Optical packet switching (OPS) performs packet switching in which a packet is sent along with its header. The packet is buffered at the node in the optical domain, while the header is being processed by an intermediate node, either all optically or electronically (after an O/E conversion). However, high-speed optical logic, optical-memory technology and synchronization requirements are major problems with this approach [2]. Optical burst switching is a new wave length division multiplexing technology that retains some of the advantages of OPS and it was first proposed by Qiao and Yoo [4].

Optical burst switching (OBS) is a new wave length division multiplexing technology that combines the advantages of both wavelength-routed (WR) networks and optical packet switching (OPS) networks [4]-[5]. As in WR networks, there is no need for buffering and electronic processing for data at intermediate nodes. At the same time, OBS increases the network utilization by reserving the optical channel for a limited time period. The basic switching entity in OBS is a burst. A burst is a train of packets moving together from one ingress node to one egress node and switched together at intermediate nodes [5]. An ingress node assembles the internet protocol packets, that are coming from the local access networks and destined to the same egress node, into large burst [4].An optical burst consists of two parts, a header and a data burst. The header is called the control burst (CB) and is transmitted separately from the data, which is called data burst (DB). After a given delay the CB is followed by the DB, which travels over the same path reserved by the control burst. The delay between sending the CB and the DB is called the burst offset time. The value of offset time is chosen to be greater than or equal to the total processing time delay encountered by the CB [5].The control burst contains an information about the sender, receiver and transmission wavelength of the corresponding burst. Its main function is to configure all the core nodes along the path to destination so that the burst travels smoothly in the optical domain without the need to be converted into the electrical domain.

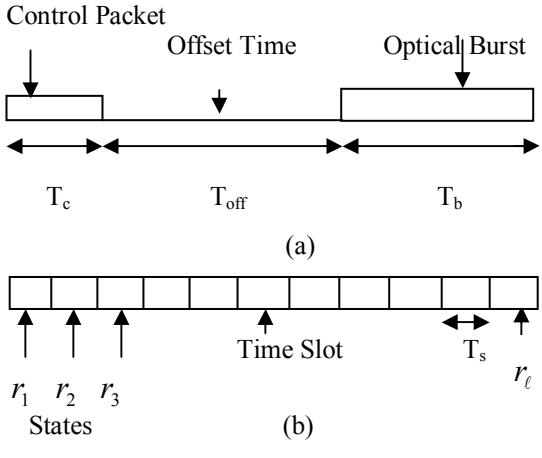


Fig. 1: (a) Transmission of an optical burst (b) Slotted timing diagram.

Consequently no buffering is needed for the data burst at intermediate nodes. It should be noticed that the optical burst arrives at the OBS node under consideration after an offset time T_{off} from the arrival of the control packet [Figure 1.(a)], which takes care of the processing and configuration times of the control packet and OXC (optical cross connect) fabric, respectively. Thus, the total time T spent from the transmission of the control packet until the end of the optical burst is

$$T = T_c + T_{off} + T_b$$

where T_c and T_b are the control packet and optical burst time durations, respectively.

The remainder of this paper is organized as follows. In section II, we introduce a mathematical model of burst blocking probability derived in [4]. We also relate the problem of calculating the burst blocking probability to the problem of calculating burst loss rate in OBS networks considering wavelength conversion capability in section II. Simulation results have been presented in section III. Finally we give our conclusion in section IV.

II. BURST LOSS RATE IN OBS NETWORKS WITH WAVELENGTH CONVERSION CAPABILITIES

In our analysis we use slotted timing model [Figure 1.(b)] in which we divide the entire period T of a burst into small time slots, each of duration T_s is called slot time. The total number of slots ℓ is calculated as ℓ =T/T_s, where ℓ is an integer. It should be emphasized that during a time slot T_s, a node would get enough information about the selected wavelength.

Let us consider an OBS network with w wavelengths. Each node in the network is equipped with u wavelength converters, u ε {0, 1, 2, . . . , w }.The factor

$$\rho = \frac{u}{w} \qquad 0 \le \rho \le 1$$

is called the network conversion capability. If ρ =0, then network has no network conversion capability, whereas if ρ =1, then the network has full conversion capability. When the arriving burst is to be served with a specific wavelength, this wavelength is removed from the pool until after the service is completed. If another arriving burst is to be served with a wavelength not available in the pool, it will be converted to another one from the pool. This wavelength is then removed and u is decreased by one. Blocking occurs whenever the pool is empty or a used wavelength is needed while u =0.

Consider an n-λ state $r_{i_n i_{n-1} i_{n-2} \ldots\ldots\ldots\ldots i_2 i_1}$ where n ε {1,2,..... $\ell \wedge w$ } and $i_1, i_2, i_3, \ldots\ldots\ldots\ldots, i_n$ ε {1,2,..., ℓ} with $i_n > i_{n-1} > \ldots\ldots\ldots > i_1$, here $\ell \wedge w$ = min(ℓ , w). The node in this state is serving slot i_n of the first burst, slot i_{n-1} of the second burst and so one. Let us consider

$$r_{i_n i_{n-1} i_{n-2} \ldots\ldots\ldots\ldots i_2 i_1} = e_n$$

Expression for e_n is available in [4], we get for any k ε {1,2,..., $\ell \wedge w$ }

$$e_k = \frac{\prod_{i=0}^{k-1} \frac{w - i(1-\rho)}{w\frac{1-A}{A} + i(1-\rho)}}{1 + \sum_{n=1}^{\ell \wedge w} \binom{\ell}{n} \prod_{i=0}^{n-1} \frac{w - i(1-\rho)}{w\frac{1-A}{A} + i(1-\rho)}}$$

Thus if $n \neq w$, the blocking probability [4] for this node is given by

$$P_b(n) = \frac{A(\ell - 1)}{w\ell}(1-\rho) n \binom{\ell}{n} e_n$$

If $n = w$, however, it [4] is given by

$$P_b(w) = \frac{A(\ell - 1)}{w\ell}(1-\rho) w \binom{\ell}{w} e_w + A\rho \binom{\ell - 1}{w} e_w$$

We denote the burst arrival probability on an input wavelength in a given time slot as A ($0 \le A \le 1$) and assume that A is independent on burst arrivals in other wavelengths and burst arrivals on in previous time slots. Let A_k ($0 \le A_k \le 1$) be the probability for k ($0 \le k \le w$) arrivals to the output fiber on a given time slot. A_k is then distributed according to a Binomial process $P_A(k/w)$ [3].

$$A_k = P_A(k/w) = \binom{w}{k} (A)^k (1-A)^{w-k}$$

The average number of burst arrivals in a time slot is E[A_k]= Aw. If two control packets are to reserve the same wavelength at a given core node for two different bursts, then only one burst will be offered to this wavelength. The other will be blocked and lost (unless there is an available wavelength converter or fiber delay line (FDL)). So we obtain the average burst loss rate as follows:

$$BLR_{avg} = \frac{1}{Aw}\sum_{k=1}^{\ell \wedge w} A_k .k.P_b(k)$$

III. SIMULATION RESULTS

The simulation results have been shown below. Fig. 2 shows the effect of burst arrival probability on burst loss rate. BLR increases with the increasing of burst arrival probability. As there is more number of bursts, more bursts will be lost. Fig. 3 shows the effect of wavelength conversion capability on the burst loss rate. With the increasing of wavelength conversion capability, burst loss rate decreases. With constant burst arrival probability, network traffic (traffic=A.ℓ) increases as the number of slots per burst increases. So fig. 4 shows burst loss rate increases with the increasing number of slots per burst or the network traffic. When the network has full wavelength conversion capability then the burst loss rate is zero until number of slots per burst is greater than the number of wavelengths. The burst loss rate decreases with the increasing of total number of wavelengths which is shown in fig. 5. Fig. 6 shows the effect of burst blocking probability on burst loss rate. The burst loss rate increases with the increasing of blocking probability and when the burst blocking probability is one then the burst loss rate is also one.

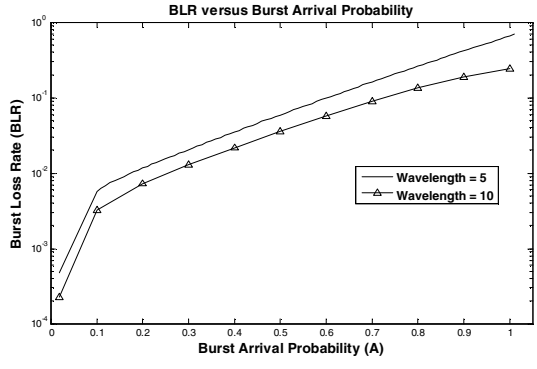


Fig. 2: The effect of burst arrival probability on average burst loss rate with different number of wavelength. Here ρ = 0.1, ℓ = 40 and $\ell > w$.

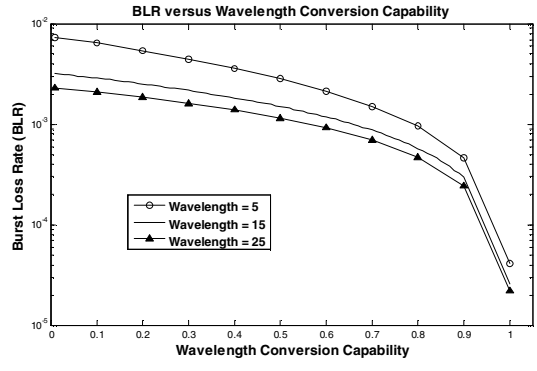


Fig. 3: The effect of wavelength conversion capability on average burst loss rate with different number of wavelengths. Here A = 0.1, ℓ = 40 and $\ell > w$.

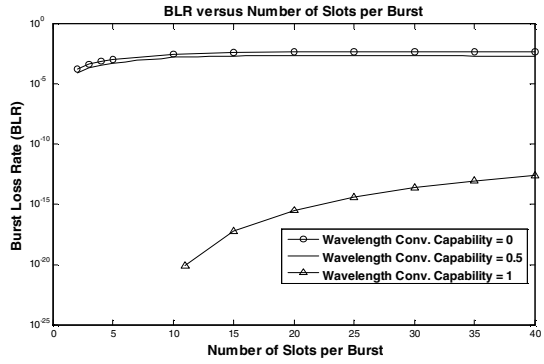


Fig. 4: The effect of number of slots per burst on average burst loss rate with different wavelength conversion capabilities. Here A = 0.1 and w = 10.

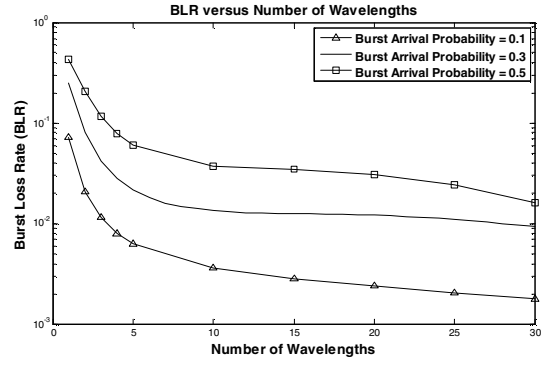


Fig. 5: The effect of wavelengths on average burst loss rate with different burst arrival probabilities. Here ρ = 0.1, ℓ = 40 and $\ell > w$.

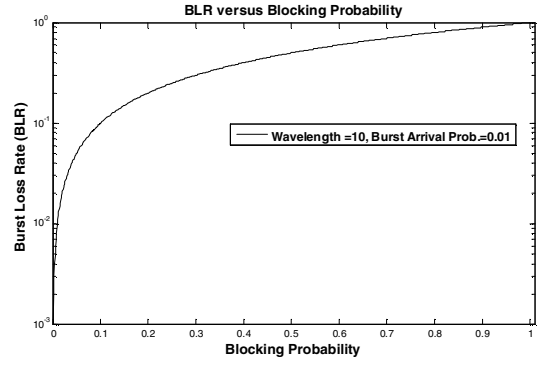


Fig. 6: The effect of burst blocking probability on average burst loss rate. Here w = 10, $\ell > w$ and A=0.01.

IV. CONCLUSION

This letter has presented an analytical model of a burst loss rate scheme for slotted burst of an optical burst switching network. The effects of several network design parameters on the system performance measures have been investigated and presented numerically. Results show that system performance improves with the increasing of wavelengths and wavelength conversion capability but degrades as the network traffic or the burst blocking probability increases.

REFERENCES

[1] V. M. Vokkarane and J. P. Jue, "Segmentation-based non-preemptive channel scheduling algorithms for optical burst-switched networks", Journal of light wave technology, vol. 23, no.10, pp. 3125-3137, October 2005.

[2] J. Phuritatkul, Y. Ji and S. Yamada, "Proactive wave length pre-emption for supporting absolute QoS in optical-burst-switched networks", Journal of light wave technology, vol. 25, no. 5, pp. 1130-1137, May 2007.

[3] H. Øverby, "Packet loss rate differentiation in slotted optical packet switched networks", IEEE Photonics technology letters, vol. 17, no. 11, pp. 2469-2471, November 2005.

[4] H. M. H. Shalabi, " A simplified performance analysis of optical burst switched networks." Journal of light wave technology, vol. 25, no. 4, pp. 986-995, April 2007.

[5] A. M. Kaheel, H. Alnuweiri and F. Gebali, " A new analytical model for computing blocking probability in optical burst switching networks." Journal of IEEE in communications, vol. 24, no. 12, pp. 120-128, December 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Survey on Adaptive Caching Techniques in Peer-to-Peer Network

Md. Tauhiduzzaman, Md. Renesa Nizamee, Sheikh Md. Rubabuddin Osmani
Md. Mohiuddin Khan, A. S. M. Ashique Mahmood

Department of Computer Science and Information Technology, Islamic University of Technology (IUT)
Board Bazar, Gazipur.
E-mail: tauhid47@yahoo.com, renesa_05@yahoo.com, fagun333@yahoo.com,
rafi867@gmail.com, shahan_iut@yahoo.com

Abstract – In this paper we discuss on adaptive caching techniques in peer-to-peer network. We have made an extensive survey on the different Adaptive caching techniques in peer-to-peer network like Squirrel, Tuxedo and PeerOLAP. Vast improvement can be made in the caching of the peer-to-peer network. We observed that combining and interchanging different methods used in different techniques results in a better performance. In this paper we discussed about the pros and cons of different techniques and proposed a solution to build a caching technique to make the performance better.

Keywords:
Peer-to-peer, adaptive caching, Tuxedo, Squirrel, PeerOLAP

I. Introduction

A peer-to-peer network is a class of systems and applications that employ distributed resources to perform a critical function (usually in a decentralized manner). On the Internet, peer-to-peer (referred to as P2P) is a type of transient Internet network that allows a group of computer users with the same networking program to connect with each other and directly access files from one another's hard drives, for example Napster [7], Gnutella [8] etc. Caching is very important for these peer-to-peer systems. There are two types of caching used in the peer-to-peer systems: adaptive and dynamic. An adaptive caching consists of multiple distributed caches which dynamically join and leave cache groups (referred to as *cache meshes*) based on content demand. Dynamic Caching is a caching technique that caches objects that change often and very infrequently.

In section II we discuss the architectures of the techniques and the pros and cons of these techniques. In section II (A) we discuss on Squirrel, in section II (B) Tuxedo and in section II (C) PeerOLAP.

In section III we also provide a table briefly differentiating these three techniques on some key factors.

Based on the advantages and disadvantages we found in these architectures, we consider on finding some solutions on caching of objects and finding delegates keeping the objects efficiently in low cost in section III. .

II. Discussion on techniques

A. Squirrel

Squirrel caching technique evaluates decentralized web caching algorithm. The key idea in Squirrel is that each client browser exports its cache to other nodes in the corporate network, thus synthesizing a large shared virtual web cache. The browser and Squirrel share a single cache managed by the Squirrel. Squirrel pools resources from many desktop machines which are called clients. Squirrel uses a self-organizing, peer-to-peer routing substrate called Pastry [6].

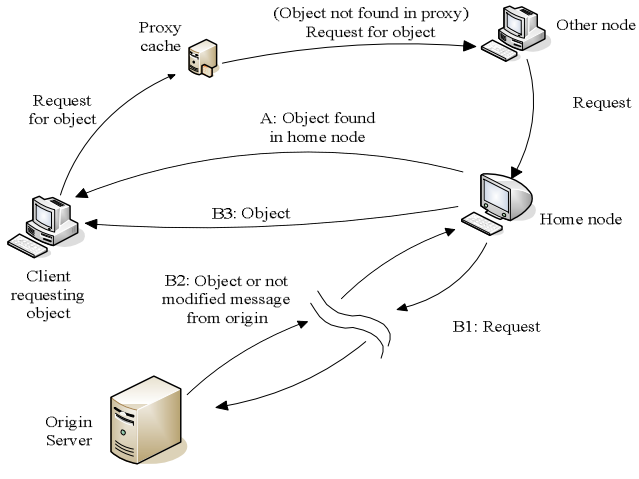


Fig. 1 A typical Home-node approach in the Squirrel caching technique: When the requested object is not found in the local cache, the request is passed to the home node. If the object is there it is dispatched to the requesting client (A). Otherwise, the request is passed to the origin server (B1), and the origin server replies with a not modified message, or the object to the home-node (B2). Then the home-node sends the object to the requesting client.

The design of Squirrel is quite simple. Web browsers issue their requests to the squirrel proxy running on the same node. If this proxy knows that the object is unreachable, it forwards the request directly to the origin server. Otherwise, it checks the local cache for fresh

copy. If the fresh copy is not found in this cache, then squirrel essentially tries to locate a copy on some other node and invokes the Pastry routing procedure to forward the request to the node with nodeID numerically closest to this objectID. It designates the recipient as the home node for this object. There are two approaches based on where the object resides: Home-node approach and Directory approach. The Home-node approach is shown in fig. 1. In [4], the architecture is described in detail. The implementation of Squirrel is shown in [5].

We found the advantages of Squirrel:

- An increase in the number of client nodes corresponds to an increase in the amount of shared resources.
- There is much less traffic around the home node (Directory approach).
- Each node's browser cache works as Squirrel cache, so a huge cache memory can be achieved.
- Workload for popular objects is controlled, so there is very low probability of overloaded traffic in a peer (in Home-store approach).
- A client can find the whole object in a delegate, rather than partial one.
- The client can always get a fresh and updated copy of the object.

The disadvantages we found:

- No bandwidth or latency is considered in case of choosing any delegate.
- No cost evaluation is considered for choosing delegate; as a result download may be costly.
- Path length is not considered, so the object may travel a longer path, which may be time consuming.
- Update is done in a non-regular basis – it's done when a request for an object comes. So it may be time consuming to get the object.
- The web browser is shared by squirrel, so the web browser's own task can be hampered by tons of data.
- Only the nodes residing in the Squirrel network can use the technique.
- Squirrel can be supported only in a limited-sized network.
- Can have a low cache memory if there is a very few nodes in the network.
- Squirrel is not scalable comparing to other techniques.

B. Tuxedo

Tuxedo includes an *adaptive neighborhood set algorithm* for different Web servers, and a *hierarchical cache digest* for sharing of transcoded [3] versions and value-added services. Tuxedo is more scalable and completely decentralized than other peer to peer caching techniques.

The CONCA proxy cache [10] is the building block of Tuxedo. Tuxedo is an overlay network which uses efficient neighborhood propagation and adaptive neighborhood mechanism for different web servers, and by using a hierarchical cache digest to store information related to transcoded content. It uses CAN [9] for locating the nodes in the network. Also, Tuxedo employs the notion of *reputation* to deal with security and trust concerns among peers. Tuxedo follows four steps, described in [2].

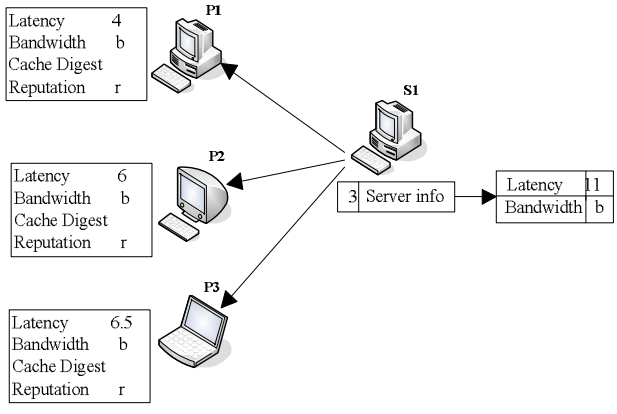


Fig. 2 Two tables maintained at each Tuxedo cache node: server table and neighbor table. Using the information at the server table the peer S1 points to three peers P1, P2 and P3 whose latency, bandwidth, cache digest and reputation information is given in the neighbor table. So, when a request comes to S1, it points to these three peers from where the contents can be fetched.

Each node (peer) maintains its own neighbor table [2] and receives query only once, it saves the network from flooding effect. But, it's difficult to decide the value of latency/bandwidth of two new peers [2].

When a peer tries to contact with his neighbor peers from the neighbor table, if it gets a false hit which means that peer is dead or it doesn't responding then its reputation is decreased by x percentage. So the next time when a peer tries to contact that peer it considers its reputation.
Tuxedo adopts a simple approach to decide whether cached copy of transcoded content should be accessed

from neighbors or it should perform transcoding locally. [2]

The advantages we found for Tuxedo are:

- The use of "reputation". It deals with the status of the neighbor based on previous update.
- Use of CONCA Proxy Cache let the users to access the content using diverse device and connection technologies.
- Maintaining the server table along with the neighbor table decreases the problem of finding a peer's neighbor and its' information.
- By storing the peer's information locally it reduces the search request for any content thus reduces the response time and network traffic.
- Taking Bandwidth and Latency into consideration makes Tuxedo scalable and more efficient than other caching techniques.
- Hierarchical cache digest used by Tuxedo makes it easier to Transcode the contents.
- Memory space required to store information about the neighbors is small enough to keep all the data structure in memory.

The disadvantages of Tuxedo we found:

- Tuxedo uses piggybacking method to update information about peers which is not efficient for dynamic updates of information.
- When there is more than one centralized server, the information from one web site may experience different values at different time which makes the server table of Tuxedo unstable.
- The cache node has to decide between transcoding an object locally and fetching the transcoded object from a neighbor cache, which is a time consuming operation.

C. PeerOLAP

PeerOLAP is developed for distributed caching of supporting On-Line Analytical Processing (OLAP) queries. A large number of low-end clients, each containing a cache with the most useful results, are connected through an arbitrary peer-to-peer network. If a query cannot be answered locally, it is propagated through the network until a peer that has cached the answer is found [1].
The PeerOLAP network is a set of peers that have local caches and implements a mechanism for publishing their local caches' contents and their conceptual capabilities. The peers' access data warehouses [1] and pose OLAP queries. Other peers can connect to the peer P_i and request a result. P_i may either answer the query (or part of it) locally, if it has the required data, or propagate the query to its neighbors. In either case, all results return directly to the peer that initiated the query. The goal of PeerOLAP is to act as a combined virtual cache, where all the components offer resources aiming at achieving lower query cost.

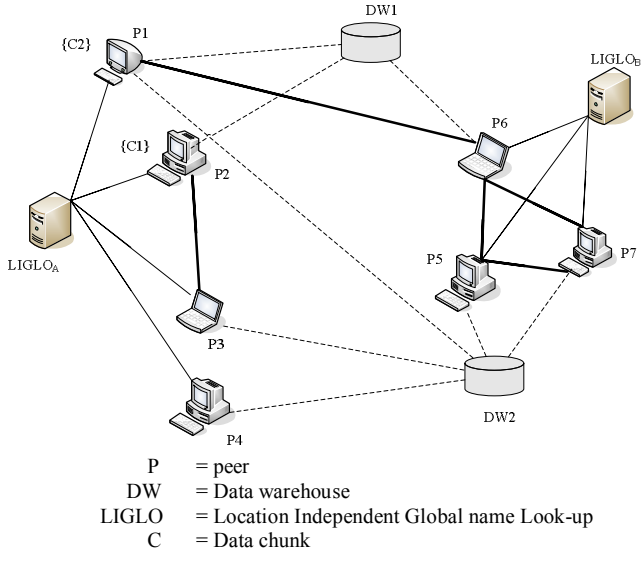


Fig. 3 A typical PeerOLAP network. Peer P2 issues query of referring to chunks C1, C2 and C3. C1 resides in P2; so it sends message requesting for C2 and C3 to neighbors P1, P3 and at the same time to another node P6. The peers who have any of the chunks, sends cost estimation of retrieving to P2. Comparing the costs, the peer is chosen for retrieving the chunk from. If the chunk is not found in the peers within the time of the expiration of the timer of the message, it is fetched from data warehouse (DW1 and DW2). For dynamic environment, LIGLO maintains the list of online peers and data warehouses along with their information to maintain administration over the network.

Fig. 3 depicts a typical PeerOLAP network and its working procedure – request messaging and object retrieving.

In order to avoid flooding the network with messages, a maximum number of hops [1] is assigned to each message. A message can't travel more than the hop. There is also a mechanism for breaking message loops: each peer keeps a queue of the recent messages and rejects the ones that had been processed before. After getting the object, instead of the whole object, a chunk of the object is kept according to the benefit value assigned to it. [1] describes the detailed algorithm.

A peer in the PeerOLAP needs a distinct architecture which is described in [1].

The advantages of PeerOLAP:

- Keeps the query cost low.
- Use query optimization techniques that determine which chunks should be requested from the warehouse, and which should be retrieved from the peers.
- It has caching policies that enable co-operation among caches and eliminate unnecessary replication of objects.
- It also uses re-configuration mechanisms that create virtual neighbors of peers with similar access patterns.
- Uses efficient mechanism for load control in peers and warehouses.
- Network flooding with message is controlled efficiently.
- In the peer architecture, by distinguishing the cache from the semantics of the data, the cache can store simultaneously data from multiple warehouses.
- For dynamic networks, administration is applied for keeping track of the online peers.

The disadvantages of PeerOLAP:

- Cached object in a peer may not be up-to-date because no updating mechanism is used.
- By using PeerOLAP architecture, we may get the query result with a low cost but it may take huge time to download the object because of the high latency and low bandwidth.
- The path of routing the object to the requesting peer may be a long one.
- Identifying the neighborhoods of peers with similar access patterns is a clustering problem because there is no complete knowledge about the whole network at any site.
- The whole object is not found in the peer, so it may be time consuming to search the whole object.
- The peers need to have a different architecture which may be costly.

III. Future work

We can summarize the above discussion into the following table:

Table 1 Comparison of different techniques.

	Criteria	Squirrel	Tuxedo	PeerOLAP
1	Object source choosing criteria	Random	Bandwidth and latency of the peer	Path cost
2	Cached object	Full object	Full object	`In chunks
3	Updating of object	Update when object is demanded	No update	No update
4	Cache	Browser's cache is shared for making the Squirrel cache	Memory	Memory
5	Transcoding	No	Yes	No
6	Multiple server (home for object) support	Yes	No	Yes
7	Adminis-tration over the network	Directory	Server and neighbour table	LIGLO
8	Download cost	May be high	May be high	Low
9	Scalability	Low	High	Low
10	Network overload control	No	Yes	Yes
11	Trust concerning scheme	No such scheme	Uses 'repu-tation' for trust concern amongst peers	No such scheme

From our findings above, we have found some efficient ways to improve the caching technique. We suggest the following criteria for efficient caching.

The cache will be shared from the memory of the peers. We have found it inefficient to share the cache of the browser.
The cached object will be stored in chunk; it ensures the availability of the peer for a longer time. It also takes less space in a peer's memory.

The peer or delegate choosing criteria will be based on bandwidth or latency. We can find out the fastest content-

residing peer in this way, although it has the drawback that it doesn't support multiple servers and the cost of fetching the object may be high.

The object will be updated dynamically. A time cycle will be assigned to each object according to which the object will be updated.

An efficient trust concerning scheme to find the peers' availability will be used. We suggest the scheme like using 'reputation' used in Tuxedo [2].

Dedicated server will be used for the administration over the network. The servers will keep information (latency, bandwidth, reputation, cache digest etc.) of the nodes and manage them dynamically such that the impact of the nodes coming in and going out of network is decreased. The table content will be updated dynamically.

The network overload will be controlled using the method in Tuxedo [2].

Using Oversim[11], we are working on to simulate our findings. We hope the results of the simulation will be helpful to build the next generation adaptive caching techniques for peer-to-peer systems.

References

[1] Panos Kalnis, Wee Siong Ng,Beng ChinOoi , Dimitris Papadias, Kian-Lee Tan. "An Adaptive Peer-to-Peer Network for Distributed Caching of OLAP Results," ACM SIGMOD conference 2002.

[2] Weisong Shi, Kandarp Shah, Yonggen Mao, and Vipin Chaudhary. "Tuxedo: A Peer-to-Peer Caching System," Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, June 2003.

[3] Anuj Maheshwari, Aashish Sharma, Krithi Ramamritham. "TranSquid: Transcoding and Caching Proxy for Heterogenous E-Commerce Environments," Research Issues in Data Engineering (RIDE) 2002: San Jose, USA.

[4] Sitaram Iyer, Antony Rowstron and Peter Druschel, "Squirrel: A decentralized peer-to-peer web cache," Proceedings of the Twenty-First Annual ACM Symposium on Principles of Distributed Computing, July 2002.

[5] Florence Cl´evenot, Philippe Nain, "A simple fluid model for the analysis of the Squirrel peer-to-peer caching system," The 23rd Annual Joint Conference of the IEEE Computer and Communications Societies, March 2004.

[6] Antony Rowstron, Peter Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," Middleware 2001: Heidelberg, Germany

[7] http://www.napster.com.

[8] http://www.gnutella.wego.com.

[9] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, "A Scalable Content-Addressable Network," ACM SIGCOMM Conference 2001.

[10] Weisong Shi, Vijay Karamcheti, "CONCA: An Architecture for Consistent Nomadic Content Access," Workshop on Caching, Coherence, and Consistency (WC3 '01), in conjunction with ACM ICS'2001, Sorrento, Italy, June 17, 2001.

[11] http://www.oversim.org

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Very Low Voltage High Duty Cycle Step-up Regulator

Mohiuddin Hafiz, Tania Ansari, Khondker Zakir Ahmed, Syed md. Jaffrey and Syed Mustafa Khelat Bari

Bangladesh University of Engineering & Technology.

E-mail: hafiz2431@hotmail.com

Abstract - A monolithic current mode CMOS DC-DC converter with new control topology and integrated power switches has been developed. System and circuit level implementation have been accomplished for the converter. Afterwards, this converter has been fabricated with a standard 0.5 µm pseudo-CMOS process. Besides having the remarkable features of complete discharge of the output voltage, pulse-by-pulse current limiting, pulse-skipping mode in light load operation etc. the control loop's stability of the converter is independent of compensating signal, unlike conventional current mode converters. The simulation and test data are in good agreement. The experiment results show that this converter can boost 0.8V to as high as 5V, which makes it suitable for low voltage application. The output ripple voltage is about 20 mV with a 4.7-µF capacitor.

I. Introduction

Power management ICs, like low voltage highly efficient DC-DC step-up converters find extensive applications in electronic equipments like MP3 players, PDAs, Digital Still Cameras, Portable Medical Equipments, Cordless Phones and so on. To make these equipments handy, the size and weight of the power modules need to be minimized. Moreover, the converters with specific loads should be much smaller than the batteries to enhance the runtime [1]. The obvious result is to focus on the CMOS implementation of low-power converters such that power management and mixed-signal circuitries can be fabricated on the same chip for low power applications.

Among the switching regulators, voltage mode DC-DC conversion was the initial approach, but it required a large number of external components for compensation and the design of the compensation network was too involved. The obvious result was the current-mode control which requires a much simpler compensation scheme compared to its voltage-mode counterpart. In current mode control topology, pulse width modulation (PWM) and pulse frequency modulation (PFM) are widely used due to their over-current protection, robust dynamic responses, and simplified voltage-loop compensator design [2]. In both cases, the inductor current is made to modulate pulse width in PWM or oscillator frequency in PFM for output voltage regulation. Current-mode has two feedback loops: an outer one which senses DC output voltage and delivers a DC control voltage to an inner loop which senses power transistor currents and keeps them constant. Fig.1 illustrates a simplified structure of a synchronous boost converter. The two switches SW1 and SW2 are turned ON alternately to have charge and boost phase of the converter. To have a regulated output voltage, a closed loop control mechanism is needed. The modulator serves that purpose. It senses the parameters like portion of output voltages or the current through the inductor or both and controls the loop operation by converting those signals to time domain signals (i.e. duty cycle) in such a way that a negative feedback is achieved to regulate the output voltage [3],[4]. The overall stability of the loop depends; to a great extent, on the stability of the modulator itself.

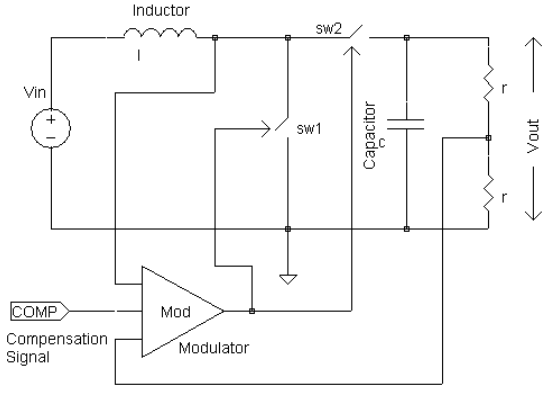


Fig.1. Synchronous step-up regulator with modulator to set close loop dynamics.

Among the several current-mode control techniques, constant-frequency current programmed mode (CPM) control, also known as peak current-mode (PCM) control, is a popular control technique for switch mode power converters. In this control scheme, power transistor peak currents are kept constant on a pulse-by-pulse basis. In spite of having most of the advantages of the current mode control scheme, the main disadvantage is found during the continuous conduction mode (CCM). This is because of the limited duty cycle that requires compensation by addition of an artificial ramp with the sensed current signal to extend the duty-ratio range beyond 50%. This issue is also known as 'Sub-harmonic Oscillation', as this occurs at half the switching frequency when the converter runs above the maximum allowable duty cycle, without being compensated [5]-[8]. A new topology of current mode synchronous step-up converter,

which can run at higher duty cycle without addition of any kind of compensating signal, has been developed and fabricated with a standard 0.5-μm pseudo BiCMOS process. The power switches, feedback and current sensing circuits are built on-chip. Off-chip elements are one inductor, one capacitor and a resistive divider. No other eternal compensating network is needed and hence the number of I/O pins is reduced. It is designed to serve low voltage applications like stepping up 1.5V to 3.3V or 5V with an output load of 100mA, which is suitable for electronic equipments powered by single cell battery.

II. Topological Features

The functional block diagram in Fig. 2 depicts the basic operational topology of the converter. This converter employs OFF time (Boost phase) modulation technique. The timing diagram in Fig. 3(a) explains the operational principle of the loop. The ON time (Charge phase) is kept constant for a fixed input voltage V_{in} , in such a way that this ON time is inversely proportional to V_{in}. Each charge phase is initiated by an error amplifier which senses the valley current through the inductor as the current sense signal and takes the Band-gap reference and feedback

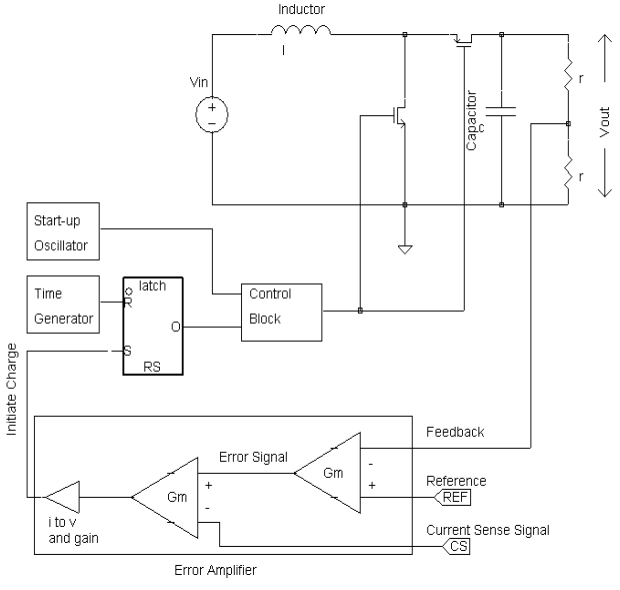


Fig. 2 The basic functional diagram of the step-up converter.

voltage as its input signals. The charge phase, however, will be terminated by a time generator circuit, which generates the required ON time for a fixed input voltage. Referring to Fig. 3(a), it is found that the current sense signal is proportional to the inductor current of the time T_{off} (i.e. during the boost phase). Besides, the difference between the reference and feedback voltages constitutes the 'error signal'. The 'current sense signal' and the 'error signal' are compared at their down slopes to yield the logic signal to initiate a charge phase. Here the current loop regulates the valley of the inductor current. The start-up oscillator starts the converter, when it is powered up by raising the output voltage to the point where the loop can regulate. This topology provides some inherent features, so that any sort of the current disturbances may die out within a cycle and does not grow any further. Hence, unlike the converter of the topology in [1], compensation of the current loop by the addition of an external ramp is not necessary.

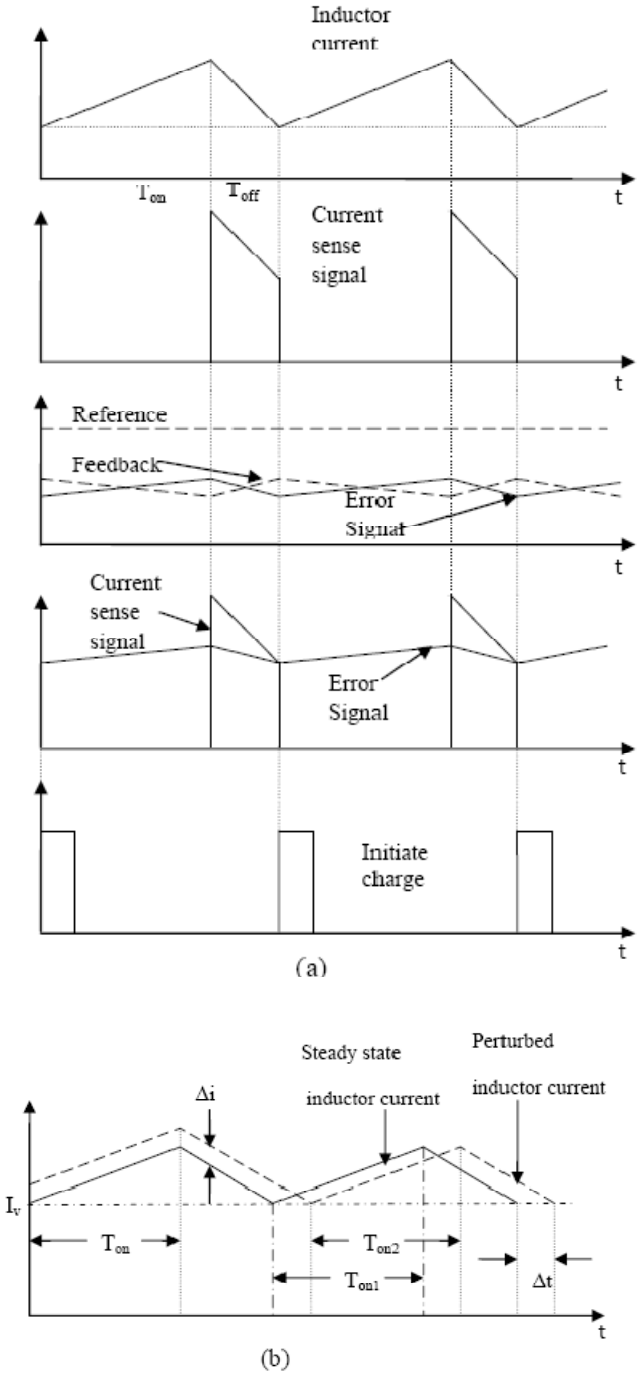


Fig. 3 The functional features of the converter: (a) operational principle of the converter and (b) topological effect on the current disturbances of the converter.

This fact is illustrated in Fig. 3(b). For some reasons, with a fixed input voltage i.e. V_{in}, a current disturbance Δi occurred (shown by dotted lines in Fig. 3(b)). As this is a valley current mode topology, the current loop will keep

the valley of the inductor current constant (I_v in Fig. 3(b)). As a result, the next cycle will be initiated only if the perturbed inductor current reaches the valley at its down-slope. Moreover, ON time is constant for a fixed DC input voltage ($T_{on}=T_{on1}=T_{on2}$) for this topology and hence the charge phase of the next cycle will terminate at that fixed ON time. As a result, the perturbed inductor current will just be a replica of the steady state inductor current with a Δt shift in time, in the next cycle. Thus the current disturbance that has got induced in the current loop is automatically eliminated by the topological features of the converter.

III. Circuit Implementation

The circuit implementation of the blocks responsible for the normal loop operation of the converter, as illustrated in Fig. 2, has been addressed here. The error amplifier along with the time generator block settles the loop dynamics. So the main focus is on the design issues of the two main blocks of the converter, along with the technique to sense the current signal.

A. The Error Amplifier with On-chip Current Sensing Technique

The Error amplifier is the heart of all the On-Off logics of the entire scheme. It is basically a open loop uncompensated comparator, which performs some sequential comparisons during the normal loop operation and yields a digital signal called 'charge' that initiates the charging cycle of the boost regulator. The basic structure of the error amplifier has been illustrated in Fig. 4. The output voltage is sensed by taking a portion of it, called feedback voltage, using an external voltage divider resistive network. This feedback voltage is compared with an internally generated bandgap reference voltage. On the other hand, unlike the conventional current-sensing

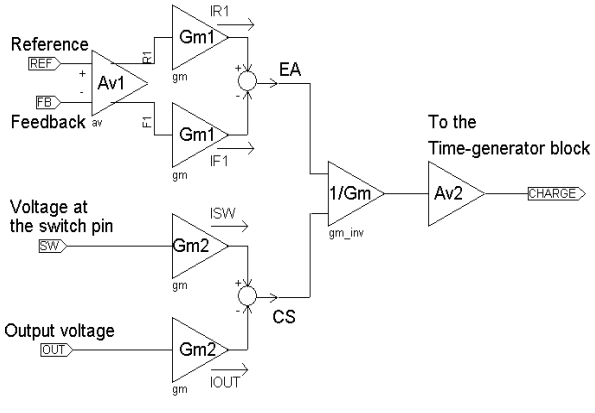


Fig. 4 The basic structure of the error amplifier of the converter.

techniques at [9],[10], the load current is sensed in the boost phase for this purpose; and with this end in view, the voltage across the PMOS switch is sensed in the boost phase, as shown in Fig.6 . The wave shape of the voltage across SW and OUT pins has been shown, which is actually the replica of the current through PSW, during the boost phase, as PSW acts as a voltage controlled resistor. These two voltages are converted to current signals by NCS and NCO and are made to beat with the current signals coming from other blocks.

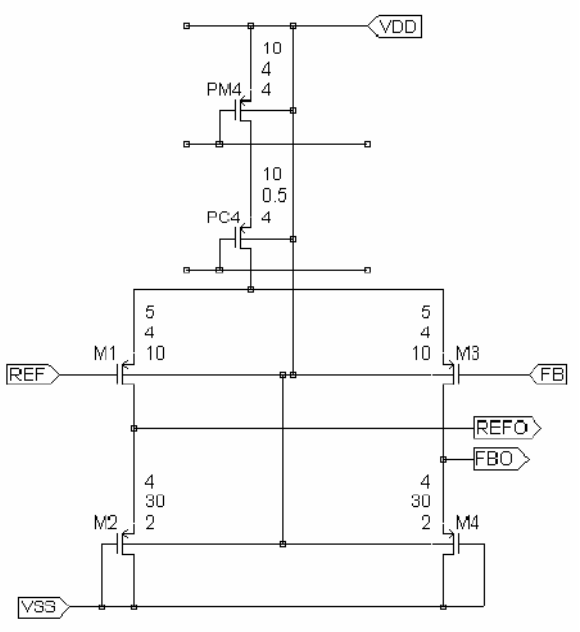


Fig. 5 The basic building block of EA_GEN

The crucial points in designing the error amplifier are to settle the relative speed and gain of the two loops, i.e. the voltage loop and the current loop, embedded within the

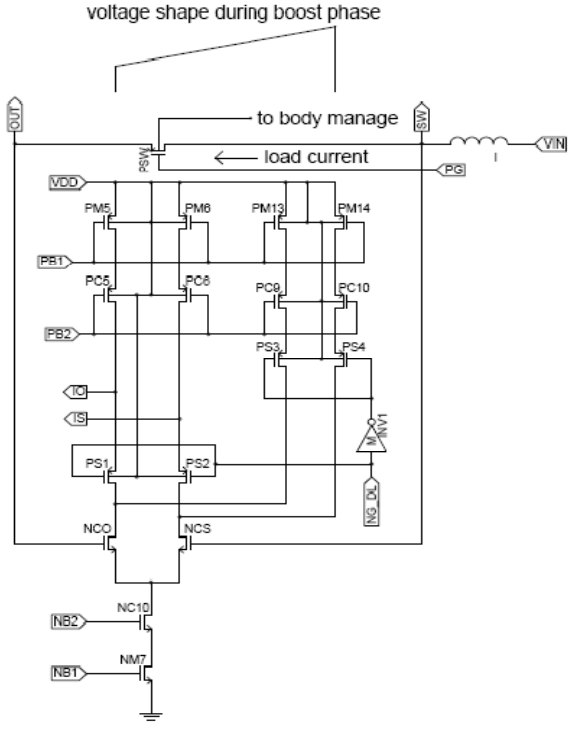


Fig. 6 The CS_GEN block with current sensing scheme

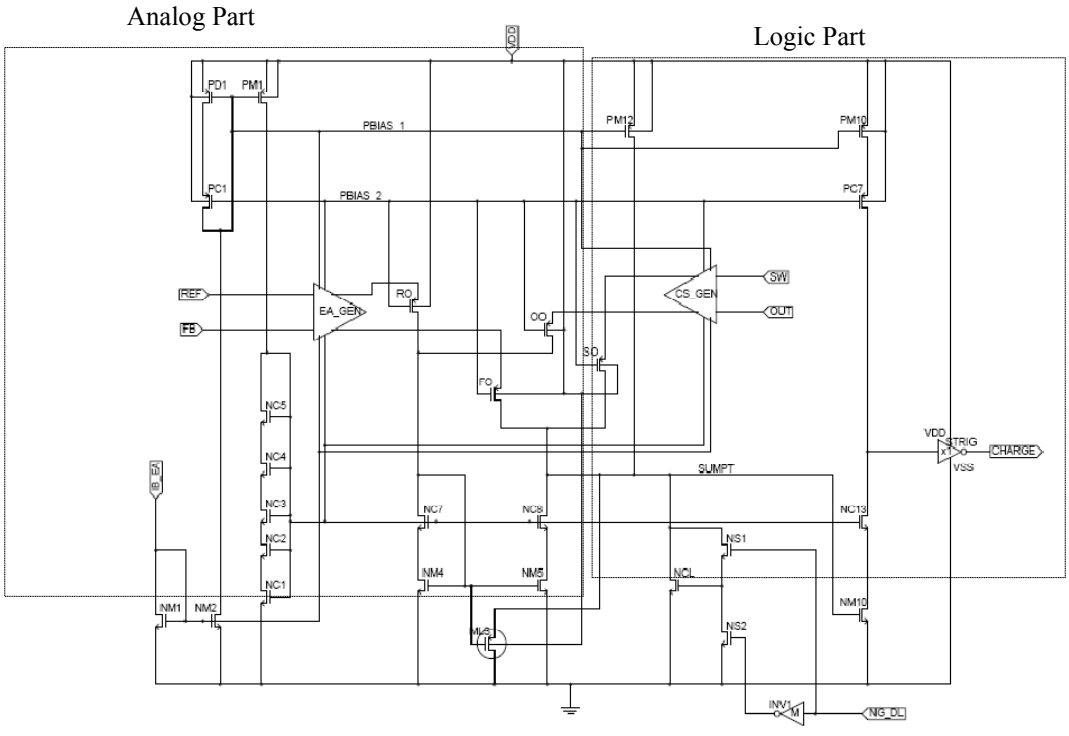


Fig. 7 The schematic of overall error amplifier

former. A dramatic simplification is obtained in stabilizing if the two loops to consider are mutually indifferent to each other. That is, one loop (current loop) is faster than the other (voltage loop) so that the slower loop does not see the faster loop. The relative gain requirements of the comparators are set by the differential values of the pairs to be considered. The gain portioning of the two loops has been settled such that for V_{in}=1.5, V_{out}=3.3, load=100mA, an opted regulation of 2% (20mV as the reference is 1V) is found. The differential value for current sense is the resistance of the PSW (0.5 Ω) times the current through it during the boost phase. The peak of the load current through PSW is 100mA + 190mA = 290mA and the valley is about 120mA at that condition. But at over load situation (say in start-up) the peak may go as high as 1A before the protective current limit scheme inhibits the error amplifier and breaks the current loop. Before that condition error amplifier has to work properly. Therefore the differential pair of the current sense comparator is set to work for voltage as high as 500mV. Hence a gain ratio of 25 is required, i.e. the difference at the voltage comparator needs to be gained up 20 times before it beats the signal from CS_GEN. Now we need to select such a topology of EA_GEN that its gain equation is independent of process, temperature and supply variation. That's why the topology shown in Fig. 5 has been chosen, whose gain is set simply by the (W/L) ratio of MOS M1 and M2 [11]. Then, finally, the comparison of the amplified version of the difference of reference and feedback signal with the difference of switch and out signal is accomplished. The whole comparison can be broken down into beating of currents at a node (SUMPT in Fig. 7). The output of CS_GEN is double ended differential currents in the folded cascode MOS devices (devices RO and FO). One of this current (RO) is steered back and compared with the other. At the same node the other folded cascode pairs (SO and OO) are also compared after steering back one of it (OO). These resultant currents beat each other at the summing node 'SUMPT'. This signal is provided a single stage gain by NM10 and ultimately the logic signal 'CHARGE' is generated.

B. The Time Generator Block

Time generator, shown in Fig.8, generates the ON time (for charge phase) that is inversely proportional to the input voltage and the minimum OFF time. Moreover, it controls the duration of the charge and the boost phases based on different conditions of the overall chip. Time generator can be divided into two parts; 1) Analog Part & 2) Logic Part. The basic principle of this time generation process depends on the well known property of a capacitor C, being charged by a current source of fixed current I, i.e. $\frac{I}{C}=\frac{\Delta V}{\Delta t}$. As it's evident from Fig.8, the current I is generated as, $I=\frac{V_{IN}-V_{BE}}{R_1}\Rightarrow I\alpha V_{IN}$.

This current is mirrored and multiplied by a factor m, by QN2, M2 and M3 and then made to charge the capacitor C1, when the logic 'CHR' is TRUE. Here R1 is 740K, C1 is 3.6pF and the ratio of M3 to M2 is 2:1. The capacitor is charged until QON is turned ON and thus $\frac{mI}{C}=\frac{V_{BE}}{T_{on}}$

$\Rightarrow T_{on}\alpha\frac{1}{I}$. Hence, we've the relationship, $T_{on}\alpha\frac{1}{V_{IN}}$,

which is one of the unique features of the loop in normal operation. When 'CHR' is FALSE, capacitor C1 discharges through NS1 and the collector of QON is pulled to supply, so that the next charge phase is not affected by stored charge and TON remains at a definite state. $T_{off\text{-}min}$ is fixed for the converter and hence it's generated from a fixed current ramping up the capacitor

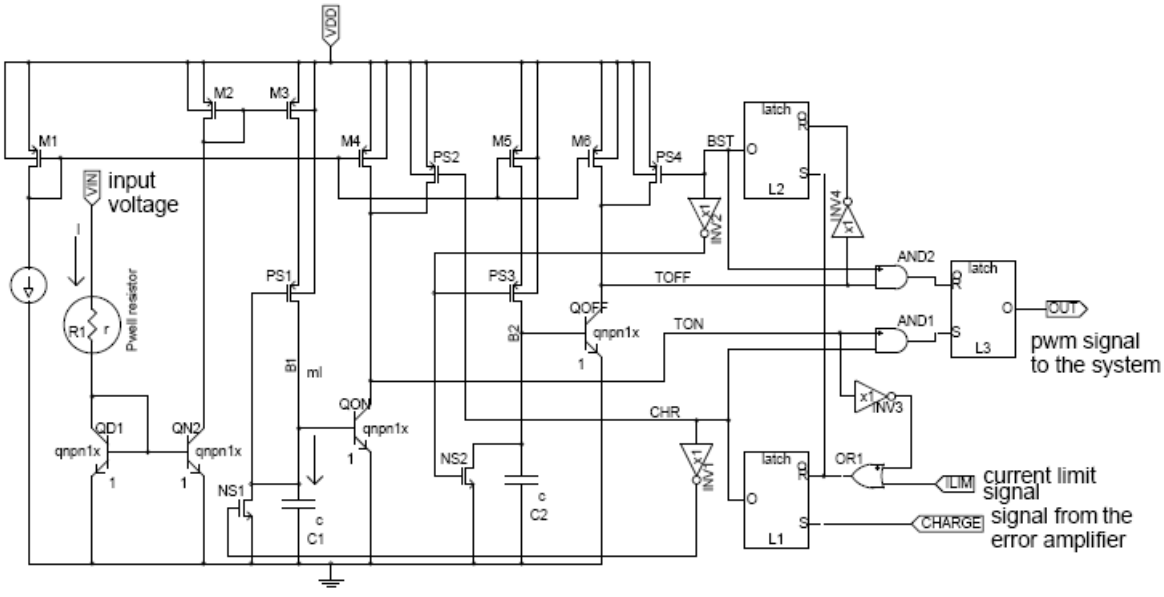


Fig. 8 The overall time-generator block, consisting of analog and logic parts.

C2 (C2=1pF), when 'BST' is TRUE. The analog block is so managed that, when there's no 'CHR' or 'BST' signal, TON or TOFF remains TRUE. When 'CHR' becomes TRUE, TON is pulled down after a specific span of time decided by the current *mI* and capacitor C1. Thus T_{on} is extracted as the duration of time when the signal TON remains TRUE after 'CHR' switches to TRUE.

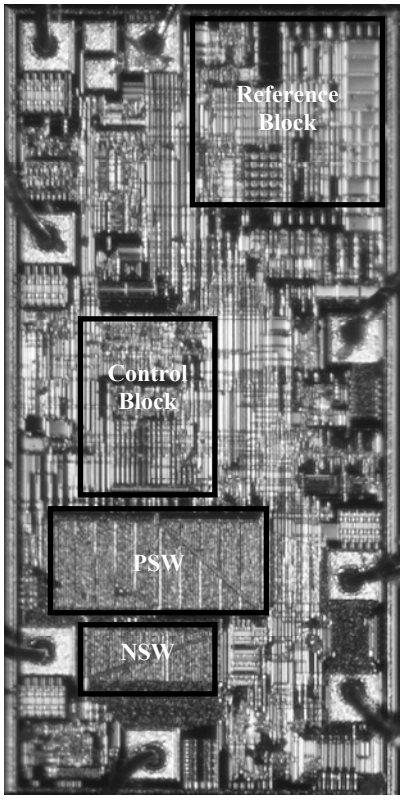


Fig. 9 The chip micrograph

The digital part consisting of some asynchronous latches and logic gates, converts the signal, generated discretely on 'TON' and 'TOFF' nodes, to the PWM signal as required by the system depending on all the prevailing conditions, e.g. the conditions when the current-loop breaks or when the peak current through the inductor reaches the currentlimit or when the converter runs at the discontinuous conduction mode etc.

IV. Measured data & real time waveforms

The DC-DC step-up converter has been fabricated with a standard (double metal, double poly) 0.5 μm Pseudo BiCMOS process. The chip micrograph (1.5mm×0.74mm) has been shown in Fig. 9. The charge time T_{on} is inversely proportional to input voltage V_{in} . However, to prevent the ON time from becoming very long, the time generator circuit has been biased with a fixed current, which is always present in charging the capacitor and this is evident from the flat nature of T_{on} vs. V_{in} curve, at the lower range of V_{in}, as shown in Fig. 10.

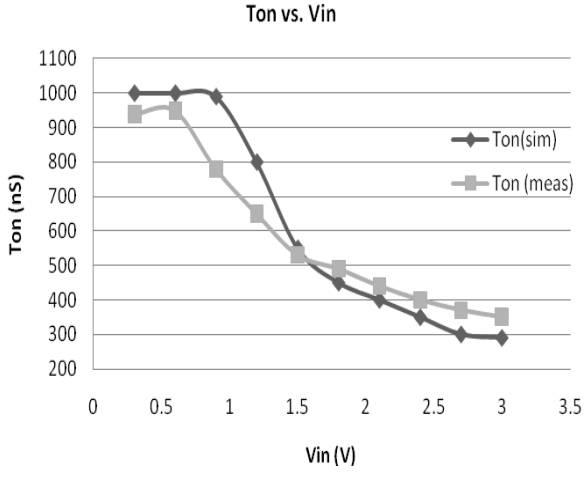
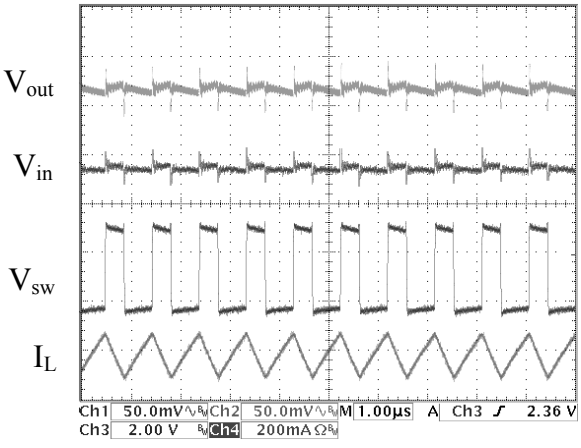


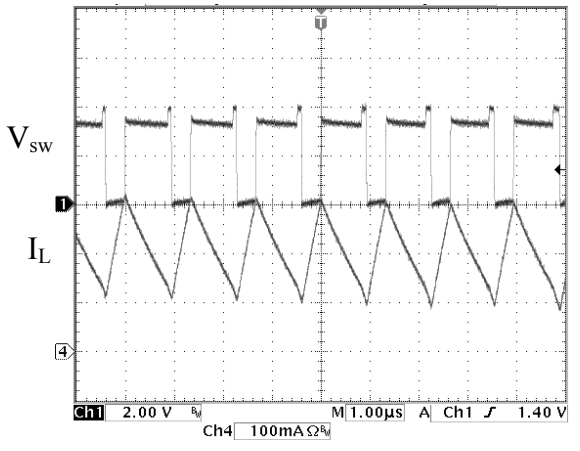
Fig. 10 Simulated & measured data of ON time

The steady state heavy load switching condition of the converter has been depicted in Fig. 11 (a). The chip has been set to supply a load of 100mA while boosting input voltage of 1.5V to 3.3V, i.e. running at 55% duty cycle. The designed converter can operate at lower duty cycle without facing any type of response speed related

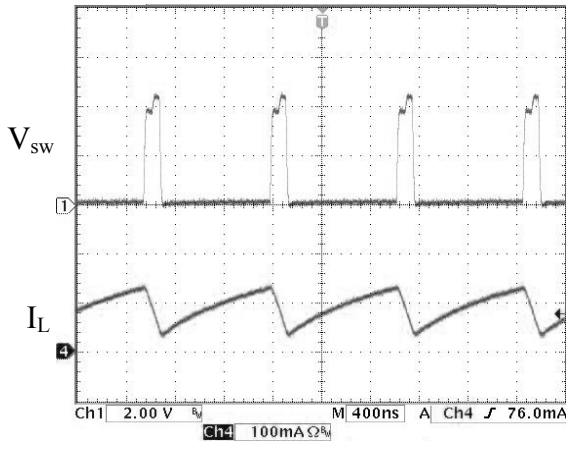
problem. Moreover, there is no sub-harmonic oscillation of the converter, at high duty cycle. The facts are illustrated in Fig. 11(b) and Fig. 11(c), respectively. Table 1 is the comparison of some of the features of the chip with that presented in [1].



(a)



(b)



(c)

Fig. 11 Waveforms of the converter at different operating conditions (a) steady state operation of the converter at V_{in}=1.5, V_{out}=3.3, Load=100mA; (b) low duty cycle (D=24%) operation at V_{in}=2.5, V_{out}=3.3, Load=115mA and (c) high duty cycle (D=84%) operation at V_{in}=0.8, V_{out}=5, Load=10mA.

Table 1 Comparison of different features

Parameters	[1]	This chip
Process	0.6 µm	0.5 µm
Topology	Fixed frequency	Variable frequency
Input voltage range	3 to 5.2 V	0.8 to 5V
Output current range	< 450 mA	< 250 mA

V. Conclusion

In this paper a low voltage high duty cycle step-up converter, that employs T_{off} modulation, has been developed. The converter doesn't need conventional slope compensation for its loop stability. Various issues relevant to the design and development of the chip have been discussed along with circuit level implementations of different functional blocks and measurement results, where the data shows that it can step-up 0.8V to 5V, making it suitable for low voltage application.

Acknowledgement

The authors are grateful to the top management of POWER IC LTD., the first semiconductor company of Bangladesh, for their support in developing the chip.

References

[1] F. Lee and P. K. T. Mok, "A monolithic current-mode CMOS DC-DC converter with on-chip current-sensing technique", IEEE J. Solid-State Circuits, vol.39, no. 1, Jan. 2004.

[2] R. Mammano, "Switching power supply topology: voltage mode vs. current mode," in *Unitrode Design Note DN-62*. Dallas, TX: Texas Instruments Incorporated, 1994.

[3] R. W. Erickson and D. Maksimovic, *Fundamentals of Power Electronics*. Norwell, MA: Kluwer, 2001.

[4] P. T. Krein, *Elements of Power Electronics*. New York, NY: Oxford Univ. Press, 1998.

[5] R. D. Middlebrook, "Topics in multiple-loop regulators and current-mode programming," IEEE PESC Record, pp. 716-732,1985.

[6] F. D. Tan and R.D. Middlebrook, "A unified model for current-programmed converters", IEEE Transactions on Power Electronics, Vol. 10, No. 4, July 1995.

[7] R. D. Middlebrook and S.M. Cuk, "A general unified approach to modelling switching converter power stages" , IEEE PESC Record, pp. 18-34, 1976.

[8] M. K. Kazimierczuk , "Transfer function of current modulator in PWM converters with current-mode control", IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, Vol. 47, No. 9, September 2000.

[9] P. Givelin, M. Bafleur, E. Tournier, T. Laopoulos, and S. Siskos, "Application of CMOS current mode approach to on-chip current sensing in smart power circuits," *IEE Proc. Circuits, Devices, Systems*, vol. 142,no. 6, pp. 357–363, Dec. 1995.

[10] M. Corsi, "Current sensing schemes for use in BiCMOS integrated circuits," *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, New York, NY, 1995, pp. 55–57.

[11] P.R. Gray, P.J. Hurst, S.H. Lewis, and R.G. Meyer, *Analysis and Design of Analog Integrated Circuits*, John Wiley & Sons, Inc. 4th Edition.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Design and Implementation of Semi-Quadratic Slope Compensation Circuit for PWM Peak Current Mode Boost Regulator

[1] Khondker Zakir Ahmed, [2] Syed Mustafa Khelat Bari, [3] Mohiuddin Hafiz and [4] Didar Islam

[1, 2, 4] Power IC Ltd., Dhaka, Bangladesh., [3] Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
E-mail: zakir.ak@gmail.com

Abstract – This paper presents a circuit implementation to fulfill the demand of the non-linear slope compensation to improve stability in a switching peak current mode DC-DC boost converter. The circuit is implemented by regular threshold voltage NMOS and PMOS and standard on-chip pico-Farad range capacitor biased by a PTAT current. Using the quadratic nature of the drain current of a MOS device in saturation the circuit generates a compensating slope signal on a cycle by cycle basis and this compensating slope signal is added with the current sense signal to compensate the disturbance that might arise. Along with the quadratic nature current another fixed current is added to compensate the noise that creates instability at lower duty cycle where the sense signal is weak. A chip is fabricated in 0.5µm technology using the proposed circuit. Simulation and test data has been presented.

I. Introduction

Peak Current Mode (PCM) Pulse Width Modulation (PWM) boost regulator is considered as an important integrated circuit for many low power mobile applications. Although current mode design offers many advantages over voltage mode design like built-in over-current protection, robust dynamic responses, simplified voltage-loop compensator design, rejection of input voltage disturbances, relatively simple current sharing for power modules operating in parallel and so on, one of the major problems of the design scheme is the inherent sensitivity to current disturbance while operating in continuous conduction mode (CCM) at a duty cycle greater than 50% [1], [2]. This sort of oscillation is also called as sub-harmonic oscillation, as this occurs at half the switching frequency when the converter runs above the maximum allowable duty cycle, without being compensated [2]-[5].The current disturbance which grows cycle by cycle if not compensated and makes the IC unstable. To enable the converter, operating in CCM, run at higher duty cycle, theoretically above 50%, a compensating ramp is added to current sense signal. Moreover to reduce the sensitivity to noise, the compensation ramp is commonly added in practical CPM (current programmed mode) designs, even when operating the converter at duty cycles less than 0.5 [6]. The compensation, commonly known as slope compensation, is theoretically expected to be non-linear with respect to duty cycle. Some work has already been published concerning the non-linear slope compensation method. One method proposes a output voltage independent second-order slope compensation technique for a buck converter [7]. Another work emphasised on piece wise linear slope compensation [8]. This paper specifically focuses on a PCM mode PWM technique boost converter stability. A semi-quadratic slope compensation signal generating circuit is proposed using which a chip is fabricated in 0.5µm technology and the stability is analyzed.

II. The Phenomenon of Instability

Basic ideas of the instability in the current loop in the absence of compensation and the remedy to it have been illustrated in this section. One of the ways the instability due to the sub-harmonic oscillation occurs, when a current disturbance gets induced in a converter running at high duty cycle, for a fixed input, as shown in Fig.1. For a fixed DC input voltage, if for some reasons there is an initial current disturbance ΔI_1, after a first down-slope the current will be displaced by an amount ΔI_2 .If the duty cycle is above 50% ($m_2 > m_1$ in Fig. 1), the output disturbance after one cycle ΔI_4 is greater than the input disturbance ΔI_3. This can be further explained from Fig 1. For a small current displacement $\Delta I1$, the current reaches the original peak value earlier in time by an amount dt, where $dt = \Delta I_1 / m_1$. ON the inductor down-slope, at the end of the ON time, the current is lower than its original value by an amount ΔI_2 which is defined by Eqn. (1),

$$\Delta I_2 = m_2 dt = \Delta I_1 \frac{m_1}{m_2} \qquad \text{----------(1)}$$

Now with m2>m1, the disturbances will continue to grow but eventually decay, giving rise to an oscillation. As mentioned earlier, this sort of oscillation is also called as sub-harmonic oscillation. This PCM (Peak Current Mode, as the peak current is regulated) control, however, is a widely used topology in industrial purposes and the above mentioned type of oscillation is eliminated by the addition of a compensating signal with slope Mc >0. The way in

which such compensating technique works, has been illustrated in Fig 2.

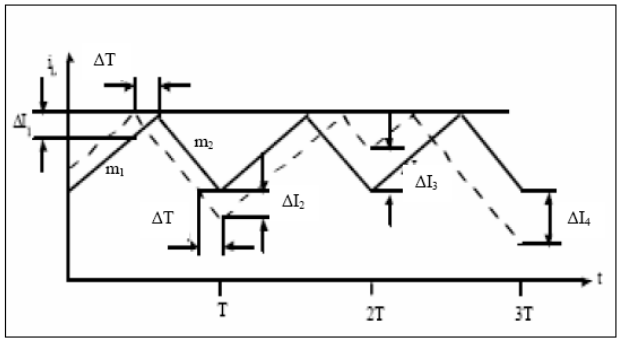


Fig. 1. Current loop oscillation for duty cycle > 50%

From a simple geometrical argument, the condition to prevent subharmonic oscillation is expressed in Eqn. (2):

$$\left|\frac{\Delta I_2}{\Delta I_1}\right| = \left|\frac{m_2 - M_c}{m_1 + M_c}\right| < 1 \qquad \text{----------(2)}$$

where, $m_1 = V_{in}/L > 0$ is the inductor current slope in the charge phase. Hence from the above equation, it is evident that, by selecting $M_c > m_2/2$, stable operation of the current loop under all steady-state conditions is possible to achieve.

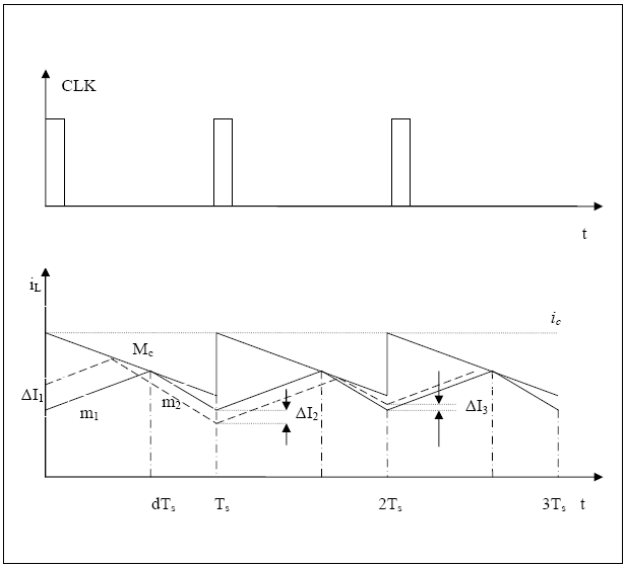


Fig. 2. Correction of oscillation by compensation ramp.

Besides, in [3],[9],[10] modulator model for the current programmed converters was derived, experimented and modified. A model for current-mode converter consists of two parts: 1) the model for the power stage, and 2) the model for the modulator [3]. The model of power stage in [11], [12] is now widely accepted, hence the actual task is reduced to the establishment of a modulator model. A modulator is expected to be expressed in terms of the average inductor current i_l, the average control current i_c and an effective voltage v_{eff}, which determines the slopes of the inductor current. Satisfying the conditions, a unified modulator model has been developed in [3] and the modulator gain F_m is expressed in Eqn. (3):

$$F_m = \frac{2L}{T_S v_{eff}\left[\left(\frac{2LM_c}{D' v_{eff}} + 2\right)D' - 1\right]} \qquad \text{----------(3)}$$

From the above equation it's evident that F_m approaches infinity for Eqn. (4)

$$D' = \frac{1}{\left(\frac{2LM_c}{D' v_{eff}} + 2\right)} \qquad \text{----------(4)}$$

The above equation indicates the minimum value of D' to maintain a finite positive value for F_m , expressed in Eqn. (5),

$$D'_{\min} \equiv \frac{0.5}{\left(1 + \frac{M_c}{D' v_{eff}/L}\right)} \equiv \frac{0.5}{\left(1 + \frac{M_c}{m_1}\right)} \qquad \text{---------(5)}$$

Hence, from the above equation, the current loop becomes unstable for min D' =0.5, if Mc=0, which implies that in the absence of Mc, the loop oscillates for $D_{max} = 1\text{-}D' = 0.5$. [3]

III. Design and Analysis of Circuit

The circuit is implemented using standard NMOS and PMOS devices. Fig. 3 shows the schematic view of the proposed slope compensation circuit.

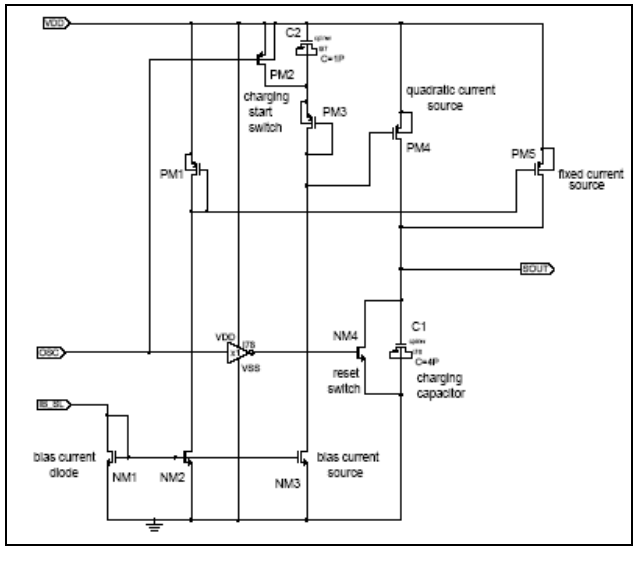


Fig. 3. Schematic of the proposed slope compensation circuit

The circuit is biased form an external PTAT current source. Inside the circuit, the current is mirrored and supplied to other nodes to bias different elements. The core of the circuit is the charging capacitor (C_1) which is charge by two types of current. One is quadratic in nature

and the other is a constant one. The quadratic current is generated by moving the gate voltage of a MOSFET linearly. The diode connected p-MOSFET (PM_3) shown at the upper portion of the circuit ensures that the quadratic current generating MOSFET is always in saturation. At the beginning of each oscillator cycle, the charging initiation switch (PM_2) turned off and the upper capacitor (C_2) is linearly charge up by the constant current sink (NM_3). As the source of PM3 drops linearly so does the gate of the PM_4 which is connected to the drain of PM_3. Linear variation of gate voltage produces quadratic current at the drain of PM_4 which is used to charge capacitor C_1. in addition another fixed current is also added at the same node of C_1. with the two charging currents the capacitor charges up in a semi-quadratic fashion voltage which is used as the slope compensating signal.

At the end of the ON-time, the oscillator signal goes low which turns ON the reset MOSFET (NM_4). The reset switch discharges the charged up capacitor C_1 and get it ready for the next cycle. Also the initiation switch PM_2 remains turned ON during oscillator OFF time and ensures the discharging of capacitor C_2. The reason for adding a fixed current to the quadratic one is to provide a fixed amount of slope compensation at lower duty cycle. Without the additional fixed current compensation at lower duty cycle when the quadratic slope generating current might be too low to provide minimum amount of compensation at lower duty cycle and the chip might go unstable. The slope compensation which is added by the fixed current provides a shield against noise at lower duty and remains the chip in stable region.

IV. Results and Discussion

Simulation result of the proposed block is shown in Fig. 4. The slope compensating signal is shown increasing non-linearly over the full TON-period. A linear slope compensation signal is shown in

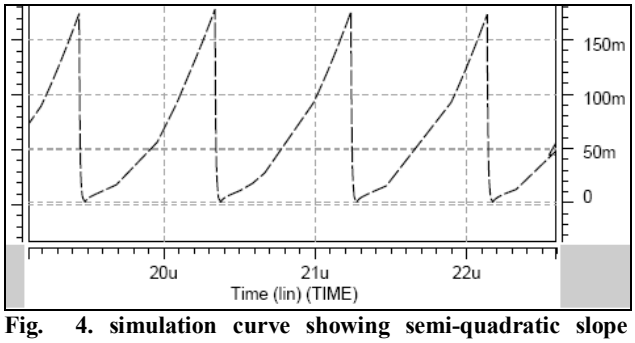


Fig. 4. simulation curve showing semi-quadratic slope compensating signal. Y-axis unit in Volts and X-axis unit in Seconds.

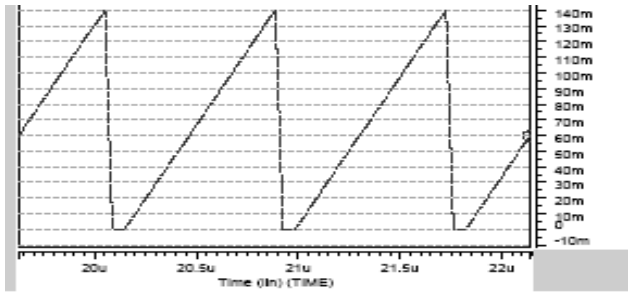
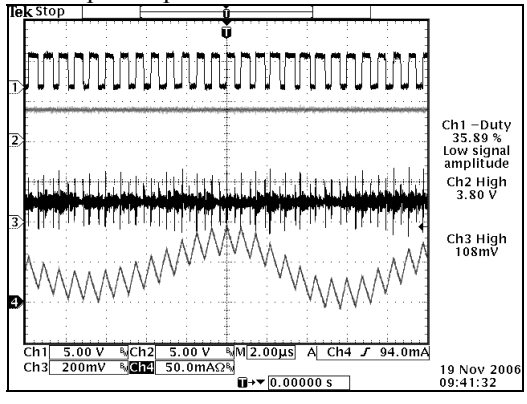
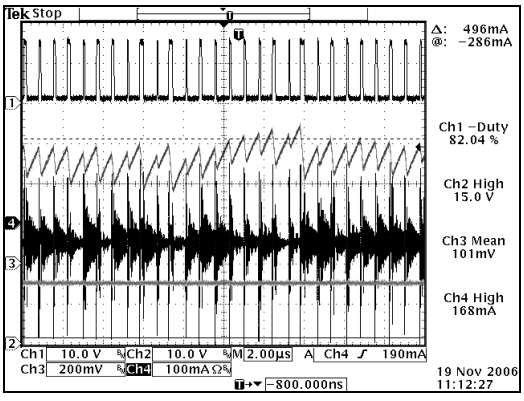


Fig. 5. simulation curve showing linear slope compensating signal. Y-axis unit in Volts and X-axis unit in Seconds.

The test data from the fabricated chip is presented below. The stability improvement due to the use of semi-quadratic slope compensation circuit is emphasised in the test data. First a chip that uses linear slope compensation is presented. The instability due to poor compensation at lower duty is shown in Fig. 6(a) and instability due to poor compensation at higher duty is shown in Fig. 6(b). This unstable version of chip employs linear slope compensation technique as shown in Fig. 5. As can be seen, the chip was unstable both in low duty and high duty cycle range. Instability was created due to the poor slope compensation signal strength. For linear increment the signal didn't get enough time to increase to provide necessary magnitude. In contrast to the unstable figures, we present test data from another chip fabricated using the proposed slope compensation circuit. The booster is found to be stable from duty cycle as low as 15% to as high as 85%. Corresponding oscilloscope snaps are shown in Fig. 7(a) and Fig. 7(b). The wide stable operating range is a significant evidence of better performance of the pseudo quadratic slope compensation circuit.

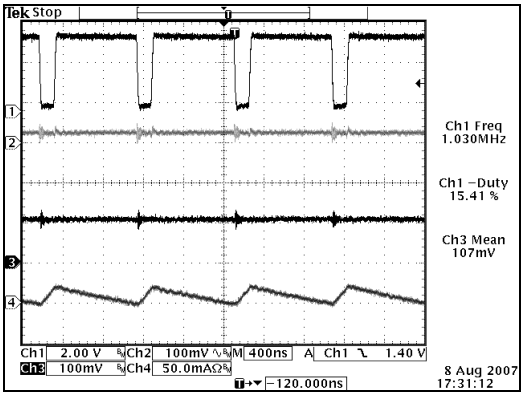


(a)

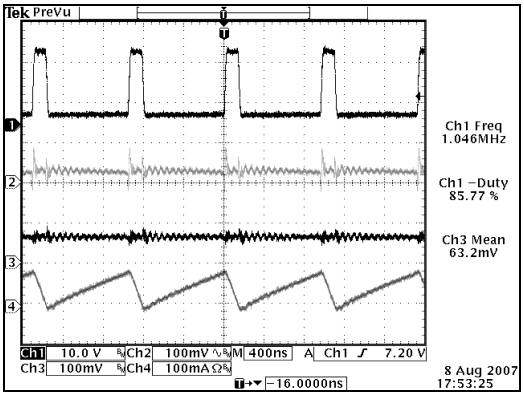


(b)

Fig. 6. Oscilloscope snaps of the poorly slope compensated booster operating at (a) low (below 50%) duty cycle and (b) high (above 50%) duty cycle. In both low and high duty cycle the booster is unstable which is obvious from the inductor current shape. Channel – SWITCH pin, Channel 2- VOUT, Channel 3 – FB, Channel 4 – inductor current



(a)



(b)

Fig. 7. Oscilloscope snaps of the rightly slope compensated booster implemented by proposed circuit operating at (a) 15% duty cycle and (b) 85% duty cycle . In both low and high duty cycle the booster is stable. Channel – SWITCH pin, Channel 2- VOUT, Channel 3 – FB, Channel 4 – inductor current

V. Conclusion

Submission in this paper we have presented a new proposal for a semi-quadratic slope compensation circuit for PCM – PWM DC-DC boost regulator. This proposed circuit is fabricated in 0.5µm technology. The test data of the fabricated chip shows very good result regarding the overall stability. The Chip fabricated with the proposed semi-quadratic slope compensation circuit operates in stable condition from very low duty cycle (15%) range to a very high duty cycle (85%) which was previously unstable when fabricated with linear slope compensation circuit.

References

[1] "Current Mode Control" Technical Paper 05, Venable Industries Inc.

[2] R. D. Middlebrook, "Topics in Multiple-Loop Regulators and Current-mode Programming," *IEEE PESC Record, pp. 716732,1985.*

[3] F.D. Tan and R.D.Middlebrook , "A Unified Model for Current-Programmed Converters", *IEEE Transactions on Power Electronics, Vol.10 ,No. 4,July 1995.*

[4] R.D. Middlebrook and S.M. Cuk, "A General Unified Approach to Modeling Switching Converter Power Stages", *IEEE PESC Record, pp. 18-34, 1976.*

[5] M.K. Kazimierczuk, "Transfer Function of Current Modulator in PWM Converters with Current-Mode Control", *IEEE Transactions on Circuits and Systems—I:Fundamental Theory and Applications, Vol. 47, No. 9, September 2000.*

[6] R.W. Erickson and D. Maksimovic, "Fundamentals of Power Electronics", *2nd Edition,Chapter 12 and Appendix B.3," Kluwer Academic Publishers, 2001.*

[7] H. Sakura and Y. Sugimoto, " Analysis and design of a current mode PWM buck converter adopting the output voltage independent second order slope compensation scheme", IEICE Trans. Fundamentals, vol.E88-A, no.2, pp.490-497, February 2005.

[8] L. Jiaying and W. Xiaobo "A novel piecewise linear slope compensation circuit in peak current mode control" IEEE conference on Electron Devices and Solid State Circuits, 20-22 Dec., 2007, pp 929-932.Sdfg

[9] G. C. Verghese, C. A. Bruzos, and K. N. Mahabir, "Averaged and sampled-data models for current mode control: a re-examination," *IEEE PESC Record, pp. 484491,1989.*

[10] C. P. Shultz, "A unified model of constant frequency switching regulators using multiloop feedback control," *PCIM Proc., 1993, (subsection 6 .3) pp .319-329.*

[11] R. Tymerski, V. Vorperian, C.Y. Fred and W.T. Baumann, "Nonlinear Modeling of the PWM Switch", *IEEE Transactions on Power Electronics, Vol. 4, No. 2, April 1989.*

[12] V. Vorperian, "Simplified Analysis of PWM Converters Using Model of PWM Switch, Part 1: Continuous Conduction Mode", *IEEE Transactions on Aerospace and Electronic Systems, Vol. 26, No. 3, May 1990.*

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Harmonic Mitigation in Transformers of Twelve-Pulse Rectifier Using Active Filter

Mohammad Rubaiyat Tanvir Hossain, Muhammad Quamruzzaman

Department of Electrical and Electronic Engineering
Chittagong University of Engineering and Technology, Chittagong-4349
E-mail: mrth94@yahoo.com, qzaman359@yahoo.com

Abstract – Researchers proposed several techniques to reduce the harmonics present in the input current of 12-pulse rectifier. But the transformers used in the rectifier contain more low order harmonics in their input currents. In this paper a novel active filtering method for input current wave shaping of the transformers used in a 12-pulse rectifier is proposed. With the proposed method both the rectifier and transformers are shown to draw near sinusoidal input currents from the utility. This active filtering technique uses high switching frequency PWM Boost converter that shapes the input currents by eliminating high frequency harmonic components with the aid of small size LC filters. The size of the LC filters is reduced to a considerable extent with the proposed technique. Detailed design and simulation results are presented to show the effectiveness of the proposed method.

I. Introduction

The advances in the power semiconductor devices have led to the increase in the use of power-electronic converters in various applications such as heating, lighting, ventilating and air conditioning applications, large rated dc drives and ac drives, adjustable speed drives (ASDs), HVDC systems, in process technology such as electroplating, welding units etc., battery charging for electric vehicles, power supplies for telecommunication systems etc [1-4]. And many of these use ac to dc conversion by various rectifiers. Rectifiers are non-linear circuit elements and generate harmonic currents. The non-sinusoidal harmonic currents drawn by the rectifiers are injected into the ac power lines /transformers /source causing a number of problems for the power distribution network and for other electrical systems in the vicinity of the rectifier deteriorating the power quality at the point of common coupling (PCC), thereby affecting the nearby consumers [5-7]. Consequently, design and development of rectifiers with improved waveforms has gained importance for stringent power quality regulation and strict limit on total harmonic distortion (THD) of input current placed by standards such as IEC 1000-3-2 and IEEE 519-1992 [8-10]. Among various methods, the most rugged, reliable and cost-effective solution is to use multi-pulse methods such as 12-pulse rectifier [11, 12]. The essence of these methods is to use multiple converters, which draw currents in a phase-staggered manner, resulting in the cancellation of certain harmonics. 12-pulse rectifiers are widely used in mid- and high-power applications such as large AC or DC drives, HVDC systems to achieve low input current harmonics [12, 13]. However, the 12-pulse rectifier input currents do not meet IEEE 519 harmonic standard without additional filtering [14, 15].Table 1 shows the harmonic components of supply current of a typical 12-pulse rectifier.

Table 1 Harmonic content of a typical 12 pulse rectifier

	12 pulse rectifier load	IEEE-519 Std.
5^{th}	3% - 6%	5.6%
7^{th}	2% - 6%	5.6%
11^{th}	5% - 9%	2.8%
13^{th}	3% - 8%	2.8%
THD	7.5% - 14.2%	7.0%

Also, the individual transformer windings carry non-sinusoidal currents containing large low order harmonics requiring over sizing of these transformers so that transformer windings are not overheated for a certain load. Harmonic filtering is thus needed for 12-pulse rectifier-utility interface to meet IEEE-519 harmonic current limits [14-16].

Several approaches were proposed to reduce input current harmonics of twelve-pulse-rectifiers. Dominant Harmonic Active Filter (DHAF) based on square-wave inverters switching at 5^{th} and 7^{th} harmonic frequencies, which are transformer coupled in series with 11^{th} and 13^{th} harmonic passive filters respectively is one method proposed to cost-effectively meet IEEE 519 harmonic current limits for 12 pulse rectifier loads [15]. Ref. [16] has proposed to eliminate harmonics drawn by a twelve-pulse rectifier through modulation of the dc bus. In addition to the increased component count, disadvantages of the proposed method include higher ripple current in the bridges and requirement of a controller for the modulator. Some approaches [17-20] have been found using and modifying inter-phase reactor results in near sinusoidal utility line currents. But the use of an inter-phase reactor is generally bulky and difficult to design. Autotransformer- based 12-pulse ac-dc converters have been reported [21- 25] for reducing the total harmonic distortion (THD) of the ac mains current, but the high dc-

link voltage, requiring inter-phase transformers and impedance-matching inductors, resulting in increased complexity and cost.

In this paper an active filtering scheme has been proposed to reduce the total harmonic distortion (THD) of input currents drawn by the transformers of a twelve-pulse rectifier. The Active filtering technique uses high switching frequency PWM Boost converter that shapes the input and transformer currents by eliminating high frequency harmonic components with small size filter. This technique has been used in single-phase and three-phase conventional rectifiers so far. No mention has been found in literature about the technique being used in 12-pulse rectifier. Also filter design has been done to eliminate the high frequency components in the current, which appears as a result of high frequency switching of the Boost converter stage.

II. Twelve-pulse AC-DC Rectifier

The schematic diagram of twelve-pulse rectifier is shown in Fig.1. The 12-pulse rectifier's input circuit consists of two six-pulse rectifiers, displaced by 30 electrical degrees, operating in series. A wye-wye and a delta-wye transformer have been used to give the necessary phase shift to produce the desired twelve-pulse output voltage and to cancel out the low-order harmonics. The turns-ratio of the wye-wye and delta-wye transformers are purposely chosen 1:1 and √3:1 respectively so that the peak output voltages of each transformer secondary are equal.

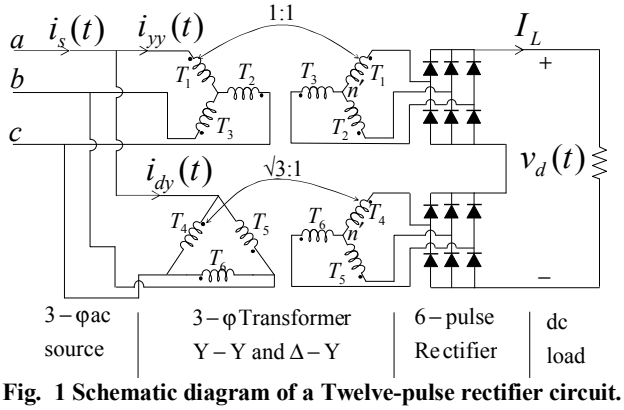


Fig. 1 Schematic diagram of a Twelve-pulse rectifier circuit.

The input currents i_{yy} and i_{dy} of wye-wye and delta-wye transformers respectively, as shown in Fig. 2 (a) & (b) are not sinusoids and have the following harmonic contents,

$$i_{yy}(t) = \frac{2\sqrt{3}}{\pi} I_L(\cos\omega t - \frac{1}{5}\cos 5\omega t + \frac{1}{7}\cos 7\omega t - \frac{1}{11}\cos 11\omega t + \cdots$$

$$i_{dy}(t) = \frac{2\sqrt{3}}{\pi} I_L(\cos\omega t + \frac{1}{5}\cos 5\omega t - \frac{1}{7}\cos 7\omega t - \frac{1}{11}\cos 11\omega t + \cdots$$

The resultant *ac* line current, i_S, as shown in Fig. 2(c), is given by the sum of the two Fourier series of the input currents of the star connected and delta connected transformers

$$i_S(t) = i_{yy}(t) + i_{dy}(t)$$

$$= 2\left(\frac{2\sqrt{3}}{\pi}\right) I_L(\cos\omega t - \frac{1}{11}\cos 11\omega t + \frac{1}{13}\cos 13\omega t - \frac{1}{23}\cos 23\omega t + \cdots$$

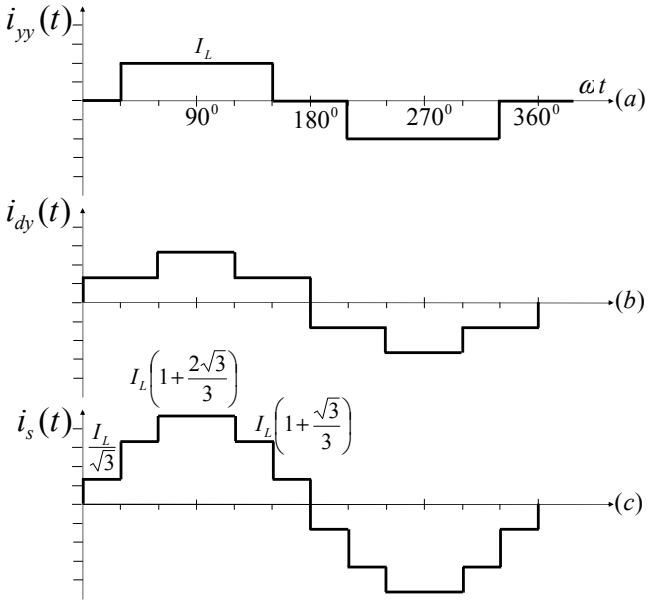


Fig. 2 Current waveforms of 12-pulse rectifier.

It is noted that the individual transformer windings of the 12-pulse rectifier circuit carry non-sinusoidal currents containing large low order harmonics such as 5th, 7th etc. This requires over sizing of these transformers so that transformer windings are not overheated for a certain load. In the resultant input *ac* current the 5th, 7th, 17th, 19th, etc. harmonics are eliminated and it contains 1st, 11th, 13th, 23rd, 25th, etc. harmonics. Consequently, the input line current for the twelve-pulse rectifier is close to sinusoidal waveform. However, 12 pulse rectifier front ends do not meet IEEE-519 harmonic standard without additional filtering. Thus filtering scheme is required to improve the performance of the twelve-pulse rectifier circuit.

III. Proposed Active Filtering Scheme

Fig. 3 shows the schematic diagram of the proposed active filtering scheme for the twelve-pulse rectifier configuration. In the proposed method a high switching frequency (5 KHz) PWM Boost converter is used at the output of the two six pulse rectifiers with ac LC filters at the 6-pulse rectifier inputs for reducing the total harmonic distortion of the input current and the transformer currents. The boost switch is turned on at constant frequency. A dc *L-C* filter is added at the output side in order to reduce the ripple content of the output voltage.

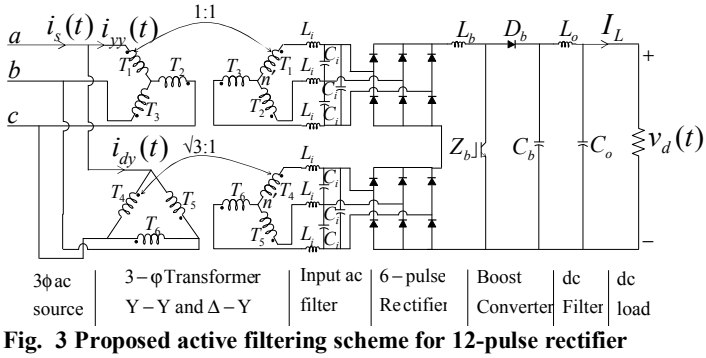


Fig. 3 Proposed active filtering scheme for 12-pulse rectifier configuration.

During the period when the boost switch is turned on, a symmetrical short circuit occurs at the rectifier input through the rectifier diodes, the boost inductor and the boost switch. Consequently the phase currents build up linearly at a rate determined by the input source voltages and the boost inductor, independently of each other, and the magnitude is proportional to the respective phase voltage amplitude. This means that the positive phase voltages cause positive currents through the upper diodes of each 6-pulse rectifier, which return as negative currents through the lower diodes that are caused by the negative phase voltages.

Again when the boost switch is turned off, the phase current through the boost inductor flows to the output capacitor decreases linearly at a rate determined by the input voltage, output dc voltage and the inductor. The input currents in all three phases contain the fundamental (50 Hz) component and a band of high frequency unwanted components centered around the PWM switching frequency of the boost switch. The discontinuous phase current pulses at high PWM frequency with the sinusoidal locus of the peak values can be filtered with a small *LC* filter to obtain a sinusoidal average current ideally.

A. PWM module design

Fig. 4 shows the schematic diagram of the PWM module which has been used to generate gating signals for switching the boost converter switch at varying duty cycles in order to enhance the continuity of the input current by providing it an alternate path through closing of the switch. The PWM module mainly consists of an opamp, an opto-coupler and BJT. The gate pulses are generated by comparing a saw-tooth wave with a reference dc voltage. Changing the reference voltage changes the duty cycle. The opto-coupler is used to provide necessary ground isolation between the PWM module and the switch while producing the pulses. The BJT amplifier is connected for increasing the voltage level at about 10 volts to drive the switch.

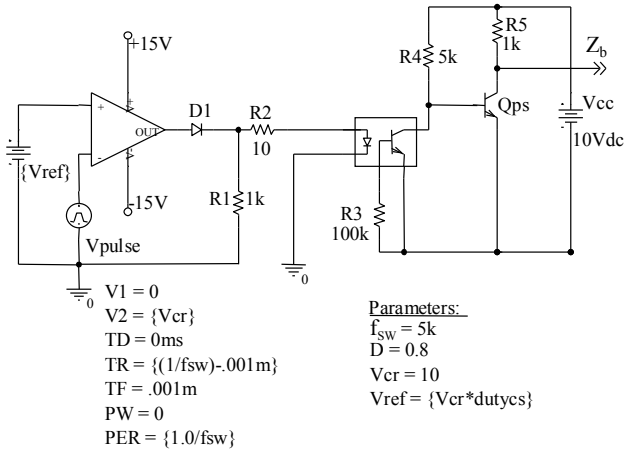


Fig. 4 Schematic diagram of PWM module for Active Filtering Scheme.

B. Output Filter design

A simple dc *LC* filter is used to reduce the 12-pulse ripple content of the output voltage of 12-pulse rectifier. Considering only the harmonic components, the equivalent circuit of rectifier with dc *LC* filter can be found as shown in Fig. 5.

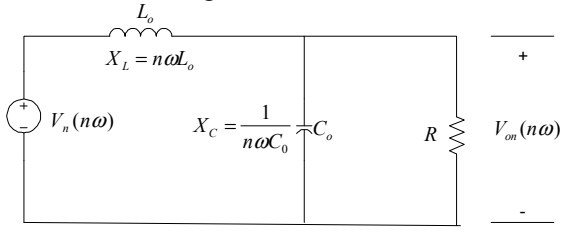


Fig. 5 Equivalent circuit for voltage harmonics.

The amount of reduction in the ripple voltage can be estimated as

$$\frac{V_{on}}{V_n} = \left| \frac{1}{1-(2\pi f_r)^2 L_o C_o} \right|$$

Where V_n is the ripple voltage before filtering, V_{on} is the ripple voltage after filtering, and f_r is the ripple frequency. Considering the ripple voltage to be reduced to 1% after filtering, the L-C constant can be found as $L_o C_o = 6.97 \times 10^{-6}$. Again, for the *n*th harmonic ripple current to pass through the filter capacitor, the capacitance value should be so chosen that the load impedance must be much greater than that of the capacitor. That is, $R >> \frac{1}{2\pi f_r C_o}$. Considering a dc load of 750Ω if $C_o = 1000\mu F$, the value of L_o is found as $6.97mH \approx 7mH$.

C. Input Filter design

To find the values of an *LC* input filter to limit the amount of input ripple current, considering only the harmonic components, the equivalent circuit per phase for the *n*th harmonic component of the rectifier system is given in Fig. 6.

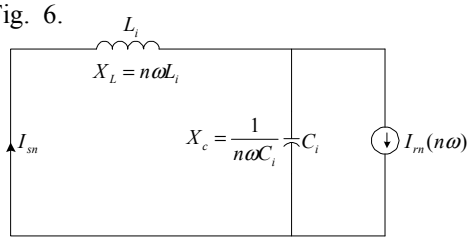


Fig. 6 Equivalent circuit for harmonic current.

Due to active filtering action all three input ac currents consist of the fundamental (50 Hz) component and a band of high frequency unwanted components at $f_{sw} - f_i = 4.95KHz$ and $f_{sw} + f_i = 5.05KHz$ centered around the PWM switching frequency, $f_{sw} = 5KHz$, where $f_i = 50Hz$ is the supply frequency. To reduce this unwanted high frequency components to 1% at the supply and transformers, the value of filter constant $L_i C_i$ is found as -

$$L_i C_i = \frac{1}{(n\omega)^2}\left[\frac{I_{rn}}{I_{sn}} - 1\right] = \frac{1}{(2\pi \times 4.95 \times 10^3)^2}\left[\frac{100}{1} - 1\right] = 10234 \times 10^{-9}$$

Where I_{sn} is the rms value of the *n*th harmonic current appearing in the supply and I_{rn} is the rms value of the nth harmonic current of the rectifier.

Without active filtering action, the 11th order harmonic is considered to be dominant. So to reduce the total harmonic distortion (THD) of the input line current to 1% and thereby reducing the THDs of transformer currents using only passive filter or ac input filter, the value of $L_i C_i$ can be found as $L_i C_i = 6.42 \times 10^{-7}$.

IV. Simulation Results

The proposed scheme is simulated using Orcad at different duty cycles. Fig. 7 shows the simulated waveforms of supply and transformer currents of 12-pulse rectifier without filter and Fig. 8 shows those with active filtering scheme under the following conditions:

Input phase voltage (peak value) = 300 V
Input frequency = 50 Hz
Duty cycle, $D = 0.4$
PWM switching frequency = 5 KHz
Input filter inductance = 5 mH
Input filter capacitance = 20 μF
Boost inductance = 1 mH
Boost capacitance = 10 μF
Output filter inductance = 7 mH
Output filter capacitance = 1000 μF

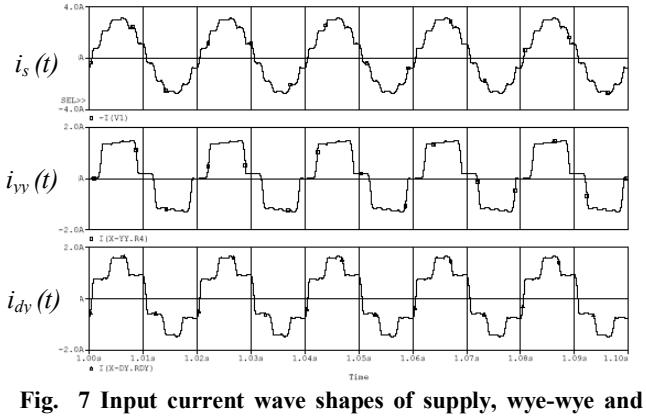


Fig. 7 Input current wave shapes of supply, wye-wye and delta-wye transformer of 12-pulse rectifier without filter.

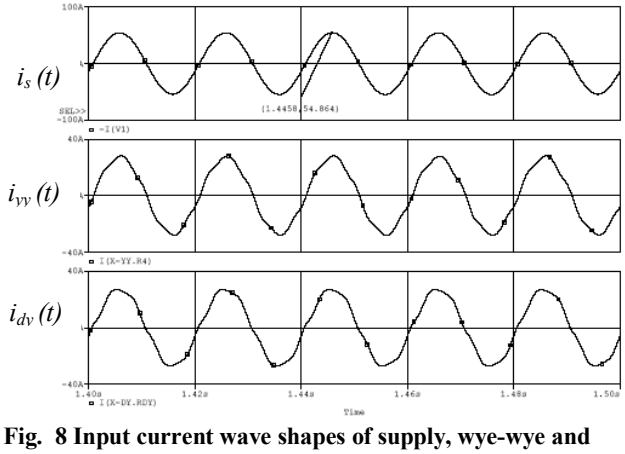


Fig. 8 Input current wave shapes of supply, wye-wye and delta-wye transformer of 12-pulse rectifier with the active filtering scheme.

From the simulation results it is found that with the proposed active filtering scheme the supply current, I_s and the transformer currents I_{yy} and I_{dy} approach sinusoidal wave shape. The total harmonic distortions (THD) of these current waveforms are significantly reduced for smaller input filter elements Li=5mH and Ci=20uF.

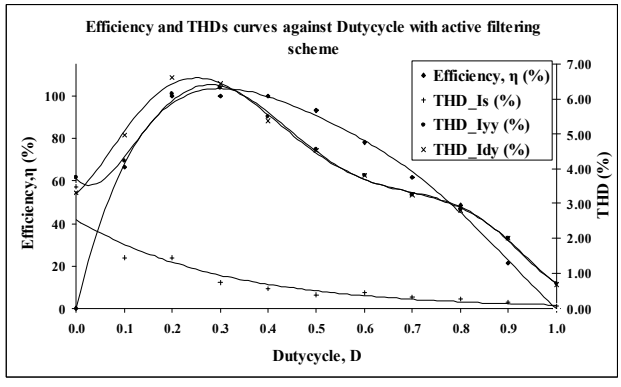


Fig. 9 Efficiency and THD curve against Duty cycle for the 12-pulse rectifier circuit with active filtering scheme.

The curves shown in Fig. 9 summarize the results showing the efficiency and THD curve at different duty cycles. It is found that duty cycle in the range of 0.2 to 0.5 gives the high efficiency (above 90%) for the module and the highest of 100% efficiency is found at 0.2 to 0.4 duty cycle. The THD of input current decreases exponentially as the duty cycle increases and THDs of the transformer currents are about 6% at 0.2 and 0.4 duty cycle and thereafter THD decreases gradually. At the highest efficiency of 100% the THDs of supply current and wye-wye and delta-wye transformer currents have been found 0.55%, 5.48% and 5.37% respectively.

V. Discussion

For harmonic mitigation in transformers of a twelve-pulse rectifier using active filter the performance of both passive and active filtering schemes have been studied through simulation. The simulation results and different performance parameters including total harmonic distortion (THD), efficiency etc. obtained thereof for different filter combinations at different positions have been compared in order to find a better harmonic mitigation method. Fig. 10 shows the THDs (%) of input currents in 12-pulse rectifier without and with different filtering schemes at respective maximum efficiency point.

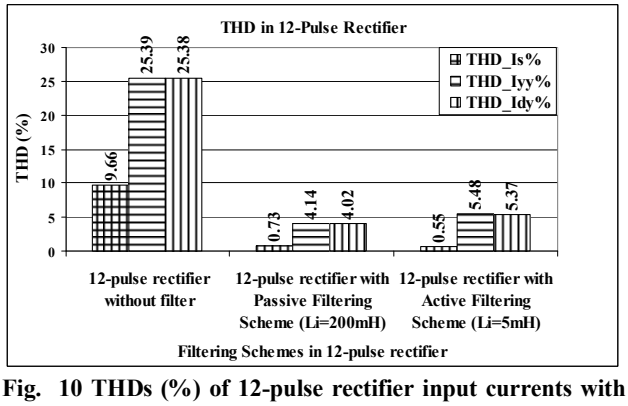


Fig. 10 THDs (%) of 12-pulse rectifier input currents with different filtering schemes at respective maximum efficiency point.

From the bar chart it is found that in 12-pulse rectifier without input-output filter the THDs of the input and transformer currents are about 9%, 25% and 25% respectively which do not meet IEEE-519 harmonic standard. Using passive filters at the rectifier inputs the minimum value of ac input LC filter inductance required is $L_i = 200\,mH$ for fixed $L_i C_i = 6.42 \times 10^{-7}$ to get the maximum rectifier efficiency of 99.06% and lower total harmonic distortion of the supply current and wye-wye and delta-wye transformer currents of 0.73%, 4.136% and 4.02% respectively. With active filtering scheme using single boost converter with ac input LC filter inductance $L_i = 5mH$ and capacitance, $C_i = 20\mu F$ the THDs of supply current and wye-wye and delta-wye transformer currents have been found 0.55%, 5.48% and 5.37% respectively at the highest efficiency of 100% at 0.4 duty cycle. Thus Incorporating active filtering scheme considerably reduces THDs of input and transformer currents with smaller size of ac input filter compared to passive filtering scheme. The passive filtering scheme would require 40 times higher value of inductance that would be more weightier than that of active filtering scheme to get lower THDs of input currents and highest efficiency of the rectifier circuit.

VI. Conclusion

In this paper an active filtering scheme using high switching frequency PWM Boost converter has been designed to reduce the total harmonic distortion (THD) of input currents drawn by the transformers of a twelve-pulse rectifier. Performance evaluation of the proposed scheme has been carried out and compared with the passive filtering scheme under similar conditions. It is evident that the proposed active wave shaping technique is a simple, cheap and efficient method for harmonic mitigation in transformers of 12-pulse rectifier. However, there are opportunities of extending this work in future to meet other goals. The proposed active filtering scheme may be implemented practically to investigate its actual potential. Investigation can be extended to regulate the output voltage. Investigation may also be made to see how twelve-pulse rectifier with proposed filtering scheme performs under actual operating conditions with unbalanced input line voltages.

References

[1] Karvelis, G. A., Manias, S. N., Kostakis, G., "A comparative evaluation of power converters used for current harmonics elimination", in Proc. IEEE HQP'98, pp. 227-232, 1998.

[2] Erickson, R. W., Fundamental of Power Electronics. New York: Chapmann & Hall, 1997.

[3] Bose, B. K., "Recent advances in Power Electronics", IEEE Trans. Power Electron., vol. 7, no. 1, pp. 2-16, Jan. 1992.

[4] Prasad, A. R., Ziogas, P.D., and Manias, S., "A comparative evaluation of SMR converters with and without active input current wave shaping", IEEE Trans. Ind. Electron., vol. 35, pp. 461-468, Aug. 1988.

[5] Bashi, S.M., Mariun, N., Noor, S.B., and Athab, H.S., "Three-phase Single Switch Power Factor Correction Circuit with Harmonic Reduction", Journal of Applied Sciences, vol. 5, no. 1, pp. 80-84, 2005.

[6] Redl, R., Tenti, P., Wyk, J. D. V., "Power electronics polluting effects", IEEE Spectr., vol. 34, no. 5, pp. 32-39, May 1997.

[7] Wyk, J. D. V., "Power Quality, Power Electronics and Control", in proceedings EPE. 93, pp. 17-32, 1993.

[8] "IEEE Recommended Practice and Requirements for harmonic Control on Electric Power Systems" IEEE Std. 519, 1992.

[9] Oscar, G., Jose, C., Roberto, P., Pedro, A., and Javier, U., "An alternative to Supply DC Voltages with High Power Factor", IEEE Trans. on Industrial Electronics, vol. 46, no. 4, pp. 703-709, August 1999.

[10] Yang Z. and Sen P. C., "Recent Developments in High Power Factor Switch –mode Converters", in IEEE proceedings CCECE 98, pp. 477-480, 1998.

[11] Paice, D.A., "Calculating and controlling harmonics caused by power converters", in Proc. IEEE Ind. Appl. Soc. Annual Meeting, vol. 1, pp. 456-463, 1989.

[12] Paice, D.A., "Power Electronic Converter Harmonics: Multipulse Methods for Clean Power", Piscataway, NJ: IEEE Press, 1996.

[13] Mohan, N., Udeland, T., Robbins, W., Power Electronics: Converters, Applications and Design. 3rd ed. New York: Wiley, 2002.

[14] Erickson, R. W., "Some Topologies of High Quality Rectifiers", Keynote paper, First International Conference on Energy, Power and Motion Control, May 5-6, 1997, Tel Aviv, Israel.

[15] Cheng, P. T, Bhattacharya, S., Divan, D. M., "Application of dominant harmonic active filter system with 12 pulse nonlinear loads", IEEE transactions on power delivery, vol. 14, no. 2, pp. 642-647, Apr. 1999.

[16] Raju, N. R., Daneshpooy, A., and Schwartzenberg, J., "Harmonic Cancellation for a Twelve-Pulse Rectifier using DC Bus Modulation", IEEE IAS Annual Conference Record, pp. 2526 – 2529, 2002.

[17] Miyairi, S., Iida, S., and Nakata, K., ''New method for reducing harmonics improved in input and output of rectifier with interphase transformer," IEEE Trans. on Ind. Appl., Vol. IA-22, NO. 5, pp.790-797, 1986.

[18] Choi, S., Enjeti, P. N., Lee, H., and Pitel, I. J., "A new active interphase reactor for 12-pulse rectifiers provides clean power utility interface," IEEE/IAS Anna. Meeting, pp.2464-2474, 1995.

[19] Tanaka, T., Koshio, N., Akagi, H., Nabae, A., "A novel method of reducing the supply current harmonics of a 12-Pulse Thyristor Rectifier with an Interphase Reactor", IEEE, pp. 1256-1262, 1996.

[20] Choi, S., Enjeti, P. N., Lee, H. H., Pitel, I. J., "A new active interphase reactor for 12-pulse rectifiers provides clean power utility interphase", IEEE Trans. on Ind. Appl., Vol. 32, No. 6, pp. 1304-1311, Novemver/December, 1996.

[21] Paice, D. A., "Multipulse converter system", U.S. Patent 4 876 634, Oct 24, 1989.

[22] Hammond, P.W., "Autotransformer", U. S. Patent no. 5 619 407, Apr. 8, 1997.

[23] Paice, D. A., "Transformers for Multipulse AC/DC Converters", U. S. Patent no. 6 101 113, Aug. 8, 2000.

[24] Kamath, G. R., Runyan, B., and Wood, R., "A compact autotransformer based 12-pulse rectifier circuit", in Proc. IEEE IECON conf., pp. 1344-1349, 2001.

[25] Singh, B., Bhuvaneswari, G., Garg, V., "Harmonic mitigation using 12-pulse AC-DC converter in vector-controlled induction motor drives", IEEE Trans. on Power Delivery, vol. 21, No. 3, pp. 1483-1492, July 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Design and Analysis of a Resonant Inverter fed from a Cûk Converter for the Conversion of Alternative Energy directly into Commercial Supply Efficiently.

Munshi Mahhbubr Rahman
Instructor Class A
Dept of Electrical Electronic and Communication Engg
Military Institute of Science & Technology
Mirpur Cantt, Dhaka-1216
Email: mahbub549@yaoo.com

Aminul Hoque
Professor and Head of dept
Dept oe Eelctrical and Electornics Engg
Bangladesh University of Engineering and Technology
Dhaka-1000
Email: aminulhoque@eee.buet.ac.bd

Abstract— Power crisis in the world has compelled mankind to think for alternative energy. But all the alternative sources have not yet been exploited. From many ongoing researches, it reveals that if all the alternative energy sources are exploited, the energy crisis of the world would be reduced to a great extent. Many alternative energy technologies today are well developed and they are reliable and cost competitive with the conventional fuel generators. There are many alternative sources of energy such as biomass, wind, solar, minihydro and tidal power. The most important advantage offered by alternative energy sources is their potential to provide sustainable electricity in areas not served by the conventional power grid. Most of the alternative energy technologies produce DC power, and hence power electronics and control equipment are required to convert the DC power into AC power.

Industry professionals are now opening their minds to alternate energy sources that can deliver quality power at reduced life-cycle costs in a variety of applications. Alternate energy sources generate power independently of the electrical grid. These sources can operate in stand-alone mode or they can be interconnected directly to the grid. With the use of power electronics, the alternative energy may be directly converted into commercial supply effectively in a cost effective way.

***Keywords*—**Resonant inverter, Cûk converter, Customised module, PWM, Fuel Cell, Snubber etc.

I. Introduction

There are areas in the country where grid electrical power is unavailable due to its geographical location and terrain. These areas have a variety of terrain like riverines, hilly areas and coastal islands. There is hardly any item commercially available which can be effectively used for those areas to run the conventional commercial equipments, except few commercial DC items which, not specially designed for the requirements of those remote areas [1].

In this research, an alternative power source (Solar system, fuel cell, wind power etc.) at 12V or 24V DC is assumed to be available in remote area. A power supply unit is to be designed and fabricated to provide conventional AC voltage at 50/60 Hz, so that the commercially available electrical and electronic equipments may be used in remote installations without specially made customised modules[2].

Commercially available electrical/electronic equipments may be used in remote areas are so chosen to obtain desired AC voltage to run conventional communication equipment. Step up DC-DC conversion will be achieved by duty cycle control of Cûk converter topology[3]. Cûk topology is the most efficient and light weight unit among all switch mode power supplies. Cûk converters have both voltage step down and voltage step up capability. In this work, the converter will be operated in step up operation. The output of Cûk converter will be fed to a resonant inverter circuit to obtain direct sinusoidal supply for electrical/electronic equipment. Resonant inverters are soft switched inverters with very low switching loss and they are usually designed and operated at very high frequency. In this work, the design and operation of a resonant inverter at 50Hz is a challenging work in terms of component size. Also, the LC of the inverter has to match with the front end DC-DC converter[4].

II. Power Supplies of Remote Area electrical/electronic Equipment

When planning remote area electrical/electronic power supply , it has a number of possible power solution options. Selection of the optimum solution will depend on the local circumstances and could include:

- Generators
- Wind power
- Solar Power
- Bio Fuels
- Fuel Cells
- Microhydro Generator

- Pico Hydro
- Power Storage

III. DC-DC Power Converters

In many industrial applications, it is required to convert a fixed-voltage dc source into another level or a variable-voltage dc source. A DC converter can be considered as dc equivalent to an ac transformer with a continuously variable turn ratio. Like transformer, it can be used to step down or step up a DC voltage source Basic converters are of four types:

- Buck
- Boost
- Buck-Boost
- Ćuk

IV. Ćuk Converter

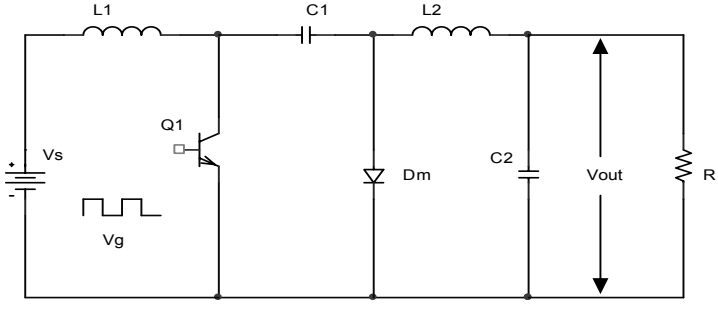


Fig 1 : Ćuk converter

With a clearly defined goal of achieving non-pulsating currents, the desired converter configuration gradually emerges**:** an inductor is needed in series with both input source and output load for either switch position. Then, energy transfer and level conversion is achieved by use of a single capacitance and a single switch, or its bipolar transistor, diode implementation, known as Ćuk converter. The Ćuk converter is formed by combining the Buck and the Boost converter. The Ćuk converter is a type of DC-DC converter that has an output voltage magnitude that is either greater than or less than the input voltage magnitude, with an opposite polarity[5]. It uses a capacitor as its main energy-storage component, unlike most other types of converter which use an inductor. It is named after Slobodan Ćuk of the California Institute of Technology, who first presented the design.

The main component that controls the flow of energy from input to output is the capacitor in between. Ćuk converter feature capacitive energy transfer to attain high efficiency.

V. Resonant Inverter

The inverter is a basic component of any independent power system that requires AC power. Inverters convert DC power stored in batteries or DC power directly obtained from alternative sources into AC power to run conventional appliances.

A resonant inverter is a high frequency inverter used in many applications. Resonant switching topologies are the next generation of power conversion circuits, when compared to traditional pulse width modulation (PWM) topologies. Resonant-based supplies are more efficient than their PWM counterparts. This is due to the zero current and/or zero voltage transistor switching (ZVS or ZCS), so that the switching loss is zero and the switches are not subjected to high-voltage stress[6]. that is inherent in a resonant supply design. This feature also provides additional benefits of eliminating undesirable electromagnetic radiation normally associated with switching supplies, reduction in thermal requirement, snubberless operation. The voltage and current are forced to pass through zero crossing by creating an LC resonant circuit.

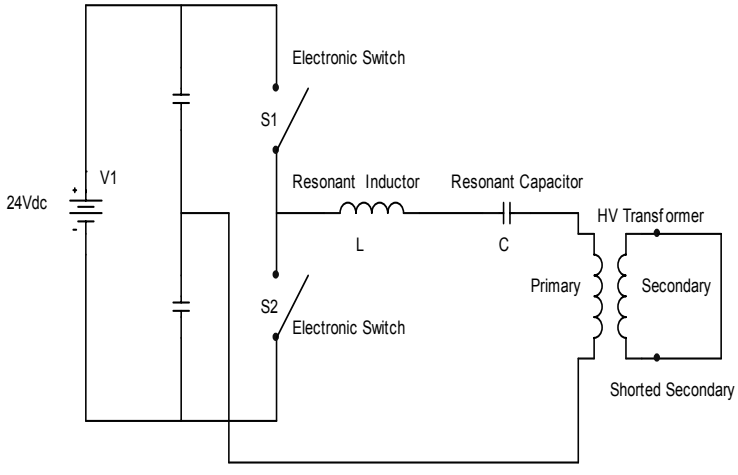


Fig 2: Resonant Inverter

Where, $I = V\sqrt{\frac{C}{L}}$; $f = \frac{1}{2\pi\sqrt{LC}}$

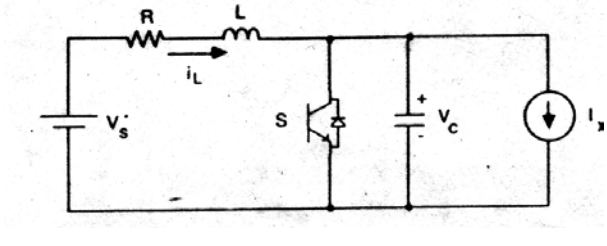


Fig 3: Equivalent circuit of resonant inverter

VI. Analysis of Ćuk converter : Need for double stage Ćuk operation

In following section, by analyzing the single stage and double stage Ćuk converter, the need for double stage is explained:

a. Efficiency for single stage Ćuk converter

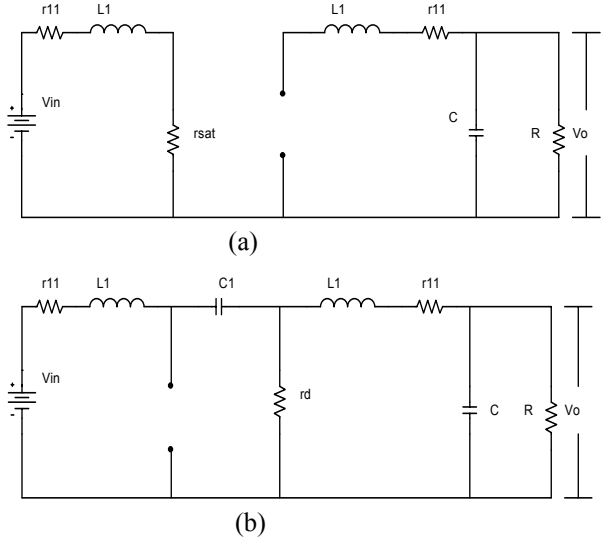


Fig 4 : Single stage Ćuk converter operation
(a) Switch off (b) Switch on

From the Fig 4,
$r_1 = r_{l1} + r_{sat}$
and $r_2 = r_{l2} + r_d$
Efficiency,

$$\eta = \frac{I_0^2 R}{I_0^2 R + I_{in}^2 r_1 + I_0^2 r_2}$$

$$\text{or, } \eta = \frac{1}{1 + \frac{I_{in}^2}{I_0^2} \cdot \frac{r_1}{R} + \frac{r_2}{R}}$$

$$= \frac{1}{1 + \left(\frac{D}{1-D}\right)^2 \cdot \alpha_1 + \alpha_2}$$

Where, $\alpha_1 = r_1/R$ and $\alpha_2 = r_2/R$

Now, r_{l1} and r_{l2} are chosen as 0.1Ω

$\alpha_1 = r_1/R$
$= (r_{l1} + r_{sat})/R$; $r_{sat} \approx 0.9$ and $R = 300\Omega$
$\alpha_1 = (0.1 + 0.9)/ 300$
$= 0.00333$

$\alpha_2 = r_2/R$
$= (r_{l2} + r_d)/R$; $r_d \approx 0.9$ and
$R = 300\Omega$
$\alpha_2 = (0.1 + 0.9)/ 300$
$= 0.00333$

Efficiency,

$$\eta = \frac{1}{1 + 0.00333 \left(\frac{D}{1-D}\right)^2 + 0.00333}$$

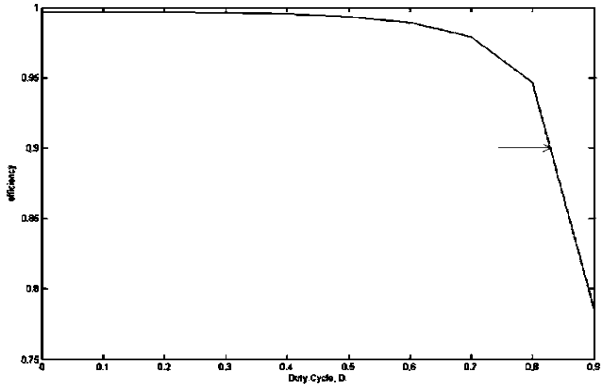


Fig 5: Efficiency, η Vs. Duty Cycle, D for single stage Ćuk converter

From the above curve (Fig 5) it is clear that for a single stage converter; the maximum efficiency practically attainable is 98.5%; where the value of D = 0.6. The input voltage is 24V DC.

The output will be $V_0 = 24 \times \frac{D}{1 - D}$
$= 24 \times 0.6/0.4$
$= 36$ V

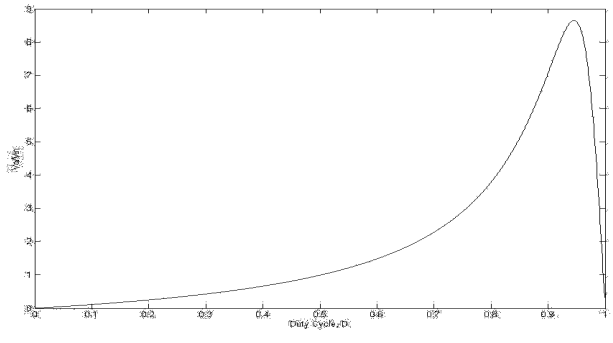


Fig 6: Voltage transfer function Vs. duty cycle V_0/Vin Vs. Duty Cycle, D for single stage Ćuk

VII. Efficiency for double stage Ćuk converter

Efficiency, $\eta = \eta_1 \times \eta_2$

Again, $\eta = \dfrac{V_0 I_0}{V_{in} I_{in}}$

$$\frac{V_0}{V_{in}} = \eta \frac{I_{in}}{I_0}$$

$$= \eta \frac{D}{1 - D}$$

$$\frac{V_0}{V_{in}} = \frac{1}{1 + 0.00333 \left(\frac{D}{1-D}\right)^2 + 0.00333} \cdot \frac{D}{1-D}$$

Voltage transfer function V_0/V_{in} with variation of duty cycle D for single stage converter is shown in Fig 6. Figure shows the V_0/V_{in} increases due to parasitic resistances and device voltage drops in practical circuit. Now, efficiency for double stage is given by:

$$\eta = \frac{1}{1+0.00333\left(\frac{D_1}{1-D_1}\right)^2 + 0.00333} \times \frac{1}{1+0.00333\left(\frac{D_2}{1-D_2}\right)^2 + 0.00333}$$

If $D_1 = D_2$ then,

$$\eta = \left(\frac{1}{1 + 0.00333 \cdot \left(\frac{D}{1-D}\right)^2 + 0.00333} \right)^2$$

Variation of efficiency of double stage converter is shown in Fig 7, which indicates that maximum efficiency of is 99% at D = 0.55. At 0.8 duty cycle, efficiency of single stage converter fall below 90%, whereas at D = 0.8 efficiency of double stage converter still remains above 95%.

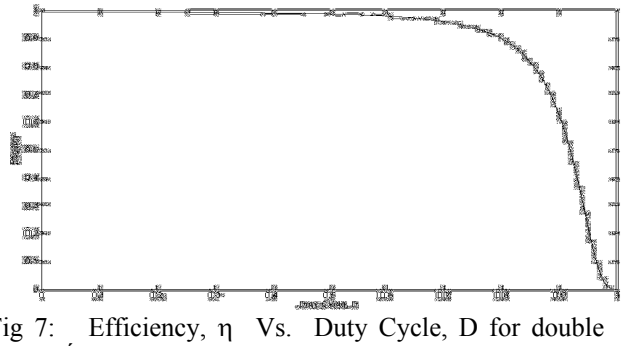


Fig 7: Efficiency, η Vs. Duty Cycle, D for double stages Ćuk converter

For double stage, taking the efficiency η = 0.9, from the efficiency curve, D = 0.8

Again,

$$\frac{I_0}{I_{in}} = \frac{I_{02}}{I_{in2}} \cdot \frac{I_{01}}{I_{in1}}$$

$I_{in2} = I_{01}$

$$\frac{I_0}{I_{in}} = \frac{I_{02}}{I_{in1}}$$

$$\frac{I_0}{I_{in}} = \frac{1 - D_1}{D_1} \cdot \frac{1 - D_2}{D_2}$$

If $D_1 = D_2$, then

$$\frac{I_0}{I_{in}} = \left(\frac{1 - D}{D}\right)^2$$

Again, $\eta = \dfrac{V_0 I_0}{V_{in} I_{in}}$

$$\frac{V_0}{V_{in}} = \eta \frac{I_{in}}{I_0}$$

$$= \eta \left(\frac{D}{1 - D}\right)^2$$

$$\frac{V_0}{V_{in}} = \left(\frac{D}{1-D}\right)^2 \left(\frac{1}{1+0.00333 \cdot \left(\frac{D}{1-D}\right)^2 + 0.00333} \right)^2$$

Taking D = 0.8 (from the curve),

$$\frac{V_0}{V_{in}} = \left(\frac{0.8}{1-0.8}\right)^2 \left(\frac{1}{1+0.00333 \cdot \left(\frac{0.8}{1-0.8}\right)^2 + 0.00333} \right)^2$$

$$\frac{V_0}{V_{in}} = 14$$

Or, $V_0 = 24 \times 14 = 336$ V

So, the output voltage can be easily obtainable as 230V or more keeping the efficiency of the converter more than 90%, which is the main reason why a double stage Ćuk converter is used in this thesis work.

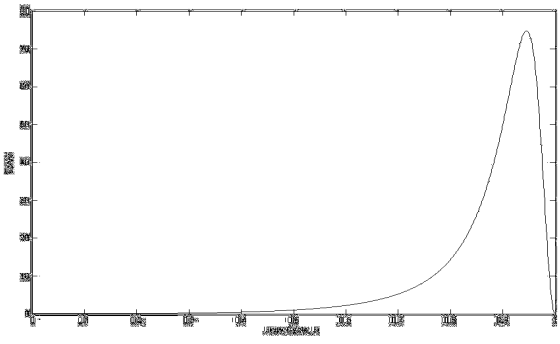


Fig 8: Voltage transfer function, V_0/Vin Vs. Duty Cycle, D for double stage Ćuk

VIII. Ćuk Converter fed Resonant Inverter

In the Ćuk Converter fed Resonant Inverter, the input selected for the Ćuk Converter is 24V DC as obtainable from the solar or other alternative source. Practical switches are replaced by IGBTs. Switching voltage for the IGBT is 20V. The output is taken from a 300Ω resistance.

The Resonant Inverter is fed from the Ćuk Converter. If the Ćuk Converter is not connected with the Resonant Inverter its output is about 220V DC; but when the Resonant Inverter is connected as a load to the Ćuk Converter its output drops to about 140V DC, and the output of the Resonant Inverter is about 100V AC. A modified Ćuk Regulator circuit is used to compensate for the drop. The components of the Resonant Inverter is so chosen to obtain the output frequency as 50Hz. The output impedance of the Resonant Inverter is taken as 300Ω.

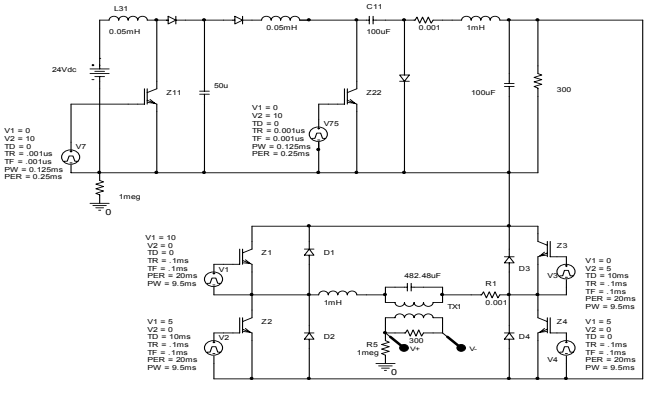


Fig 9: Ćuk Regulator fed Resonant Inverter

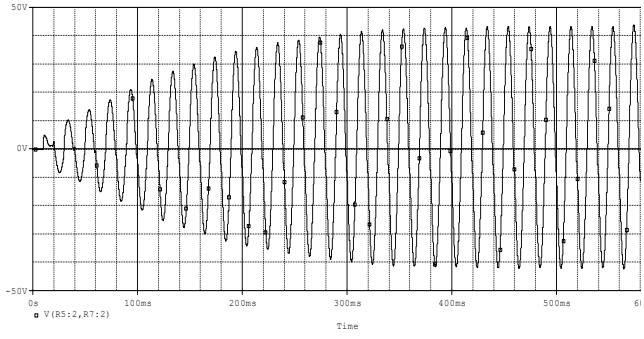


Fig 10: Wave shape of Cûk Regulator fed Resonant Inverter

The desired frequency 50Hz is attainable in the single stage Ćuk Regulator (Fig 10) but it will not provide 300V as required for the output of 300Ω load. So, double stage modified Ćuk Regulator is required to be used as in the next section.

IX. Modified Ćuk Regulator fed Resonant Inverter

The modified Ćuk Regulator fed Resonant Inverter is shown in Fig 11. In this circuit, the switches in the Cûk Regulator are replaced by Darlington pair and that of Resonant Inverter is replaced by IGBTs for practical considerations. Like the modified Ćuk Regulator fed Resonant Inverter, two lift circuits are used for compensation. The input is taken as 24V DC. The load in the output of the Resonant Inverter is 300Ω. The output wave shape in the Resonant inverter as taken in the 300Ω load is shown in Fig 12. From Fig 12, it is evident that the output voltage is 240V AC and frequency is 50Hz.

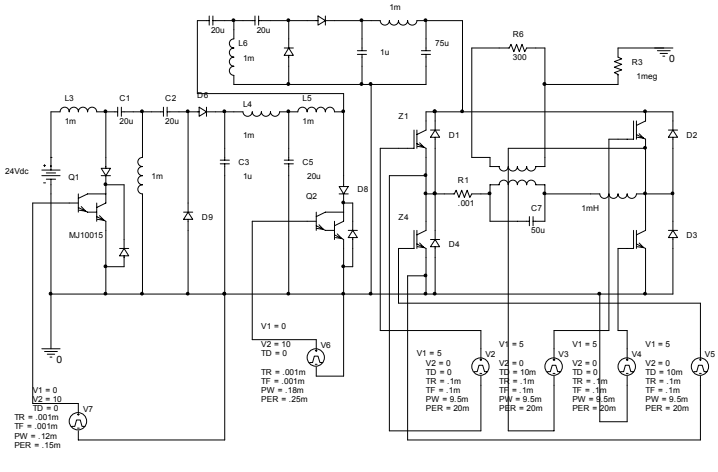


Fig 11 : Modified Ćuk Regulator fed Resonant Inverter

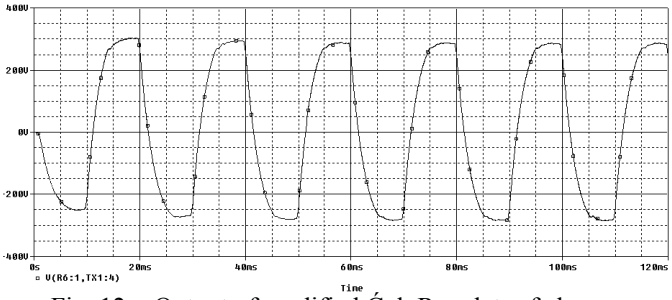


Fig 12: Output of modified Ćuk Regulator fed Resonant Inverter

The deviation from the theoretical one is for the need of lift circuit for resonant inverter and other practical constraints.

X. Discussion

The objective has been to design and analyze a power supply to the load is a 240V AC at 50Hz frequency from renewable/alternative energy sources such as solar system, fuel cell and microwind turbines etc. The maximum frequency of the zero current switching based inverter is as high as the resonant frequency, because a switching device can be turned on or off once in a resonant cycle. In the simulation works, IGBT's are used as switching devices, so that the resonant frequency f_R is designed to be 50Hz is given by,

$$f_R = \frac{1}{2\pi\sqrt{L_R C_R}} = 50Hz$$

The parameters are so chosen to obtain desired AC voltage to run conventional commercial electrical/electronic equipments designed for 230V, 50Hz or 60Hz. Step up DC-DC conversion has been achieved by duty cycle control of Ćuk converter topology. In this work, Ćuk converter is used for step up operation. The output of Ćuk converter is fed to a resonant inverter circuit to obtain direct sinusoidal supply for commercial equipments. Resonant inverters are soft switched inverters and they are usually designed and operated at very high frequency. But in this thesis work the inverter is designed for 50Hz output.

A 2-stage Ćuk Regulator has been used. The negative output of the Ćuk Regulator is made positive with suitable circuit configuration. If a transformer would be used in the first stage of the circuit to step-up the voltage, the circuit would have to withstand high current, which would require heavy transformer. Calculation shows that the designed circuit has reduced the wire the wire size to at least $\frac{1}{25}$ times than it would be if a transformer was used. It is shown that, the efficiency of a double stage converter is above 90%, which is more than the efficiency achievable by single stage Ćuk converter[7].

XI. Recommendation

The proposed design "Ćuk Converter fed Resonant Inverter" to convert the alternative supply (DC) into commercial supply (240V, 50/60Hz) being a successful and efficient one, may be implemented practically in laboratory and be made commercially available in future.

References

[1] Gregory J., "Financing Renewable Energy Projects," IT Publications, UK, 1997.

[2] Okoro O. I., and Madueme T.C., "Solar energy: a necessary investment in a developing economy", International Journal of Sustaunable Energy" Volume 25, Number 01/March 2006.

[3] Middlebrook R. D., and Slobodan Ćuk, "A New Optimum Topology Switching DC -DC converter," IEEE Power Electronics Conference Record, Palo Alto, CA, USA, pp 170-179, 1977.

[4] Pipattanasomporn M., and Rahman S., "Information and Communication Technology Infrastructure and Its Distributed Generation Solutions in Remote Areas", Department of Electrical and Computer Engineering, Alexandria Research Institute, Virginia Tech, Alexandria, VA 22314, U.S.A

[5] Rashid M H.. and Rashid H. M. "SPICE FOR POWER ELECTRONICS AND ELECTRIC POWER", 2nd Edition,Taylor & Francis Group, New work.

[6] Rashid M H., Power electronics, "Circuits, Devices and Applications", 3rd Edition, Prentice Hall of India Pvt limited, New Delhi.

[7] Bose B. K., " Introduction to Power Electronics," University of Tennessee, Published in 1992.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Self-Tuned NFC Based Speed Ripple Minimization of a Faulty Induction Motor

M. Nasir Uddin[1], *Senior Member IEEE,* and Z. R. Huang[1]

Department of Electrical Engineering
Lakehead University
Thunder Bay, Ontario P7B 5E1, Canada
E-mail: muddin@lakeheadu.ca

Abstract - **This paper presents a self-tuned neuro-fuzzy controller (NFC) based speed ripple minimization of a vector controlled faulty induction motor (FIM) with broken rotor bars. First, the performance of the FIM is investigated in terms of speed ripple under the open-loop condition. Then, a new mechanical model of induction motor is developed incorporating the speed ripple. Based on this model a new NFC is proposed to tolerant the effect of the fault under an indirect field oriented control scheme. The proposed NFC compensates the faulty condition by minimizing the supply frequency related speed ripples instead of directly working on the low frequency signature speed ripples which a FIM exhibits. Based on the knowledge of motor control and intelligent algorithms an unsupervised self-tuning method is developed to adjust weights of the proposed NFC. The convergence of the weights is also discussed. The complete drive is experimentally implemented using a digital signal processor board DS-1104 for a laboratory 250W faulty IM. The effectiveness of the proposed NFC is tested both in simulation and experiment.**

I. Introduction

The induction motor (IM) has been considered to be workhorses for industries of manufacturing, oil, chemical, etc. due to its simple and robust construction. The vector control is generally adopted for precise speed tracking. However, the IM tends to produce speed ripple and hence the fault tolerant control is highly desired for IM drive [1]. Among many causes of speed ripples in an IM drive the unevenness of air gap and finite number of stator slots and rotor bars are causes of fundamental frequency speed ripple (i.e., the frequency of the stator supply, f_e) [2-3]. Sixth harmonic speed ripple is the effect of dead time of an inverter-fed motor drive [4-5]. Under the field oriented control the DC bias current measurement error causes f_e speed ripple and the scaling error causes the $2f_e$ speed ripple [6]. Stator asymmetry (e.g. stator winding fault or unbalance power supply) also causes $2f_e$ speed ripple [7]. The faulty motor, particularly the broken rotor bar, produces specific speed ripple, which is called signature ripple of low frequency modulation in speed response [7-8].

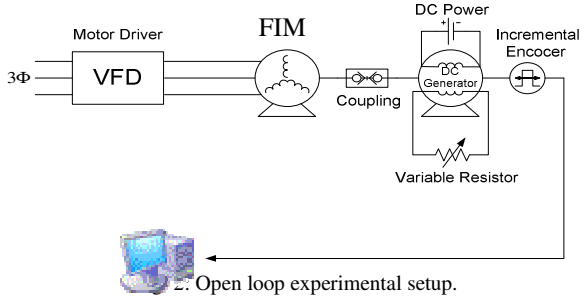
The speed or torque ripple compensation methods can be found in [8-11]. Authors in [8] only compensated the low frequency ripples $2sf_e$ of a FIM by using a fault-tolerant controller, which is not the major cause of speed ripple. In order to reduce the speed ripple in [9] authors developed a torque compensation method, which requires torque sensor that may not be acceptable in many applications. In [6] the authors utilized a compensation method by correcting current measurement error in order to reduce the f_e and $2f_e$ speed ripples. But the current measurement error is not the only source of speed ripples.

In this paper first a mathematical model is developed for the faulty induction motor (FIM) incorporating speed ripple, which assists to analyze the control scheme. Based on this model a NFC is developed to minimize f_e and $2f_e$ speed ripples of a FIM. The proposed NFC produces a constant torque plus a reverse ripple torque to fulfil two control targets: speed set point and lower speed ripple. The inputs of the proposed NFC are only rotor position and velocity, which is also calculated from rotor position and hence no system parameters or any extra hardware are needed. The complete drive is successfully implemented in laboratory using a digital signal processor (DSP) board DS-1104 for a prototype 250W faulty IM. Simulation and experimental results validate the capabilities of the proposed NFC to reduce the speed ripple and hence the fault tolerant ability of the controller for FIM.

II. Modelling of IM Incorporating Speed Ripple

An open-loop experiment was conducted on a FIM with broken rotor bar (as shown in Fig. 1) at different load and speed conditions as per Fig. 2. Figure 3(a) shows a particular speed response at 22Hz and its Fast Fourier Transform (FFT) is shown in Fig. 3(b). It is shown from Fig. 3(a) that a low frequency of $2sf_e$ modulation exists in rotor speed and Fig. 3(b) shows that the f_e and $2f_e$ speed ripples are dominant parts among all of the components of speed ripples. The experimental results are summarized in Table 1. At the initial faulty condition, the magnitude of signature ripples ($2sf_e$) related to different fault conditions is much lower than that of f_e and $2f_e$ speed ripples. Further

Fig. 1: The Faulty rotor of the IM.



2. Open loop experimental setup.

Table 1: FFT analysis of faulty IM with broken rotor bars

Load Level	Speed ref. (rpm)	Speed ripple		
		f_e(rad/s)	$2f_e$(rad/s)	$2sf_e$
Low	10	0.266	0.252	0.014
↓	15	0.384	0.319	0.016
	20	0.358	0.389	0.022
	25	0.519	0.387	0.025
Medium	30	0.733	0.381	0.033
Medium	10	0.214	0.211	0.021
↓	15	0.414	0.369	0.035
	20	0.512	0.334	0.034
	25	0.626	0.396	0.043
High	30	0.740	0.434	0.070

analysis has been done to reveal the relationship between speed ripples and rotor position. A fault-tolerant controller is supposed to be able to minimize signature frequency ripples which belong to particular fault condition as well as the f_e speed ripple in order to achieve good speed tracking and compensate the effect of the fault condition. Figure 4(a) illustrates the relationship between the f_e speed ripple and the rotor position angle θ, while the Fig. 4(b) for $2f_e$ speed ripple and 2θ. It gives us the theory support that the nth order speed ripple could be modeled by,

$$\omega_{nth\_ripple} = K_a \cos(n\theta) + K_b \sin(n\theta) \qquad (1)$$

The experimental observation suggests that: 1) The signature low frequency speed ripple $2sf_e$ can be neglected because of too small magnitude at the initial stage of fault; 2) The source of speed ripples can be modeled as a linear summation of sinusoidal functions whose frequencies are multiple of supply frequency f_e.

As the torque ripple causes the speed ripple, the modified IM mechanical model incorporating the torque ripple can be written as follows:

$$J_m \frac{d\omega_m}{dt} + B_m \omega_m = T_e^* - T_L + (T_e^* - T_L) \sum_{n=1}^{\infty} [K_{an} \cos(n\theta) + K_{bn} \sin(n\theta)] \qquad (2)$$

where ω_m is rotor speed, T_L is a constant load torque, T_e is the developed torque, J_m is moment of inertia, B_m is the coefficient of viscous friction, K_{an} & K_{bn} are scale factors representing magnitudes and phases of frequency components for torque ripples. In order to separate the speed ripple from the reference set speed, developed torque can be split into two parts: T_{ss} and T_{ripple}. T_{ss} is a constant torque regarding

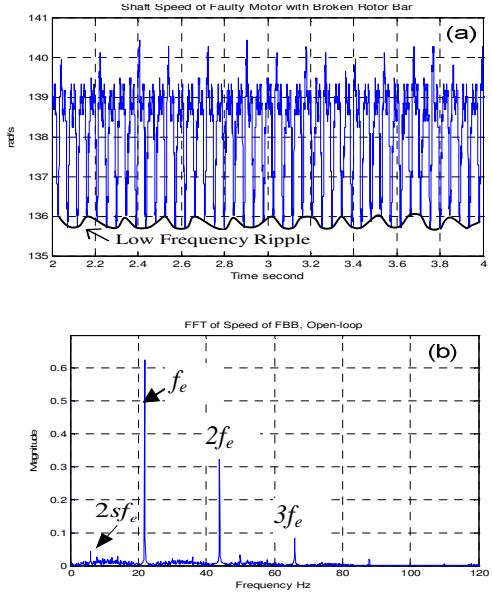


Fig. 3: (a) Shaft Speed of a FIM with broken rotor bars, (b) FFT Analysis of speed signal.

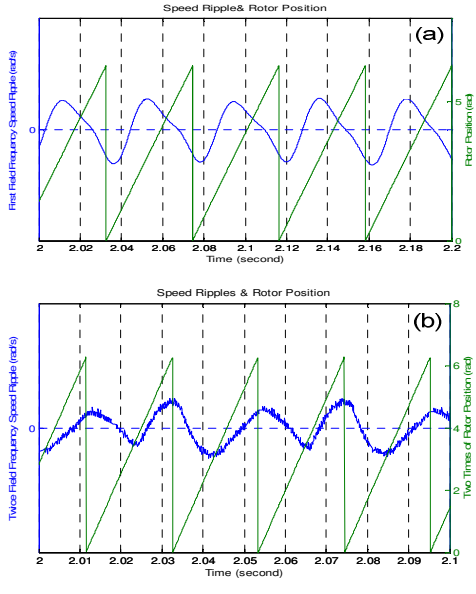


Fig. 4: Relation between speed ripple and position: (a) f_e speed ripple & θ, (b) $2f_e$ speed ripple & 2θ.

Reaching the reference set speed, and T_{ripple} is the reverse ripple torque regarding eliminating the speed ripples. Thus, (3) can be re-written as:

$$J_m \frac{d\omega_m}{dt} + B_m \omega_m \tag{3}$$
$$= \{T_{ss} - T_L\} + \left\{ (T_{ss} - T_L + T_{ripple}) \sum_{n=1}^{\infty} [K_{an} \cos(n\theta) + K_{bn} \sin(n\theta)] + T_{ripple} \right\}$$

To eliminate the speed ripple in steady-state,

$$(T_{ss} - T_L + T_{ripple}) \sum_{n=1}^{\infty} [K_{an} \cos(n\theta) + K_{bn} \sin(n\theta)] + T_{ripple} = 0$$
$$T_{ss} - T_L = B_m \omega_m \tag{4}$$

Then T_{ripple} and T_{ss} can be obtained by:

$$T_{ripple} = \frac{-(T_{ss} - T_L) \sum_{n=1}^{\infty} [K_{an} \cos(n\theta) + K_{bn} \sin(n\theta)]}{1 + \sum_{n=1}^{\infty} [K_{an} \cos(n\theta) + K_{bn} \sin(n\theta)]} \tag{5}$$
$$T_{ss} = B_m \omega_m + T_L$$

Equation (5) shows great complexity and nonlinearity. Traditional techniques are difficult to fit such kind of equation. However, neural network or neuro-fuzzy control (NFC) techniques could be a good candidate to approximate nonlinear functions due its inherent capability to learn [13-15].

II. Neuro-Fuzzy Controller and Tuning Algorithm

The proposed NFC incorporates fuzzy logic and a learning algorithm with a five-layer artificial neural network (ANN) structure as depicted in Fig. 5. Based on the modeling of speed ripple in (5) speed error, $\omega_m^* - \omega_m$, and rotor position angle, θ are chosen to be the two inputs of the proposed NFC based speed controller. The learning algorithm modifies the NFC to closely track the speed reference, and at the same time minimize the speed ripples.

A. Self Tuning Algorithm

An unsupervised on-line self-tuning method is introduced in this paper. The objective of the proposed NFC is to generate the correct T_e^* in the indirect field orientation, and then produce correct torque to counteract torque ripples. In this paper, system output error is employed to guide the control action in the right direction. The NFC parameters are directly tuned to reduce the output error. The Kaczmarz's projection algorithm is used to update the weights as the convergence of weight is very slow in back propagation algorithm [10,12]. The update rules are given as follows:

$$w_j(n) = w_j(n-1) + \gamma (u_d^{(n)} - u^{(n)}) \frac{O_j^{III}(n-1)}{\sum (O_j^{III} \wedge 2)} \tag{6}$$

where $u_d^{(n)} - u^{(n)}$ is controller output error, $\gamma > 0$ is a learning factor. In terms of the Jacobean matrix *J(n)* of the system, the system error and controller output error can be expressed as

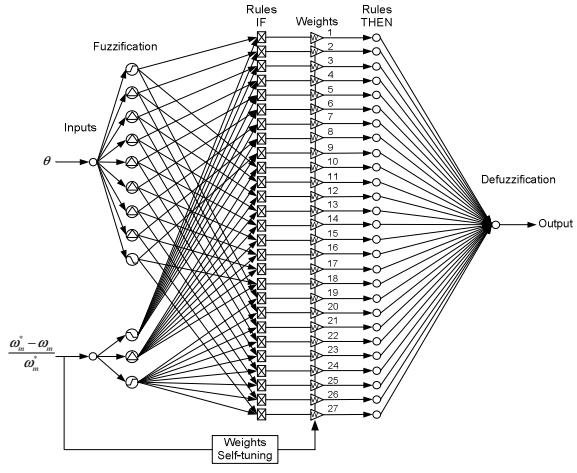
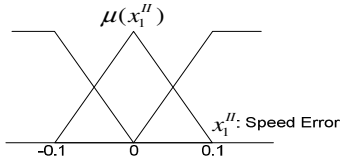
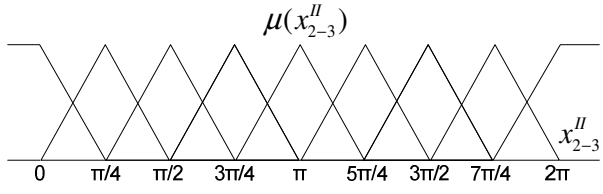


Fig 5: Structure of the proposed NFC.



(a)



(b)

Fig. 6: (a) Membership functions for input-1, (b) membership functions for input-2 and input-3.

$$u_d^{(n)} - u^{(n)} = \frac{1}{J(u)} (\omega_m^* - \omega_m) \tag{7}$$

In (6) the Jacobean matrix *J(n)* is not easily found directly. In FOC the IM system can be viewed as a single input single output system, then the *J(n)* can be estimated as a constant value $K_J > 0$. Then (5) can be rewritten as:

$$w_j(n) = w_j(n-1) + \eta (\omega_m^* - \omega_m) \frac{O_j^{III}(n-1)}{\sum (O_j^{III} \wedge 2)} \tag{8}$$

Where, $\eta = \frac{\gamma}{K_J}$. However this weight-updating method may cause problems. For the weights $w_{1\sim 9}$, the updating

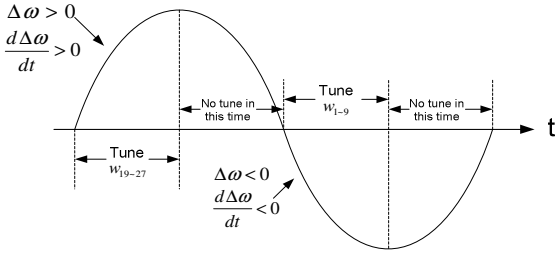


Fig. 7: Weights $w_{1\sim 9}$ and $w_{19\sim 27}$ tuning time.

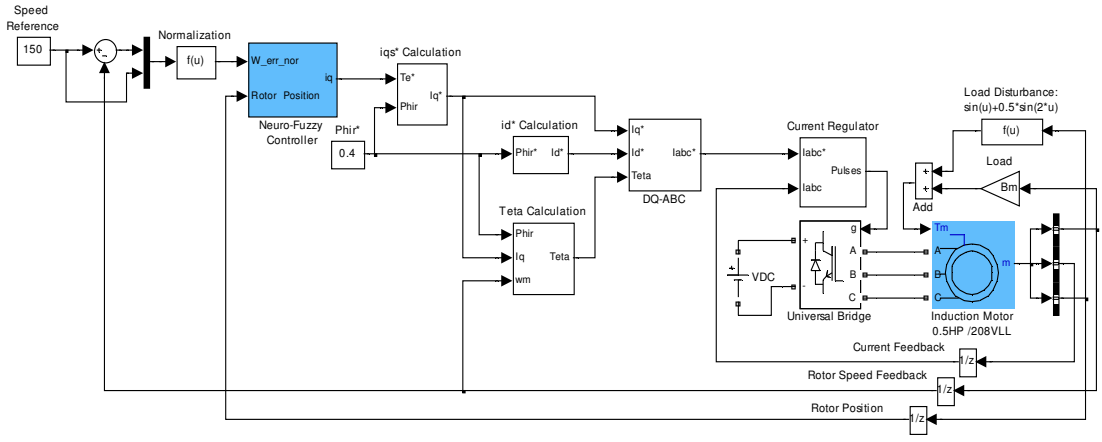


Fig 8: Block diagram of the proposed NFC based FIM drive.

is always happening at the time when $\omega_m^* - \omega_m \le 0$. And because $\eta > 0$ and $\dfrac{O_j^{III}(n-1)}{\sum (O_j^{III} \wedge 2)} \ge 0$, the weights $w_{1\sim 9}$ are consistently decreasing. On the contrary, the weights $w_{19\sim 27}$ are consistently increasing. In this paper we developed an algorithm to fix the problem, which consists of the following three steps:

(i) Assign appropriate initial value to the weights $w_{1\sim 9}$ and weights $w_{19\sim 27}$. Since the weights $w_{1\sim 9}$ relate to the time when $\omega_m \ge \omega_m^*$, the NFC is desired to produce a smaller T_e^* than the T_L, while weights $w_{19\sim 27}$ relate to the time when $\omega_m \le \omega_m^*$, the NFC is desired to produce a larger T_e^* than the T_L, the initial values are given: $w_{1\sim 9} = -PB$, $w_{19\sim 27} = +PB$, PB is a positive constant.

(ii) Tune the weights $w_{1\sim 9}$ and weights $w_{19\sim 27}$ at different specific time. The IM mechanical equation is as following:

$$J_m \frac{d\omega_m}{dt} + B_m \omega_m = T_e^* - T_L \qquad (9)$$

When the system reaches the steady state, the speed is varying around set point. At this time one can say,

$$\omega_m = \omega_m^* - \Delta\omega_m \qquad (10)$$

Meanwhile we split NFC output T_e^* into two parts: T_e^{ss} and T_e^{ripple}, and load torque into: T_L^{ss} and T_L^{ripple}. T_e^{ss} and T_L^{ss} are constant values, and T_e^{ripple} and T_L^{ripple} are varying values. The equation (10) can be rewritten as:

$$-J_m \frac{d\Delta\omega_m}{dt} - B_m \Delta\omega_m + B_m \omega_m^* = -\Delta T + T_e^{ss} - T_L^{ss} \qquad (11)$$

where $\Delta T = T_L^{ripple} - T_e^{ripple}$. Obviously, $B_m\omega_m^* = T_e^{ss} - T_L^{ss}$. Thus (11) is reduced to,

$$J_m \frac{d\Delta\omega_m}{dt} + B_m \Delta\omega_m = \Delta T \qquad (12)$$

In (17) $\Delta\omega_m$ is system error, ΔT is controller output error. From (12) it can be noted that:

$\dfrac{d\Delta\omega_m}{dt} \le 0$ & $\Delta\omega_m \le 0$, $\Delta T \le 0$;

$\dfrac{d\Delta\omega_m}{dt} \ge 0$ & $\Delta\omega_m \ge 0$, $\Delta T \ge 0$.

This gives the specific time when the weights $w_{1\sim 9}$ and weights $w_{19\sim 27}$ will be tuned as illustrating in Fig. 7. Since speed fluctuation happens at the same rotor position at different time, $w_{1\sim 9}$ / $w_{19\sim 27}$ are decreasing / increasing at different rate. Stop tuning when the system error is smaller than a threshold.

III. Simulation and Experimental Setup

Based on the NFC control algorithm explained in section III the block diagram of the proposed IM drive is shown in Fig. 8. The complete drive has been simulated using Matlab/Simulink according to this figure [16]. For simulation the f_e and $2f_e$ speed ripples are added in as load disturbances,

$$Load\ disturbance = \sin(\theta) + 0.5\sin(2\theta) \qquad (13)$$

The performance of the proposed NFC is compared to a fine tuned PI controller based drive at different operating speeds. To make a fair comparison, PI controller is readjusted whenever reference speed is changed and P-gain is set as large as possible. For simulation the parameters of PI and NFC are listed as follows:

PI: $K_p = 0.9$, $K_i = 0.6$ $\omega_m^* = 100 rad/s$

$K_p = 0.5$, $K_i = 0.3$ $\omega_m^* = 150 rad/s$

$K_p = 0.4$, $K_i = 0.3$ $\omega_m^* = 180 rad/s$

NFC: *Learning rate = 0.1,* Initial values $w_{1\sim 9} = -2.5$, $w_{19\sim 27} = +2.5$

The parameters of the FIM are shown in Appendix. The motor was damaged by introducing a rotor fault as shown in Fig. 1.

The proposed self tuned NFC based vector control of FIM drive system has been implemented in real-time using the DSP board DS1104 [17] for a laboratory 250 W faulty IM with broken rotor bars11. The block diagram experimental system is shown in Fig.9, respectively. The actual motor currents are measured by the Hall-effect sensors and fed back to the DSP board through the A/D channels. The rotor position is sensed by an optical incremental encoder of 1000-line resolution and is fed back to the DSP board through the encoder interface. The test FIM is coupled to a dc machine. The dc machine is

operated as a generator in order to adjust the mechanical load to the FIM. The indirect field orientation control scheme incorporating the proposed self-tuned NFC is implemented through developing a real-time Simulink model. Then the model is compiled and downloaded to the DSP board utilizing ControlDesk software and real-time workshop (RTW). For comparison purpose, a PI-controller-based system is also developed and experimentally implemented. After trial and error, the PI controller gains K_p and K_i are readjusted in real-time as, 0.4 and 0.3, respectively, so that the magnitudes of speed ripples can be comparable to the proposed NFC. The sampling frequency is set to be 10 kHz.

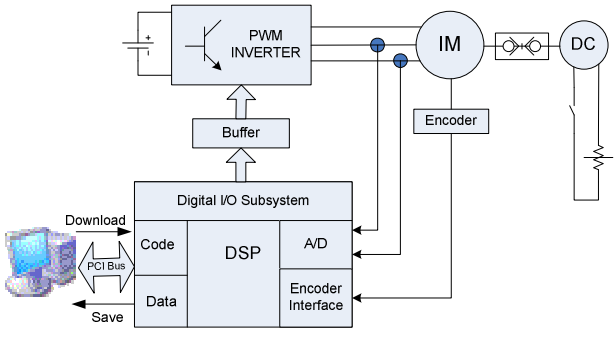


Fig 9: (a) Block diagram of the experimental system.

in order to achieve good current tracking. Due to the limitation of the computability of the processor, in real-time the weights $w_{1\sim 9}$ and $w_{19\sim 27}$ are kept as fixed values as: $w_{1\sim 9} = -3$, $w_{19\sim 27} = 3$. These are found based on simulation results.

IV. Simulation and Experimental Results

A. Simulation results

Figure 10 shows the comparative simulated speed response of the FIM drive with PI and proposed NFC at a reference speed of 150rad/s. It is seen in this figure that the speed ripple has been significantly minimized by the proposed NFC as compared to the conventional PI controller. Thus, the proposed self tuned NFC is capable to minimize the fluctuation of speed and illustrate fault tolerant ability. The effectiveness of tuning the weights can be seen in Fig. 11. It is seen that the speed ripple for the self tuned NFC is further reduced as compared to the NFC without tuning the weights.

B. Experimental results

The simulated results are verified by the experimental results. Figure 12 shows the experimental speed responses for PI and proposed self tuned NFC based FIM drive. It shows that the proposed NFC has eliminated the speed ripple by up to 50%. The FFT analysis of experimental speed response shown in Fig. 13 further verify that the proposed NFC controller can reduce the f_e and $2f_e$ speed ripples of the FIM as compared to the conventional PI controller. Thus, the proposed NFC compensates the faulty condition by minimizing the supply frequency related speed ripples instead of directly working on the low frequency signature speed ripples which a FIM exhibits. In order to reduce the computational burden for the DSP the weights $w_{1\sim 9}$ and $w_{19\sim 27}$ are kept constant based on the simulation study.

Therefore, it is evident from both simulation and experimental results that the proposed self tuned NFC is capable to minimize the speed ripple in steady-state and illustrate fault tolerant ability of a faulty IM with broken rotor bars.

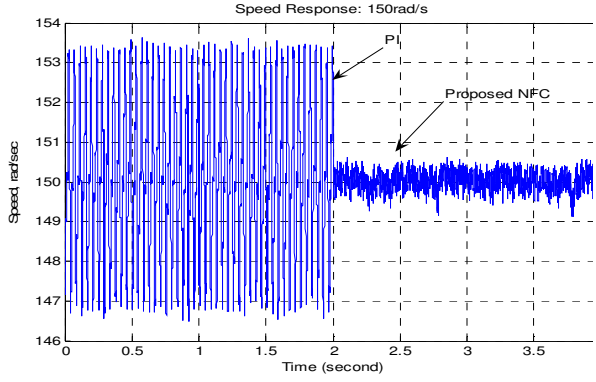


Fig 10: Simulated comparative speed response for FIM drive with PI and proposed NFC at reference speed of 150 rad/s.

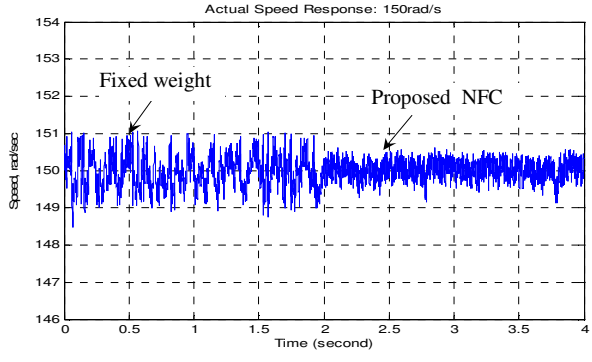


Fig 11: Simulated comparative speed response of the NFC based FIM drive with and without fixed weights.

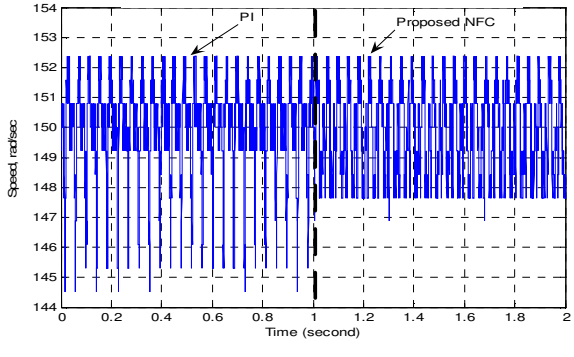


Fig 12: Experimental speed responses: PI (left side), NFC (right side).

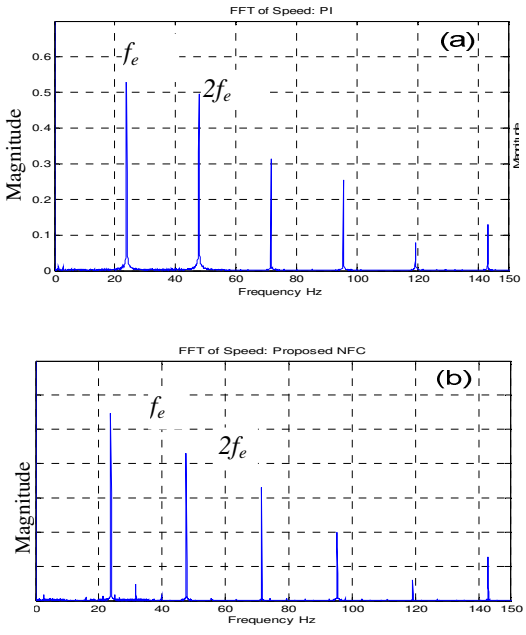


Fig 13: FFT analysis of experimental speed response at 22 Hz; (a) PI and (b) NFC.

V. Conclusion

The modeling and minimization technique for speed ripple of a vector controlled faulty induction motor with broken rotor bars have been presented in this paper. A new mechanical model of IM incorporating the speed ripple has been developed based on an open loop test. Based on this model a novel NFC is proposed to tolerant the effect of the fault under an indirect field oriented control scheme. The proposed NFC compensates the faulty condition by minimizing the supply frequency (f_e) related speed ripples instead of directly working on the low frequency ($2sf_e$) signature speed ripples which a FIM exhibits. Based on the knowledge of motor control and intelligent algorithms an unsupervised self-tuning method is developed to adjust the weights of the proposed NFC. The complete drive has been successfully implemented in real-time using a DSP board DS-1104 for a laboratory 250W faulty IM. The effectiveness of the proposed NFC is tested both in simulation and experiment. There is a close agreement between simulation and experimental results. The proposed NFC exhibits better performance in terms of speed ripple minimization as compared to a conventional PI controller. The proposed self tuned NFC is found capable to minimize the speed ripple in steady-state and illustrate the fault tolerant ability of a FIM with broken rotor bars.

APPENDIX: Induction Motor Parameters for Simulation and Experiment

Power	250 W
Stator resistance	6.5 Ω
Rotor resistance	3.4 Ω
Number of pole pairs	2
Stator inductance	0.0103 H
Rotor inductance	0.0154 H
Mutual inductance	0.2655 H
Inertia	0.0012 kgm^2
Rated speed	1725 rpm

References

1. O.V. Thorsen, and M. Dalva, "Survey of faults on induction motors in offshore oil industry, petrochemical industry, gas terminals, and oil refineries", *IEEE Transactions on Industry Applications*, vol. 31, no. 5, pp. 1186-1196, Sept-Oct. 1995.
2. B. K. Bose, *Power Electronics & AC Drives,* Englewood Cliffs, NJ: Prentice Hall International, 1986).
3. S. J. Chapman, *Electric Machinery Fundamentals*, New York: McGraw-Hill Inc., 1991.
4. R. B. Sepe and J. H. Lang, "Inverter nonlinearities and discrete-time vector current control", *IEEE Trans. Industry Applications*, vol. 30, no.1, pp. 62–70, Jan./Feb. 1994.
5. D. Antic, J. B. Klaassens, and W. Deleroi, Side effects in low-speed AC drives, in Conf. record of IEEE-PESC, pp. 998–1002, 1994.
6. D-W Chung, S-K SUI and D-C Lee, "Analysis and Compensation of Current Measurement Error in Vector Controlled AC Motor Drives", *IEEE Trans. Ind. Applications*, vol. 34, pp. 340-345, March/April 1998.
7. J. W. Choi, S-S Lee, S-Y Yu, S-J Jang, "Novel periodic torque ripple compensation scheme in vector controlled AC motor drives" Conference Proceedings of IEEE APEC, 1998, pp. 81-85.
8. C. Bonivento, A. Isidori, L. Marconi, A. Paoli, "Implicit fault-tolerant control: application to induction motors", Automatica, vol. 40, no. 3, pp. 355-371, March 2004.
9. R. Barro, P. Hsu, "Torque ripple compensation of vector controlled induction machines" Conf. record of IEEE PESC, pp. 1281-1287, 1997.
10. C. T. Lin, C. S. G. Lee, *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*, Prentice Hall, 1996.
11. B. Liang, A. D. Ball, S. D. Iwnicki, "Simulation and fault detection of three-phase induction motors", Proceedings of IEEE Region 10 Annual International Conference (TENCON), vol. 3, pp. 1813-1817, 2002.
12. Jhy-Shing Roger Jang, Chuen-Tsai Sun, Eiji Mizutani, *Neuro-Fuzzy and Soft Computing: Computational Approach to Learning and Machine Intelligence*, Prentice Hall, 1997.
13. A. Rubaai, D. Ricketts, and M. D. Kankam, **"**Development and implementation of an adaptive fuzzy-neural-network controller for brushless drive" *IEEE Transactions on Industry Applications*, Mar/Apr, 2002, pp. 441-447.
14. M. N. Uddin and Hao Wen, "Development of a Self-Tuned Neuro-Fuzzy Controller for Induction Motor Drives", *IEEE Transactions on Industry Applications*, Vol. 43, No. 4, July/August 2007, pp. 1108 - 1116.
15. B. Karanayil, M. F. Rahman and C. Grantham, "Online Stator and Rotor resistance Estimation Scheme using Artificial Neural Networks for Vector Controlled Speed Sensorless Induction Motor drive", *IEEE Trans. on Industrial Electronics*, Vol. 54, No. 1, Feb. 2007, pp. 167-176.
16. Matlab, Simulink User Guide. The Math Works Inc., 2003.
17. dSPACE Implementation Guide, Paderborn, Germany, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Novel Approach on Poly Phase to Single Phase Buck Boost Matrix Converter

Pushpakaran S, Umamaheswari B

EEE Department, Anna University
Chennai, India
E-mail: spushpakaran@gmail.com,umamahesb@annauniv.edu

Abstract - Matrix Converter is an array of controlled bidirectional switches connecting directly between the source and the load. This paper presents a comparison between the different modulation strategies employed in the matrix converter and presents an approach to increase the voltage transfer ratio beyond unity. Here we aimed at obtaining a single phase output from a poly phase input supply. A new fuzzy modulation scheme is implemented for the different configurations of poly phase to single phase matrix converter to increase the voltage gain compared to the previous method.

I. Introduction

The matrix converter is a forced commutated direct AC – AC converter which uses an array of controlled bidirectional switches that directly connects input to output, without any dc link in the intermediate stage, to create a variable output voltage system with unrestricted frequency. The main advantages of the Matrix converter are the absence of bulky reactive elements that are subject to ageing and reduce the system reliability. Furthermore, matrix converter provides bidirectional power flow, nearly sinusoidal input and output waveforms and a controllable input power factor. Unlike conventional converter it does not have any dc link in the intermediate stage of power conversion which eliminates the need for large energy storage elements. Matrix converters can contribute to the realization of down sized, light-weight, long-life and high efficiency power supplies.

II. Early Stages of Modulation of Matrix Converter

The development of matrix converters starts with the work of Venturini and Alesina (1980).They developed a rigorous mathematical analysis to describe the low-frequency behavior of the converter which can provide a voltage transfer ratio of 0.5 and through improved venturini modulation method it has been improved to 0.866 by adding common mode voltages to the target output. Roy (1989) proposed the scalar modulation method to minimize the complexity of venturini.This method yields virtually identical switch timings compared with the modified venturini method. At maximum voltage gain of 86.6% both scalar modulation and Modified Venturini methods are identical. The above said direct matrix converter has limitation on maximum voltage gain, which can be improved to 105.3% by indirect modulation method. Rodriguez (1983) introduced the idea of "fictitious dc link" by introducing a concept of indirect transfer function approach. Ziogas *et al. (*1985 –1986) expanded the "fictitious dc link" idea of Rodriguez and provided a rigorous mathematical explanation. Braun (1983), Kastner and Rodriguez (1985) introduced the use of space vectors in the analysis and control of matrix converters. In 1989, Huber *et al.* published the first of a series of papers in which he gave the principles of space-vector modulation applied to the matrix converter modulation problem. Casadei (2002) proposed a modulation algorithm to increase the maximum voltage transfer ratio to 115.5% using a new duty cycle space vector modulation. Fang Lin Luo et al. (2006) proposed sub envelope modulation strategies to reduce the total harmonic distortion.

Here we present a new fuzzy modulation scheme to simplify the complications present in it and to improve the voltage gain compared to the previous methods.

III. Basic Criteria for Matrix Converter

The input signals should be switched such that the input should not be short circuited and as the load is normally inductive in nature the output should not be open circuited.

IV. Poly Phase to Single Phase Matrix Converter

Many modulation strategies which were generalized will have a limitation on the voltage gain for matrix converter other than three phase to three phase structure. Here we have proposed a new fuzzy logic modulation strategy to maximize the voltage gain for different configurations of poly phase to single phase matrix converter. We have selected the per unit values of the desired output voltage signal and the input voltage signals as input to the fuzzy logic modulator. Fuzzy rules were formed by comparing the per unit values of the desired output voltage and the input voltage signal. The poly phase to single phase matrix converter configuration using fuzzy logic modulation shown here can be extended to multi phase to multi phase configuration of matrix converter.

A. Single Phase to Single Phase Matrix Converter

The switching arrangement of single phase to single phase matrix converter (SPMC) using four bidirectional switches

is shown in Fig. 1, where the switches S1 and S4 or S2 and S3 would turn on at the same instant so that current can flow to the load in both directions.

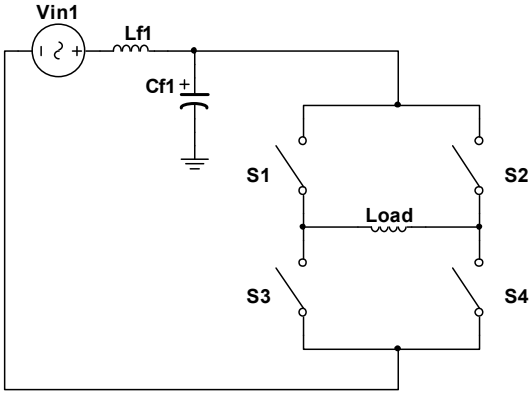


Fig. 1 Single Phase to Single Phase Matrix Converter

Let the input voltage is given by

$$V_{in}(t) = V_{im}\cos(w_i t) \tag{1}$$

and the desired output voltage be

$$V_o(t) = V_{om}\cos(w_o t) \tag{2}$$

The output voltage is obtained by modulating the input signal.

$$V_o(t) = S(t)V_{in}(t) \tag{3}$$

The switching function S(t) can be obtained through fuzzy modulator. The input to the fuzzy modulator is the per unit values of the desired output voltage $V_o(t)\ p.u.$ and the input voltage $V_{in}(t)\ p.u.$ The fuzzy membership function for the input signal to the fuzzy modulator is shown in Fig. 2.

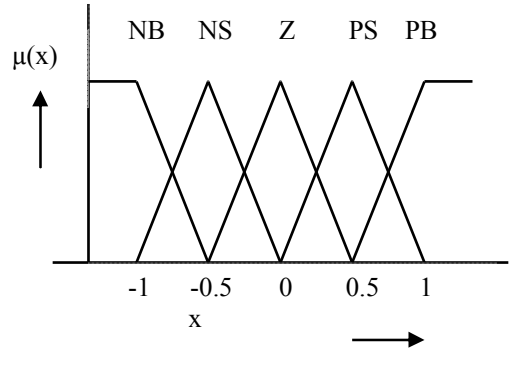


Fig. 2 Membership functions for V_{in} & V_o in p.u.

Table 1 Fuzzy rule table for SPMC.

V_{in} / V_0	PB	PS	Z	NS	NB
PB	P	P	N	N	N
PS	P	P	P	N	N
Z	N	P	P	N	P
NS	N	N	N	P	P
NB	N	N	P	P	P

The output of the fuzzy modulator is either P (+1) or N(-1) so that the load is connected to the input at any instant. The fuzzy rules were formed by considering the p.u. magnitude of V_{in} (t) & V_o (t) as given in table 1.The two possible switching combinations based on the fuzzy modulator output for a 1X1 matrix converter are

- S1 and S4 (+1)
- S2 and S3 (-1)

By proper modulation using fuzzy modulator the output voltage can be obtained to have a maximum voltage transfer ratio of 0.8611 which is higher than the previous mentioned methods for SPMC.

B. Two Phase to Single Phase Matrix Converter

The two phase to single phase matrix converter using six switches is shown in Fig. 3 where atleast one of the input signal is connected to the load at any instant.

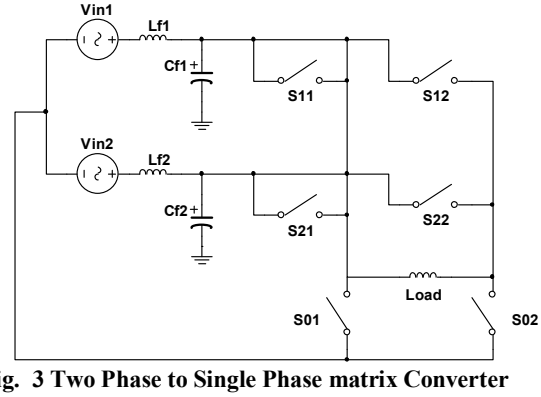


Fig. 3 Two Phase to Single Phase matrix Converter

Let the input voltage be given by

$$\begin{aligned} v_{in1}(t) &= V_{im}\cos(w_i t) \\ v_{in2}(t) &= V_{im}\cos(w_i t + 2\pi/3) \end{aligned} \tag{4}$$

and the p.u. values are

$$\begin{aligned} v_{i1}(t)_{\ p.u.} &= \cos(w_i t) \\ v_{i2}(t)_{\ p.u.} &= \cos(w_i t + 2\pi/3) \end{aligned}$$

The input current is given by

$$\begin{aligned} i_{i1}(t) &= I_{im}\cos(w_i t) \\ i_{i2}(t) &= I_{im}\cos(w_i t + 2\pi/3) \end{aligned} \tag{5}$$

Let the desired output voltage be

$$V_0(t) = V_{om}\cos(w_0 t) \tag{6}$$

Let the p.u. value of the output voltage be given by

$$V_o(t)_{\ p.u.} = \cos(w_0 t)$$

The output voltage can be obtained by modulating the input voltage.

$$V_o(t) = S(t)\begin{bmatrix} v_{in1}(t) \\ v_{in2}(t) \end{bmatrix} \quad (7)$$

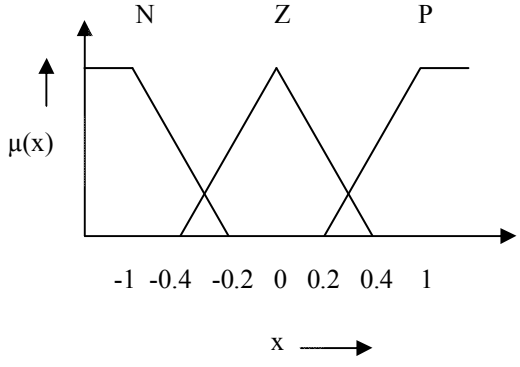


Fig. 4 Membership functions of $V_{i1}(t)$,$V_{i2}(t)$ & $V_o(t)$.

The switching function S(t) can be obtained through Fuzzy modulator and the membership function for the three inputs of the fuzzy modulator, $V_{i1}(t)$, $V_{i2}(t)$ & $V_o(t)$ p.u. is shown in Fig. 4. Fuzzy rules are framed by comparing Vo p.u. with V_{i1} p.u. & V_{i2} p.u. 27 possible fuzzy rules are framed for two phase to single phase matrix converter as shown in table 2.

Table 2 Fuzzy rule variable for two phase to single phase matrix converter.

V_i / V_0	$V_{i1}(t)$ is P			$V_{i1}(t)$ is Z			$V_{i1}(t)$ is N		
	$V_{i2}(t)$			$V_{i2}(t)$			$V_{i2}(t)$		
	P	**Z**	**N**	**P**	**Z**	**N**	**P**	**Z**	**N**
P	P Z	P Z	P N	Z P	P N	Z N	N P	N Z	Z N
Z	P N	Z P	P Z	P Z	N P	N Z	P Z	Z N	P N
N	Z N	N Z	N P	Z N	P N	Z P	P N	P Z	Z P

The two outputs of fuzzy modulator, the modulating signal which has values of +1 or 0 or -1 (i.e., P, Z, N) are used to switch any one of the six possible combinations.

The six possible switching combinations for the corresponding fuzzy modulator output are

- S11 & S22 (+1 & -1)
- S12 & S21 (-1 & +1)
- S11 & S02 (+1 & 0)
- S21 & S02 (0 & +1)
- S12 & S01 (-1 & 0)
- S22 & S01 (0 & -1)

By using fuzzy modulator for a two phase to single phase matrix converter, the output voltage can be obtained with a maximum voltage transfer ratio of 1.4832.

C. Three Phase to Single Phase matrix Converter (3X1 MC)

The three phase to single phase matrix converter using eight bidirectional switches is shown in Fig. 5

Let the input voltage is given by

$$\begin{aligned} v_{in1}(t) &= V_{im}\cos(w_i t) \\ v_{in2}(t) &= V_{im}\cos(w_i t + 2\pi/3) \\ v_{in3}(t) &= V_{im}\cos(w_i t + 4\pi/3) \end{aligned} \quad (8)$$

and the p.u. values with V_{im} as base value are

$$\begin{aligned} v_{i1}(t)_{p.u.} &= \cos(w_i t) \\ v_{i2}(t)_{p.u.} &= \cos(w_i t + 2\pi/3) \\ v_{i3}(t)_{p.u.} &= \cos(w_i t + 4\pi/3) \end{aligned}$$

The input current is given by

$$\begin{aligned} i_{i1}(t) &= I_{im}\cos(w_i t) \\ i_{i2}(t) &= I_{im}\cos(w_i t + 2\pi/3) \\ i_{i3}(t) &= I_{im}\cos(w_i t + 4\pi/3) \end{aligned} \quad (9)$$

Let the desired output voltage be

$$V_o(t) = V_{om}\cos(w_o t) \quad (10)$$

Let the p.u. value of the output voltage be given by

$$V_{o\ p.u.} = \cos(w_o t)$$

The output voltage can be obtained by modulating the input voltages.

$$V_o(t) = S(t)\begin{bmatrix} V_{in1}(t) \\ V_{in2}(t) \\ V_{in3}(t) \end{bmatrix} \quad (11)$$

At any instant a minimum of one input to a maximum of two inputs is connected to the load to satisfy the basic criteria. The switching configurations S(t) are generated using fuzzy modulator. The membership function for the inputs to the fuzzy modulator is given in Fig. 6.The membership function for the three outputs of fuzzy

modulator have values +1, 0, -1. (i.e., P, Z, N). Eighty one possible fuzzy rules are framed by comparing the p.u. magnitude of the desired output voltage with the three input p.u. voltages.

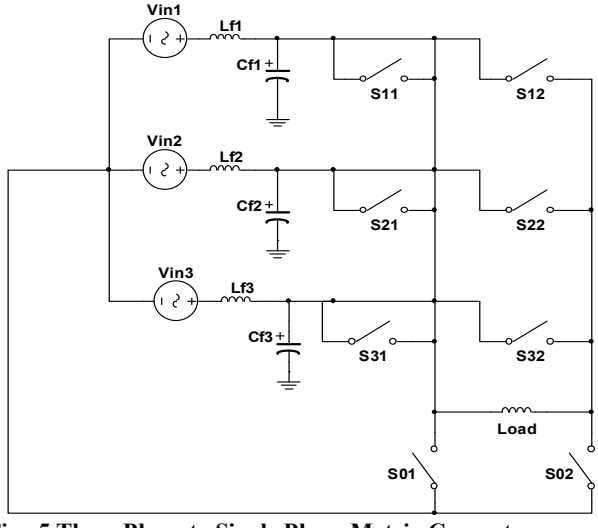


Fig. 5 Three Phase to Single Phase Matrix Converter

As it is difficult to place here all the possible combinations in 4 dimensions, a few rules are shown here.

- if V_o is P and V_{in1} is P and V_{in2} is N and V_{in3} is Z then V_{o1} is P and V_{o2} is N V_{o3} is Z.
- if V_o is N and V_{in1} is P and V_{in2} is N and V_{in3} is Z then V_{o1} is N and V_{o2} is P V_{o3} is Z.
- if V_o is Z andV_{in1} is P and V_{in2} is N and V_{in3} is P then V_{o1} is P and V_{o2} is Z V_{o3} is N.

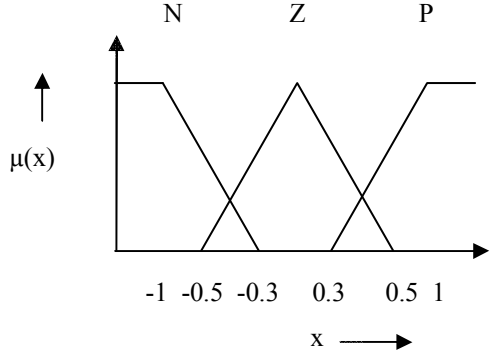


Fig. 6 Membership functions for input signals to the fuzzy modulator

The twelve possible switching configurations based on the three outputs of the fuzzy modulator are given as follows.

- S11 & S22 (+1 -1 0)
- S11 & S32 (+1 0 -1)
- S21 & S32 (0 +1 -1)
- S12 & S21 (-1 +1 0)
- S12 & S31 (-1 0 +1)
- S22 & S31 (0 -1 +1)
- S11 & S02 (+1 0 0)
- S21 & S02 (0 +1 0)
- S31 & S02 (0 0 +1)
- S12 & S01 (-1 0 0)
- S22 & S01 (0 -1 0)
- S32 & S01 (0 0 -1)

For a two phase to single phase matrix converter, by using fuzzy modulation the voltage transfer ratio has been improved to a maximum value of 1.762315.

V. Simulation Results

The three different configurations of poly phase to single phase matrix converter are simulated using MatLab/simulink and the simulation results are shown with harmonic analysis.

A. Single Phase to Single Phase MC

From equations (1), (2) & (3) with

$$V_{im} = 100V$$

$$w_i = 2\pi * 50\, rad/\sec \&$$

$$w_o = 2\pi * 1000\, rad/\sec$$

The simulated output voltage using MatLab/Simulink is shown in Fig. 7 where we get a voltage gain of 86.11%

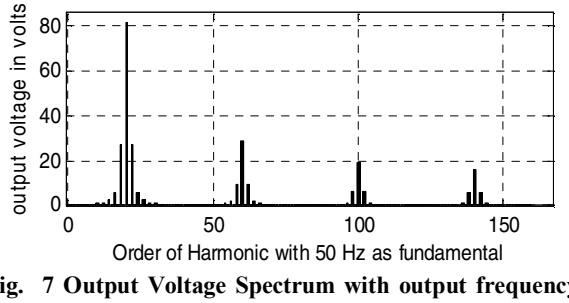


Fig. 7 Output Voltage Spectrum with output frequency of 1000 Hz & input frequency 50 Hz

The output current spectrum is shown in Fig. 8.

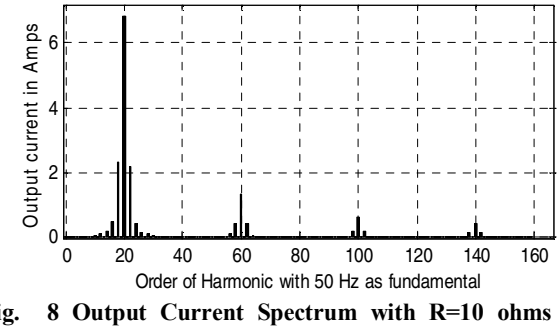


Fig. 8 Output Current Spectrum with R=10 ohms and L=1.1 mH

The output voltage waveform is shown in Fig. 9.

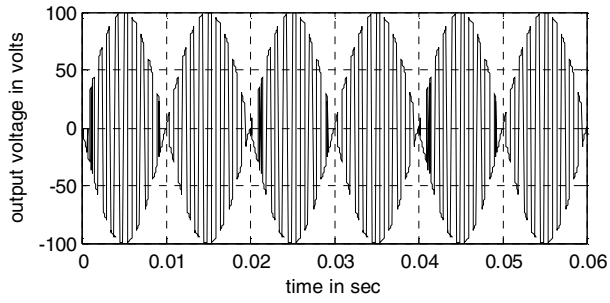


Fig. 9 output voltage waveforms for the above setup

B. Two Phase to Single Phase MC

From equations (4), (5), (6) & (7) with

$$V_{im} = 100V$$
$$w_i = 2\pi * 50\, rad/\sec \&$$
$$w_o = 2\pi * 1000\, rad/\sec$$

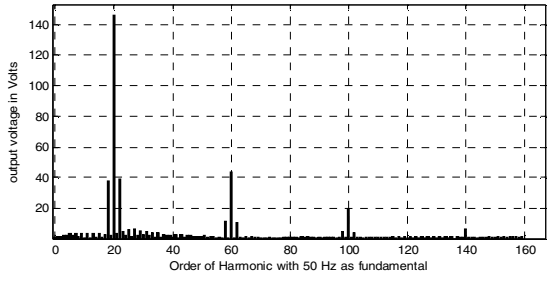


Fig. 10 Output Voltage Spectrum with output frequency of 1000 Hz & input frequency 50 Hz

The simulated output voltage and current spectrum is shown in Fig. 10 where we get a voltage gain of 148.32%. The output current spectrum is shown in Fig. 11 and the output voltage waveform is shown in Fig. 12.

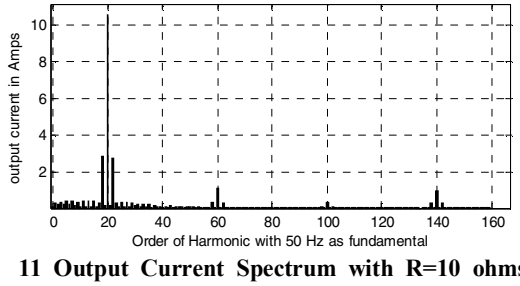


Fig. 11 Output Current Spectrum with R=10 ohms and L=1.1 mH

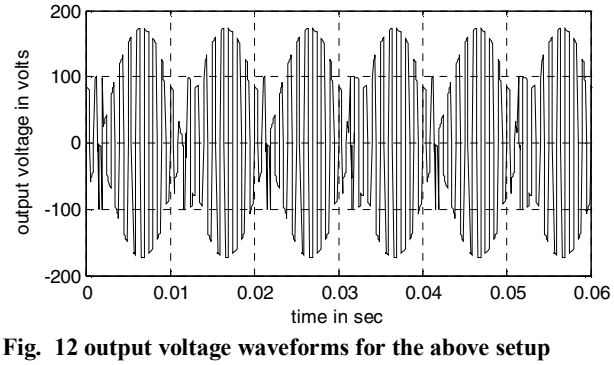


Fig. 12 output voltage waveforms for the above setup

C. Three Phase to Single Phase MC

From equations (8), (9), (10) & (11) with

$$V_{im} = 100V$$
$$w_i = 2\pi * 50\, rad/\sec \&$$
$$w_o = 2\pi * 1000\, rad/\sec$$

the simulated output voltage and current spectrum is shown in Fig. 13 where we get a voltage gain of 176.2315%.

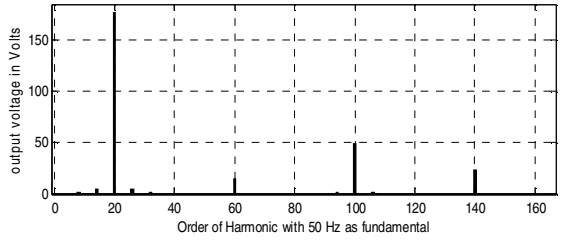


Fig. 13 Output Voltage Spectrum with output frequency of 1000 Hz & input frequency 50 Hz

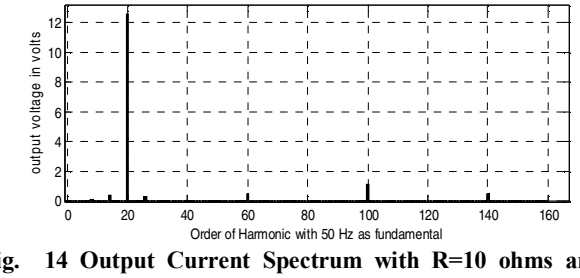


Fig. 14 Output Current Spectrum with R=10 ohms and L=1.1 mH

The output current spectrum is shown in Fig. 14 and output voltage waveform is shown in Fig. 15.

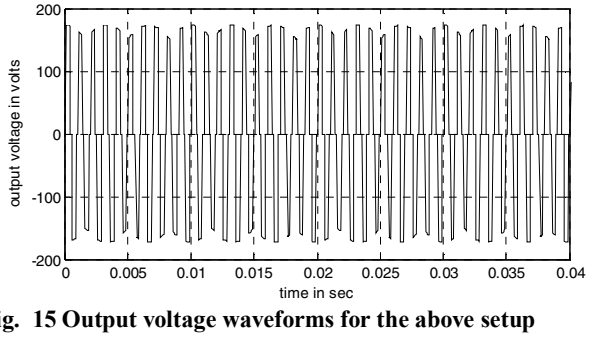


Fig. 15 Output voltage waveforms for the above setup

VI. Analysis of harmonics

Of the harmonics present in the above said poly phase to single phase matrix converter the third and fifth harmonic content is very important and the other harmonics are suppressed. The harmonics present in the different configurations of poly phase to single phase matrix converter are given in table 3.

Table 3 Harmonic contents present in poly phase to single phase matrix converter.

Order of Harmonic	Harmonic content present in output as p.u. of the input		
	1X1 MC	2X1 MC	3X1 MC
1	0.8611	1.4832	1.762315
3	0.283213	0.42958	0.145206
5	0.19103	0.20069	0.287697
7	0.160629	0.060045	0.121486

VII. Conclusion

The fuzzy modulated poly phase to single phase matrix converter gives better voltage gain than the previous methods and can be extended to poly phase to poly phase matrix converters. The performance can still be improved by increasing the membership functions for each input.

References

[1] Wheeler P.W., Rodriguez J, Clare J.C, Empringham L, Weinstein A, "Matrix converters: a technology review," IEEE Transactions on Industrial Electronics, Vol. 49 (2), pp. 276 – 288, April 2002.

[2] M. Venturini and A. Alesina, "Solid state power conversion: A Fourier analysis Approach to generalized Transformer Synthesis." in Proc. IEEE CAS, vol. 28, No. 4, pp. 319–330, April 1981.

[3] G. Roy and G. E. April, "Cycloconverter operation under a new scalar control algorithm," in Proc. IEEE PESC'89, , pp. 368–375, 1989.

[4] G. Roy and G. E. April, "Direct Frequency Changer Operation Under a New Scalar Control Algorithm", IEEE Trans, on Power Elect., vol. 6, no. 1, pp. 100-107, 1991.

[5] P. D. Ziogas, S. I. Khan, and M. H. Rashid, "Analysis and design of forced commutated cycloconverter structures with improved transfer characteristics," IEEE Trans. Ind. Electron., vol. IE-33, pp. 271–280, August 1986.

[6] H. Mohd Hanafi," Modelling & Simulation of Single-phase Matrix Converter as a Frequency Changer with Sinusoidal Pulse Width Modulation Using MATLAB/Simulink", First International Power and Energy Coference PECon 2006, Putrajaya, Malaysia, pp 482 -487, November 28-29, 2006.

[7] D. Casadei, G. Serra, A. Tani, and L. Zarri, "Matrix converter modulation strategies: A new general approach based on space-vector representation of the switch state," IEEE Trans. Ind. Electron., vol. 49, pp 370–381, April 2002.

[8] Wheeler, Clare,, Empringham, Bland, Erris, "Matrix converters," IEEE Industry Applications Magazine,Vol. 10 (1), pp. 59 – 65, Jan-Feb2004.

[9] A.Zuckerberger, D.Weinstock, A.Alexand rovitz, "Single-phase matrix converter", IEE Pmc.-Electr. Power Appl., Vol. 144, No. 4, July 1997.

[10] L Empringham, P W Wheeler, J C Clare, "Matrix Converter Bi - directional switch commutation using Intelligent Gate Drives" in Power Electronics and Variable Speed Drives, Conference Publication No.456, pp. 626 – 631, 21-23 September 1998.

[11] L. Huber, D. Borojevic, and N. Burany, "Voltage space vector based PWM control of forced commutated cycloconvertors ," in Proc. IEEE IECON'89, pp. 106 – 111, 1989.

[12] M. Braun and K. Hasse, "A direct frequency changer with control of input reactive power," in Proc. IFAC Control in Power Electronics and Electrical Drives Conf., Lausanne, Switzerland, pp. 187–194, 1983.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Modified Kalman Filter based Direct Torque Control of Induction motor for Ripple Free Torque and Flux Estimation

G. Pandian
Research Scholar
Electrical and Electronics Engg Dept
Sathyabama University
Chennai, India
E-mail: pandian1960@yahoo.co.in

S. Rama Reddy
Professor
Electrical and Electronics Engg Dept
Jerusalem College of Engineering
Chennai, India
E-mail: srr_victory@yahoo.com

Abstract*—This paper presents a robust sensorless Direct Torque Control (DTC) method for induction motor (IM) for estimation of stator flux components and rotor speed based on the Modified Kalman Filter (MKF). The model of IM and its MKF models in Matlab /Sirnulink simulation environment are developed. The proposed MKF speed and flux estimation method is also proved insensitive to the IM parameter variations. Simulation results demonstrate a good performance and robustness.*

Keywords*: DTC, Modified Kalman Filter (MKF), IM, sensorless control, anti-windup PI.*

1. Introduction

In recent years, several studies have been developed which propose alternative solutions to the FOC control of a PWM inverter-fed motor drive with two objectives: first, achievement of an accurate and fast response of the flux and the torque, and second, reduction in the complexity of the control system. Among the various proposals, Direct Torque and Flux Control (DTFC) also called Direct Torque Control (DTC), has found wide acceptance, [2].

Since its introduction in 1985, the direct torque control (DTC) [1], principle was widely used for IM drives with fast dynamics. Despite its simplicity, DTC is able to produce very fast torque and flux control, if the torque and the flux are correctly estimated, is robust with respect to motor parameters and perturbations. As it is well known, speed sensors like tachometers or incremental encoders increase the size and the cost of systems unnecessarily. Similar problems arise with the addition of search coils or Hall Effect sensors to the motor for the measurement of flux, hindering functionality in terms of implementation. Thus, to improve the overall system performance, state estimators or observers are usually more preferable than physical measurements, [3, 4, 7, 11, and 13].

The objectives of sensorless drives control are:

- Reduction of hardware complexity and cost,
- Increased mechanical robustness,
- Operation in hostile environments,
- Higher reliability,
- Unaffected machine inertia.

The Modified Kalman Filter is an optimal stochastic observer in the least-square sense for estimating the states of dynamic non-linear systems, and provides optimal filtering of the noises in measurement and inside the system if the covariances of these noises are known. Hence it is a viable candidate for the on-line determination of the speed of IM, MKF is considered to be suitable for use in high-performance induction motor drives, and it can provide accurate speed estimates in a wide speed-range, including very low speed and seem to be between the most promising methods thanks to their good performance. They have the advantage to provide both flux and mechanical speed estimates without problems of open-loop integration.

For the speed regulation, the saturation of the manipulated variable can involve a phenomenon of racing of the integral action during the great variations (starting of the machine), which is likely to deteriorate the performances of the system or even to destabilize it completely, the solution consists in correcting the integral action.

The contribution of this study is the development of an MKF based speed sensorless DTC system for an improved performance, the especially against variations in the load torque. The developed MKF algorithm involves the estimation of speed and stator flux. The performance of the control system with the proposed MKF algorithm has been demonstrated with simulations *using* Matlab /Sirnulink.

2. Principle of the DTC

DTC is a control philosophy exploiting the torque and flux producing capabilities of ac machines when fed by a voltage source inverter that does not require current regulator loops, still attaining similar performance to that obtained from a vector control drive.

2.1. Stator flux orientation

In the reference (α, β), the stator flux can be obtained by the following equation:

$$\overline{V}_s = R_s \overline{I}_s + \frac{d}{dt}\overline{\Psi}_s \quad (1)$$

By neglecting the voltage drop due to the resistance of the stator to simplify the study (for high speeds), we find:

$$\overline{\Psi}_s \approx \overline{\Psi}_{s0} + \int_0^t \overline{V}_s dt \quad (2)$$

For one period of sampling, the voltage vector applied to the asynchronous machine remains constant, we can write:

$$\overline{\Psi}_s(k+1) \approx \overline{\Psi}_s(k) + \overline{V}_s T_e \quad (3)$$

3.2. Torque Production

The electromagnetic torque is proportional to the vector product between the stator and rotor flux according to the following expression:

$$C_e = k(\overline{\Psi}_s \times \overline{\Psi}_s) = k|\Psi_s||\Psi_r|\sin(\delta) \quad (4)$$

3.3. Development of the commutation strategy

Table I, shows the commutation strategy suggested by Takahashi, [1] to control the stator flux and the electromagnetic torque of the IM.
The Fig. 1 gives the partition of the complex plan in six angular sectors $S_{I=1\ldots6}$.

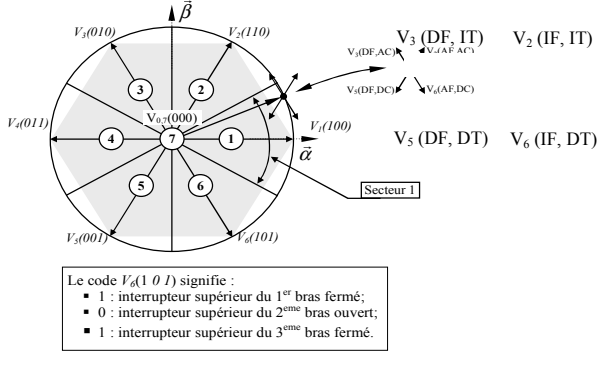


Fig. 1. Partition of the complex plan in six angular sectors $S_{I=1\ldots6}$.

IT: Increase the Torque, DT: Decrease the Torque.
IF: Increase the Flux, DF: Decrease the Flux.

Table. 1 Selection table for direct torque control.

$\Delta\Psi_s$	ΔC_e	S_1	S_2	S_3	S_4	S_5	S_6
1	**1**	V_2	V_3	V_4	V_5	V_6	V_1
	0	V_7	V_0	V_7	V_0	V_7	V_0
	-1	V_6	V_1	V_2	V_3	V_4	V_5
0	**1**	V_3	V_4	V_5	V_6	V_1	V_2
	0	V_0	V_7	V_0	V_7	V_0	V_7
	-1	V_5	V_6	V_1	V_2	V_3	V_4

4. Development of the MKF algorithm

The Kalman filter is a well-known recursive algorithm that takes the stochastic state space model of the system together with measured outputs to achieve the optimal estimation of states [8,9,10,12]. The optimality of the state estimation is achieved with the minimization of the mean estimation error. MKF, is used for the estimation of $\hat{I}_{s\alpha}$, $\hat{I}_{s\beta}$, $\hat{\Psi}_{s\alpha}$, $\hat{\Psi}_{s\beta}$, and $\hat{\omega}_r$.

Fig. 2 shows the structure of a Kalman filter.

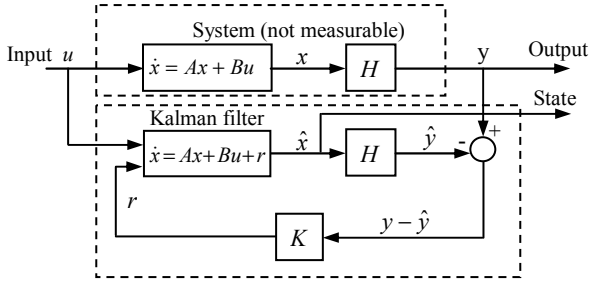


Fig. 2. Structure of Kalman filter estimator

The discrete model of the IM can be given as follows:

$$\begin{aligned} \overline{x(k)} &= f(x(k), u(k), k) + G(k)w(k) \\ \overline{y(k)} &= h(x(k), k) + v(k) \end{aligned} \quad (5)$$

With $w(k)$: is the measurement noise and v (k): is the process noise.

5. Application of the Modified Kalman filter

The speed estimation algorithm of the Modified Kalman filter can be simulated by the Matlab/Simulink software, which consists of an S-Function block as shown in Fig.3.

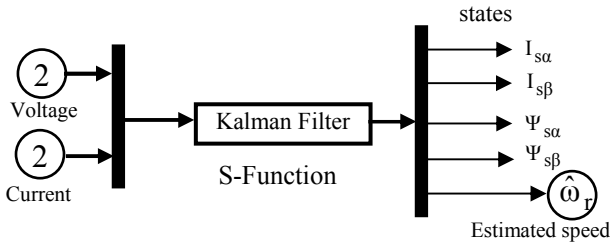


Fig. 3. Simulink model of MKF speed estimation.

5.1. Prediction of the state vector

Prediction of the state vector at sampling time (k+1), from the input u (k), state vector at previous sampling time x(k/k) .

$$\hat{x}(k+1/k) \hat{=} F(\hat{x}(k/k), u(k)) \quad (6)$$

Where:

$$F = \begin{bmatrix} (1-T_s\gamma)i_{s\alpha} + T_s\frac{MR_r}{L_r^2K}\psi_{r\alpha} + T_s\frac{M\omega_r}{L_rK}\psi_{r\beta} + T_s\frac{1}{K}V_{s\alpha} \\ (1-T_s\gamma)i_{s\beta} - T_s\frac{M\omega_r}{L_rK}\psi_{r\alpha} + T_s\frac{MR_r}{L_r^2K}\psi_{r\beta} + T_s\frac{1}{K}V_{s\beta} \\ T_s\frac{M}{T_r}i_{s\alpha} + \left(1-T_s\frac{1}{T_r}\right)\psi_{r\alpha} - T_s\omega_r\psi_{r\beta} \\ T\frac{M}{T_r}i_{s\beta} + Tw_r\psi_{r\beta} + \left(1-T\frac{1}{T_r}\right)\psi_{r\beta} \\ w_r \end{bmatrix}$$

5.2. Prediction covariance computation

The prediction covariance is updated by:

$$P(k+1/k) = F(k)P(k)F(k)^T + Q \quad (7)$$

Where:

Q: covariance matrix of the system noise

$$F(k) = \left.\frac{\partial f}{\partial x}\right|_{x(k)=\tilde{x}(k/k)} \quad (8)$$

With:

$$\frac{\partial F}{\partial x} = \begin{bmatrix} 1-T_s\gamma & 0 & T_s\frac{MR_r}{L_r^2K} & T_s\frac{M\omega_r}{L_rK} & T_s\frac{M}{L_rK}\psi_{r\beta} \\ 0 & 1-T_s\gamma & -T_s\frac{M\omega_r}{L_rK} & T_s\frac{MR_r}{L_r^2K} & -T_s\frac{M}{L_rK}\psi_{r\alpha} \\ T_s\frac{M}{T_r} & 0 & 1-T_s\frac{1}{T_r} & T_s\omega_r & T_s\psi_{r\beta} \\ 0 & T_s\frac{M}{T_r} & T_s\omega_r & 1-T_s\frac{1}{T_r} & T_s\psi_{r\alpha} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\frac{\partial h}{\partial x} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (9)$$

5.3. Kalman Gain Computation

The Kalman filter gain (correction matrix) is computed as:

$$L(k+1) = P(k+1/k).C(k)^T.(C(k)P(k+1/k)C(k)^T + R)^{-1} \quad (10)$$

With:

$$C(k) = \left.\frac{\partial c(x(k))}{\partial x(k)}\right|_{x(k)=\hat{x}(k)} \quad (11)$$

5.4. State Vector Estimation

The predicted state-vector is added to the innovation term multiplied by Kalman gain to compute state-estimation vector. The state-vector estimation (filtering) at time (k)
is determined as:

$$\hat{x}(k+1/k+1) = \hat{x}(k+1/k) + L(k+1)(y(k+1) - C\hat{x}(k+1/k)) \quad (12)$$

6. Determination of the noise and state covariance matrices

The goal of the Kalman filter is to obtain unmeasurable states (i.e. covariance matrices Q, R, P of the system noise vector, measurement noise vector, and system state vector (x) respectively). In general, by means of noise inputs, it is possible to take computational inaccuracies, modelling errors, and errors in measurements into account in modelling the system. The filter estimation ($\hat{x}$) is obtained from the predicted values of the states (x) and this is corrected recursively by using a correction term, which is product of the Kalman gain (L) and the deviation of the estimated measurement output vector and the actual output vector ($(y - C\hat{x})$). The system noise covariance matrix (Q) is [5x5], and the measurement noise covariance matrix (R) is [2x2] matrix, Q and R are diagonal, and only 5 elements must be known in Q and 2 elements in R.

7. System of speed regulation

The saturation of the manipulated variable can involve a phenomenon of racing of the integral action during the great variations (starting of the machine), which is likely to deteriorate the performances of the system or even to destabilize it completely. To overcome this phenomenon, a solution consists in correcting the integral action according to the diagram of Fig. 4. [5, 6]

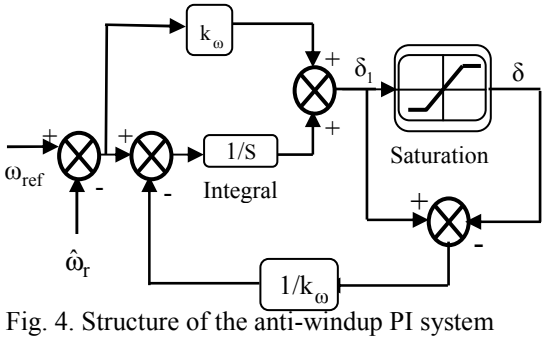


Fig. 4. Structure of the anti-windup PI system

The stator flux is a function of the rotor flux which represented by:

$$\begin{cases} \Psi_{s\alpha} = \sigma L_s i_{s\alpha} + \dfrac{M}{L_r}\Psi_{r\alpha} \\ \Psi_{s\beta} = \sigma L_s i_{s\beta} + \dfrac{M}{L_r}\Psi_{r\beta} \end{cases} \quad (12)$$

8. Proposed sensorless drive

The proposed sensorless IM drive is shown in Fig. 5. The drive operates at constant stator flux uses DTC to provide torque control. The speed controller is an anti-windup PI regulator that generates the reference torque. The stator flux is estimated by the MKF and used in the DTC control.

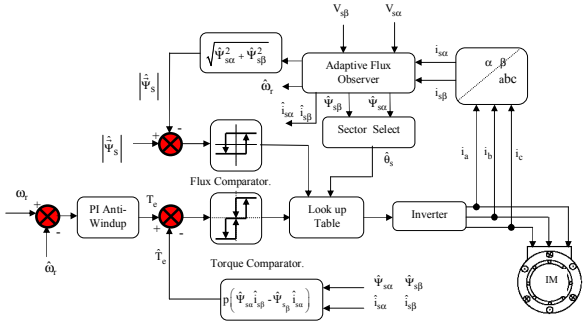


Fig. 5. Speed Sensorless Direct Torque Control System Using Modified Kalman Filter.

9. Sensitivity study and simulation results

The sensorless IM drive of Fig. 5 was verified using simulations. In order to show the performances and the robustness of the MKF algorithm, we simulated different cases, which are presented thereafter. The static and dynamic performances of the MKF are analyzed according to the simulation of the following transients:

9.1. Comparison on the level of the regulation speed

Fig. 6 presents the actual, estimated speeds ω_r , $\hat{\omega}_r$ respectively. The estimated speed follows the real speed.

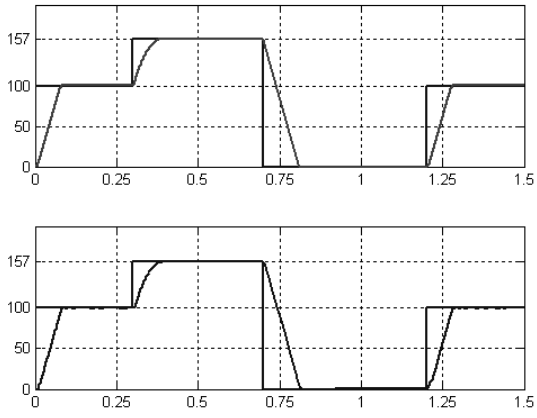


Fig. 6. Reference, Actual and Estimated Speed.

9.2. Inversion of the speed

To test the robustness of the system, we applied a changing of the speed reference from 157 rad/sec to -157 rad/sec at t=1.2s. Fig. 7 presents the actual, estimated speeds ω_r , $\hat{\omega}_r$ respectively and the estimation error (er) .The estimation algorithm is robust because the variation of the speed is important and the estimated speed follows the real speed when the motor starts and at the moment of speed inversion.

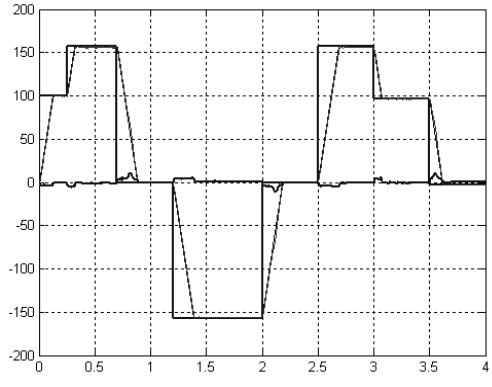


Fig. 7 High Speed, No-Load, Four Quadrant Speed Estimation with MKF

9.3. Operation at low speed

To test the speed estimation, simulation was established in low speed. Fig. 8. Illustrate simulation results of the process of speed estimation with a speed reference equal $W_{ref} = \pm 50$ (rad/sec). We can see that the speed follow perfectly the speed reference however Fig. 9 illustrates the trajectory of the estimated stator flux; the deviation detected is caused by the instantaneous reversal of the speed at the zero crossing of the speed. Fig. 10 presents actual flux $|\Psi_s|$. The estimated flux $|\hat{\Psi}_s|$ and estimation error (er).

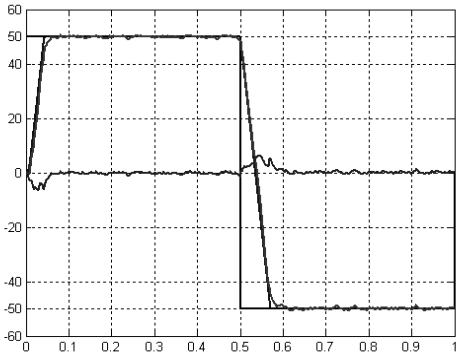


Fig. 8.Low Speed, Actual, Estimated and Estimation Error

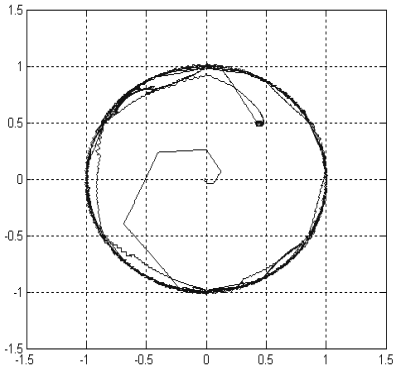


Fig. 9. Trajectory of the Estimated Stator Flux Components.

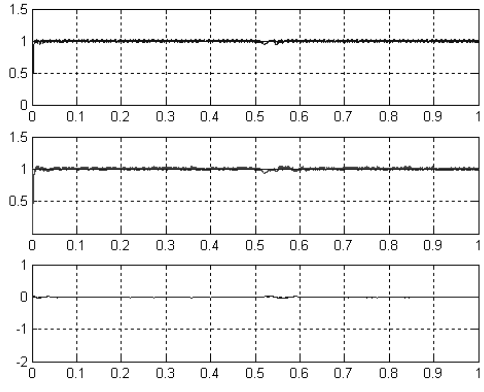


Fig. 10. The Actual, Estimated Stator Flux Magnitude and Estimation Error.

Fig. 11 illustrates the response of the estimated stator current and the electromagnetic torque. It should be noted that the amplitude of the torque ripple is slightly higher.

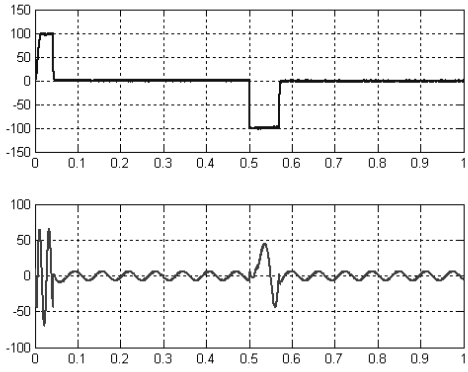


Fig. 11. Response of the Electromagnetic Torque and Estimated Stator Current.

9.4. Variation of load torques

In Fig.12, Fig. 13, rated mechanical load is applied to the motor between 0.5s-1s after a leadless starting. To verify the performance of MKF under loaded conditions. As shown above MKF works properly even under fully loaded case. We can see the insensibility of the control algorithm to load torque variation.

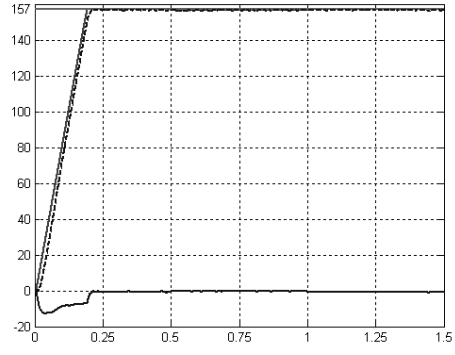


Fig. 12. High Speed, Full-Load, Speed Estimation.

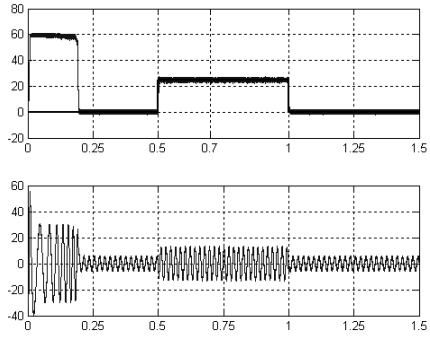


Fig. 13. Response of the Electromagnetic Torque and Estimated Stator Current.

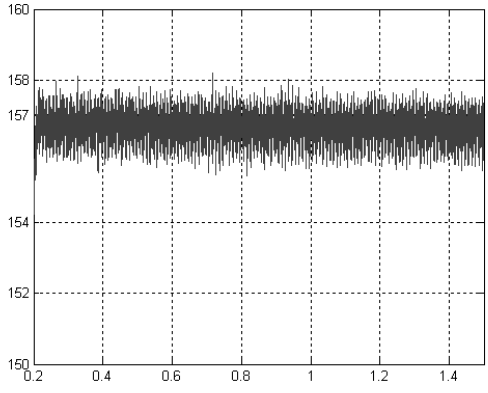


Fig. 14. Estimated Rotor Speed with Measured Noisy Current.

9.5. Injected noise to the stator currents

The aim of the current injection is to observe the low pass filter characteristics of MKF. As shown in Fig. 14, the estimated speed is not affected too much from the injected noise. The speed estimation accuracy may be increased by increasing the measurement noise covariance under noisy conditions thus; the system model will have more importance.

10. Conclusion

We have presented in this paper sensorless direct torque control of induction motor based on Modified Kalman Filter for stator flux and speed estimation. The filtering action of MKF improves the system performance, especially at low speeds. Simulation results reveal that the flux and speed tracking are good and error convergence is guaranteed. However an anti-windup PI regulator has been used to replace the classical PI controller in the speed control of a direct torque control. In conclusion, it seems that the anti-windup PI controller outperforms the classical PI controller in speed control of high performance DTC motor drive. This association makes the induction motor based DTC more robust and more stable.

ANNEXURE

Parameters

P = 5 HP f = 50 Hz W_r = 1440 rpm V_r = 380 V
R_s = 1.2 Ω R_r = 1.8 Ω L_s = 0.1554 H
L_r = 0.1568 H
T_r = 0.0871s M = 0.15 H J = 0.07 kgm^2 P = 2

12. References

1. I. Takahashi, and T. Noguchi, « A new quick-response and high efficiency control strategy of an induction machine, » IEEE Trans. Industry Appl, vol. 22, pp. 820-827, Sep/Oct 1986.
2. G. Bottiglieri, G. Scelba, and G.Scarcella, « Sensorless speed estimation in induction motor drives, » IEEE Electric Machines, IEMDC'03, pp. 624– 630, June 2003.
3. S. Belkacem, F. Naceri A. Betta, and L. Laggoune « Speed sensorless DTC of induction motor based on an improved adaptive flux observer, » IEEE Trans. Industry Appl, pp. 1192–1197, 14-17 Dec 2005. Hong Kong.
4. B. Akin, « State estimation techniques for speed sensorless field oriented control of induction motors, » M.Sc. Thesis EE Dept, METU, Aug 2003.
5. Y. Cao, Z. Lin, and D. G. Ward, « An anti-windup approach to enlarging domain of attraction for linear systems subject to actuator saturation, » IEEE Trans. Automat Control, vol. 47, pp. 140–145, Jan 2002.
6. L. Zaccarian, and A. Teel, « Nonlinear scheduled anti-windup design for linear systems, » IEEE Trans. On Automat Control, vol. 49, pp. 2055–2061, Nov 2004.
7. M. Barut, S. Bogosyan, and M.Gokasan, « EKF based sensorless direct torque control of IMs in the low speed range, » IEEE Trans. On Industry Electronics, ISIE2005. , pp. 969 - 974, June 2005.
8. A.V .Leite, R. E Araujo, and D. Freitas, « Full and reduced order Extended kalman filter for speed estimation in induction motor drives: a comparative study, » Power Electronics Specialists Conference, PESC 04. IEEE Vol. 3, pp. 2293 – 2299, 20-25, June 2004.
9. A.V Leite, R. E Araujo, and D. A Freitas, « new approach for speed estimation in induction motor drives based on a reduced-order Extended Kalman filter, » IEEE Industry Electronics, pp. 1221 – 1226, 4-7 May 2004.
10. M. Barut, O. S. Bogosyan, and M. Gokasan, « An EKF based reduced order estimator for the sensorless control of IMs, » IEEE pp.1256 - 1261, 23-25 June 2003.
11. S. Chávez Velázquez, R. Alejos Palomares, and A. Nava Segura, « Speed Estimation for an Induction Motor Using the Extended Kalman Filter,» IEEE Trans. On Electronics, Communications, 2004.
12. Sensorless Control with Kalman Filter on TMS320 Fixed-Point DSP, Texas Instruments July 1997. Available: http://www.ti.com.
13. S. Belkacem, L. Louanasse, H. Tamrabet, S. Zaidi, and B. Kiyyour," Performance Analysis of a Speed Sensorless Induction Motor Drive based on DTC scheme, «First International Conference on Electrical Systems PCSE'05,Oum El-Bouaghi, Algeria, pp. 267-272. May 7-9 2005

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Voltage Sag Compensation Utilizing Autotransformer Switched by Hysteresis Voltage Control

Amir Ahmad Koolaiyan, Abdolreza Sheikholeslami, Reza Ahmadi Kordkheili

Electrical and Computer Department, Nooshirvani Institute of Technology
Shariati Avenue, Babol, Mazandaran, Iran
E-mail: amir.koolaiyan@gmail.com

Abstract-This paper presents a new voltage sag compensator based on an autotransformer and an IGBT switched by hysteresis voltage control method. Faults occurring in power distribution systems generally cause the voltage sag or swell. Sensitivity to voltage sags and swells varies within different applications. For sensitive loads, even a voltage sag of short duration can cause serious problems. Hysteresis Voltage control method suggested in this paper is based on comparing the real voltage of system with a reference voltage. The proposed scheme is able to quickly recognize the voltage sag condition, and it can correct the voltage by boosting the input voltage during voltage sag events. One of the advantages of this topology is that it uses only one controlled switch per phase to boost the input voltage. Different voltage sag events have been simulated by MATLAB/Simulink software. The results of simulations verify the ability of proposed method to mitigate voltage sag events.

I. Introduction

With an increase in the use of sensitive loads, the power quality issues have become an increasing concern. Poor distribution power quality results in power disruption for the user and huge economical losses due to the interruption of production processes. Different power quality surveys done by researchers identify voltage sags as the most serious power quality problem for industrial customers.

A voltage sag is defined as a momentary decrease of the voltage rms value during 0.5 - 30 cycles. A sag can cause serious problem to sensitive loads that use voltage-sensitive components such as adjustable speed drives, process control equipment, and computers [2]. Some applications such as automated manufacturing processes are more sensitive to voltage sags and swells than other applications. For sensitive loads, even a voltage sag of short duration can cause serious problems. In order to increase the reliability of a power distribution system, many methods of solving power quality problems, especially voltage sag, have been suggested. Many voltage sag mitigation schemes are based on inverter systems consisting of energy storage and power switches. Existing methods of voltage sag mitigation using gate turn-off switches for PWM need at least two switches per phase. Other methods use a direct AC-AC converter topology. In addition to requiring at least two switches per phase, they require energy storing reactive components. In an effort to achieve the advantages of a fast time response, a hysteresis voltage control autotransformer is proposed here. The proposed system has only one power switch per phase with no energy storage. Any power electronic switch for a high voltage application is expensive, and the peripheral circuits such as gate drivers and power supplies are even more expensive than the device itself. The overall cost of power electronics-based equipment is nearly linearly dependent on the overall number of switches in the circuit topology. This paper suggests a scheme with a new control method that uses only one power switch with no energy storage. Since fewer components are required in this scheme, the system becomes more reliable and less expensive. To efficiently mitigate the voltage sag event with the proposed scheme, many subsystem designs are required. In this paper, the analysis and design of the overall system and simulation results are presented.

II. System Configuration

Various AC converter technologies have been proposed for AC output voltage control [1]. The well-known DC to DC conversion technology has been adapted to AC to AC conversion technology. AC converters consist of two solid-state switches per phase and require reactive elements such as a capacitor and an inductor. Since the current in the AC converter flows in both directions, static switches and diodes are serially connected to allow both directions current.

This paper suggests a new compensation scheme with a new control method. In this scheme, only one power switch is used [5]. The scheme configuration is shown in Fig.1. In this configuration an autotransformer is used as a boosting transformer instead of a two-winding transformer. The autotransformer in Fig.2 does not offer electrical isolation between primary side and secondary side but has advantages of high efficiency with small volume. The relationships of the autotransformer voltage and current are expressed in equation (1), where ‘a’ is the turns ratio. In this paper, a transformer with ratio N1 : N2 = 1 : 1 is used to boost up to a 50% voltage sag. In this configuration, the switch current is two times the load current.

This voltage sag supporter works for only a few seconds and remains in the off-state most of its operation time. Since

the switch remains in the off-state for most of the time and must withstand the voltage across it. Therefore, the voltage across the switch becomes an important factor. The voltage across the switch in the off-state is equal to one half of the input voltage.

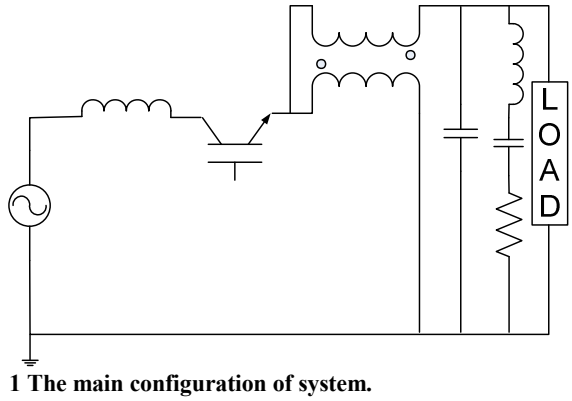


Fig. 1 The main configuration of system.

As the turns ratio equals 1:2, the magnitude of the load current I_H (high voltage side) is the same as that of the primary current I_L (low voltage side). From equation (1), it is clear that $V_H = 2V_L$ and $I_L = 2I_H$.

$$\frac{V_L}{V_H} = a = \frac{I_H}{I_L} \quad , \qquad a = \frac{N_1}{N_1 + N_2} \tag{1}$$

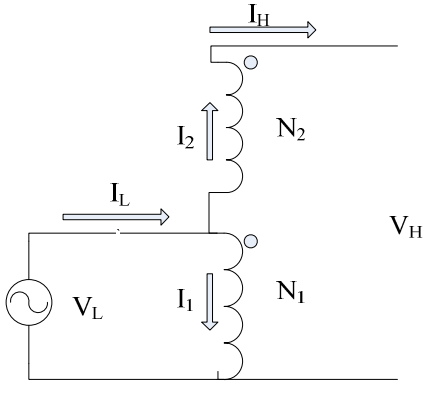


Fig. 2 General structure of autotransformer.

The basic configuration of the proposed Method is presented in Fig.3. The model consists of a single IGBT switch in a bridge Configuration, a thyristor bypass switch, output filters, an autotransformer, and the system controller.
In normal conditions, the IGBT switch is off and the power flows through anti-parallel thyristor as a bypass switch. This bypass switch connects the input power to the load unless the sag condition is present.
When voltage sag occurs, the bypass switch turns off and the switching command turns on the IGBT switch. Utilizing autotransformer, the IGBT improves the output voltage in such a way that the voltage across the load remains constant.
Since the IGBT switch operates only under sag conditions, the efficiency of the system is high. However, switching the power devices always produces noise in the system, which can cause the waveform having harmonic contents. To prevent the switching noise effect of IGBT and thyristor on the load current and voltage waveforms, two filters (a capacitor filter and a notch filter) utilized.

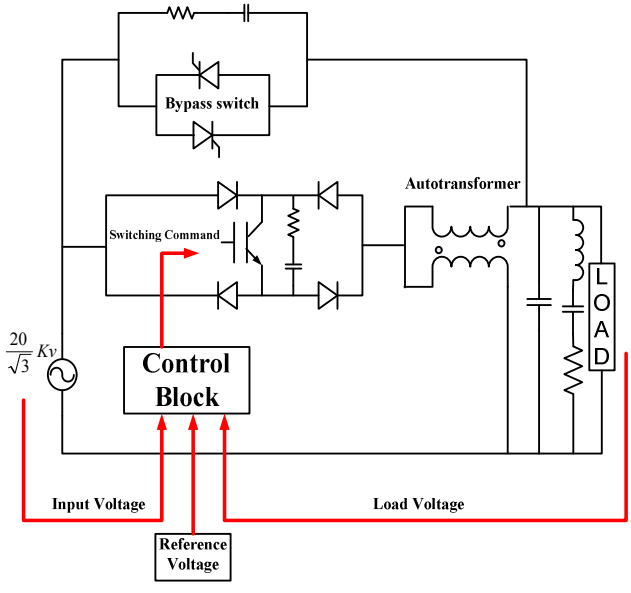


Fig. 3 Overall structure of voltage sag compensator.

When the switches turn off, a high di/dt occurs, which can damage the switches.
To keep the power switches safe, an RC snubber circuit is designed. The snubber circuit of IGBT switch consists of a resistor and a capacitor. The snubber suppresses the peak voltage across the IGBT switch when the IGBT turns off. When the IGBT switch turns off, the current flowing in the IGBT in the on-state instantly diverts to the snubber circuit. The energy stored in the current path is transferred into the snubber capacitor and resistor.

III. System Characteristics

The system can be divided in six parts: the source, the load, the IGBT switch, the thyristor bypass switch, the RC snubber circuit, and the output filters.
In this system, the magnitude of the input three-phase voltage is 20 kv rms. So, the rms value of line to neutral single-phase voltage is 11.5 kv (16.3 kv peak). The load of system is a passive RL load.

A. Output filters

To reduce harmonic components of the output voltage, two filters are used. The first one is a notch filter and the other is a capacitor filter. Usually less than 5% THD (Total Harmonic Distortion) of the voltage is acceptable in power system. To select the filter values, firstly the equivalent circuit is derived as shown in Fig.4. In Fig.4, the total effective inductance L consisting of the source and the leakage inductances of the transformer represents $4L_{source}+2L_{leakage}$.

From Fig.4, it can be observed that the combination of the effective inductance L and the output capacitor filter named C_{filter} form a low-pass filter. As the source and leakage inductance work as a low-pass filer, it seems large source impedance is preferred to reduce the harmonics. The leakage inductance helps to reduce the harmonics, but the source inductance does not. The impedance of the filter is given by equation (2).

$$Z = R + j\left(\omega L - \frac{1}{\omega C}\right) \quad (2)$$

To select the capacitor value of the notch filter, the common design rule is to choose the capacitor kVA about 25%_30% of the total rating kVA. Since the output filter is always energized regardless of the operation mode, it is desirable that the output capacitor has a lower capacitance. The capacitor reactive power (VAR) is given by equation (3).

$$VAR = \frac{V^2}{X} \quad (3)$$

Using equation (3), the total capacitance is obtained as $C = 2\ \mu F$.

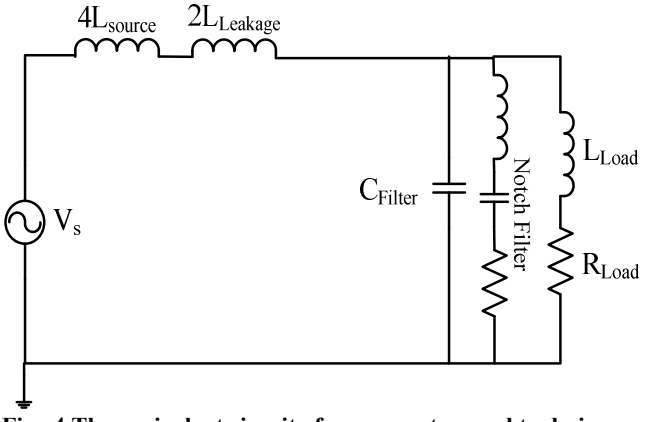


Fig. 4 The equivalent circuit of compensator used to design filter.

IV. Voltage Detection Method

Voltage detection is important because it determines the dynamic performance of the voltage sag compensator. The magnitude of output voltage determines the behavior of hysteresis voltage control method and the starting moment of compensation. Therefore, precise and fast voltage detection is an essential part of the voltage sag supporter. Several voltage detection methods have been documented for use in various voltage compensation schemes. Many approaches use the DQ transformation of the voltage in the synchronous reference frame.

If three-phase parameters such as currents and voltages are balanced, the value of the DQ transformation results in constant DC values. In addition, the resulting DC values make the voltage controller design easier. The DQ transformation uses instantaneous values. Therefore, the detection time is much faster than other methods such as average, rms, and peak detection. However, if voltage sag is unbalanced, a ripple will appear in the DC component. To solve this problem, it is necessary to use a filter. However, using the filter will cause delay in the voltage detection. So, this method is not appropriate for unbalance voltage sag event. Moreover, this method requires the values of all phases. To control and detect the voltage sag, the voltage control scheme used in this paper requires only the peak values of output voltage. Therefore, a simple method called the "peak detection method" is used for detecting voltage [4]. While the DQ transformation needs three-phase information, the peak detection method needs only a single phase value. The peak detection method is implemented as shown in Fig.5, and equation (4) forms the peak detection value as follows.

$$(V_m \sin\theta)^2 + (V_m \cos\theta)^2 = {V_m}^2 \quad (4)$$

The process of measuring the peak value can be explained as follows. The single-phase line-to-neutral voltage is measured, and the cosine value of this voltage is determined using a 90° phase shifter. Assuming a fixed value (50Hz) for the line frequency, the 90°-shifted value can be found. The signal waveforms of The single-phase line-to-neutral voltage and the 90°-shifted value of this voltage are shown in Fig.6.

Both components of voltage are squared and summed to yield V_m^2. Obtaining the square root of V_m^2 results in the peak value of the detected voltage.

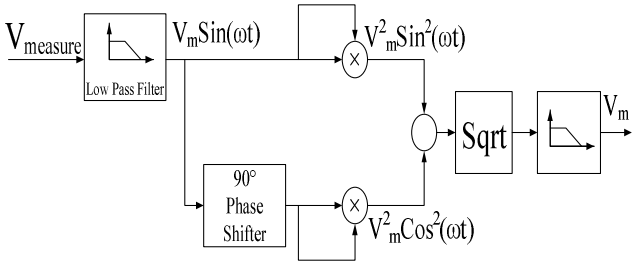


Fig. 5 The block diagram of peak detection method.

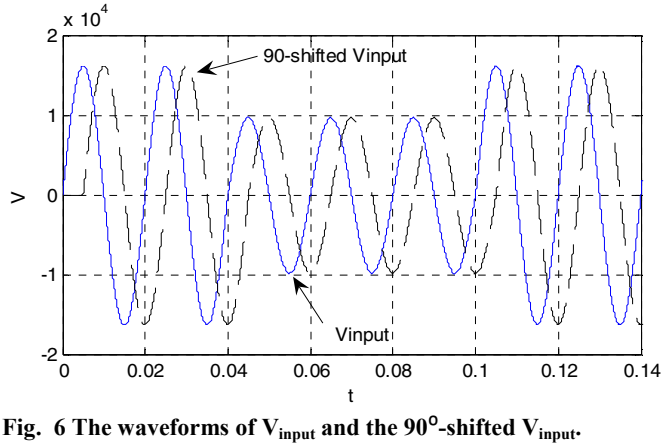


Fig. 6 The waveforms of V_{input} and the 90°-shifted V_{input}.

The output value of peak detection block is compared with the input voltage (V_{input}) in Fig.7. The comparison verifies the ability of method in detecting the peak value of the input signal in the least possible time. The detection time in this method used to be less than a quarter of a cycle.

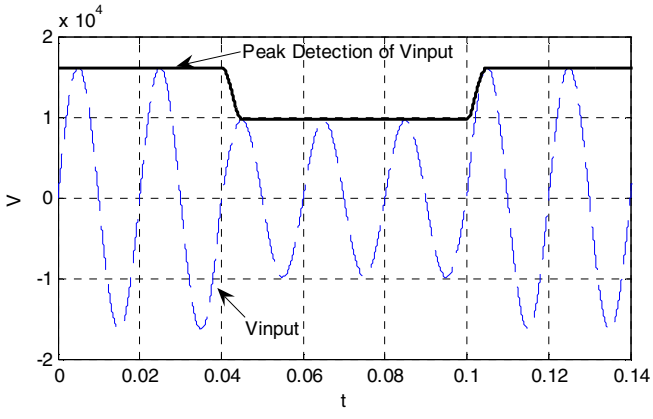


Fig. 7 Peak detection signal and V_{input} versus time.

V. Hysteresis Voltage Control Method

A hysteresis voltage control technique is implemented with a closed loop system where an error signal, e(t), is used to determine the switching states and to control the load voltage [3,6]. e(t) is the difference between the reference voltage, V_{Ref}, and the actual voltage, V_{actual}.

As shown in Fig.8, there are bands above and under the reference voltage. When the error reaches to the upper limit, the voltage gets forced to decrease and when the error reaches to the lower limit, the voltage gets forced to increase.

The input signals of the control block are V_{input}, V_{load} and V_{ref}. When a voltage sag occurs, the magnitude of the input voltage decreases. As mentioned, the hysteresis method is based on the difference between the actual voltage and the reference voltage. On the other hand, this scheme is based on utilizing an autotransformer with 1:2 turn ratio. So, to compensate voltage sag and keep the load voltage constant, it is necessary to implement half of nominal source voltage in sag conditions to the primary of autotransformer.

This will be done by IGBT switching. So, the signal ½ V_{load} will play the role of actual voltage in control block. Therefore, in this method the error signal (e(t)) is defined in equation (5).

$$e(t) = \frac{1}{2}V_{load} - V_{ref} \quad (5)$$

Which V_{ref} is presented in equation (6).

$$V_{ref} = \frac{1}{2}V_{source} \quad (6)$$

The block diagram of the proposed control method is presented in Fig.9.

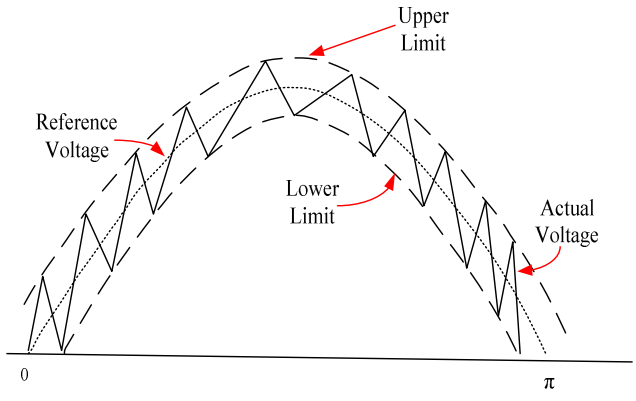


Fig. 8 Hysteresis switching method.

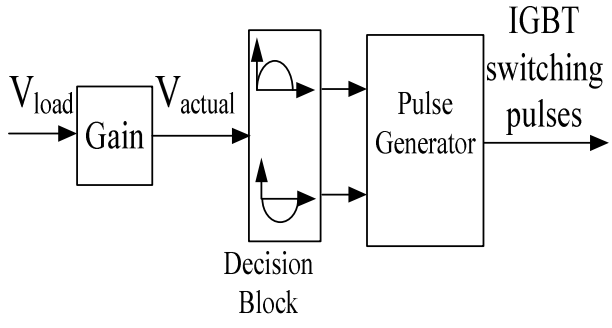


Fig. 9 The block diagram of the proposed control method.

As shown in Fig.9, the actual signal (V_{actual}) acts as the input of decision block. If V_{actual} is positive, the upper path of decision block produces IGBT switching pulses, and if V_{actual} is negative, the lower path of decision block produces IGBT switching pulses.

The control method described above is simulated in MATLAB. The simulink block of hysteresis voltage controller is shown in Fig.10.

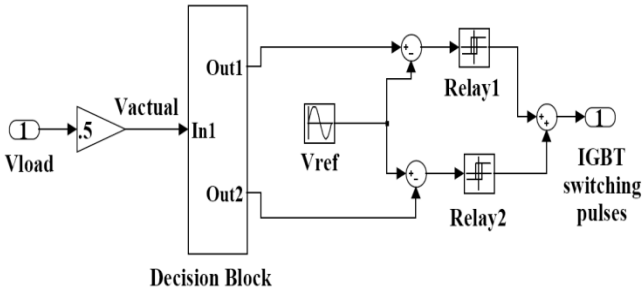


Fig. 10 The simulink block of hysteresis voltage controller.

The switchig pulses produced by Relay1 and Relay2 are shown in Fig.11 and Fig.12. Moreover, the IGBT switching pulses, which is the sum of switching pulses produced by Relay1 and Relay2, is shown in Fig.13.

The output signals of the upper and lower paths of decision block is shown in Fig.14 and Fig.15. In this figures, solid lines are the outputs of decision block, and dashed line is the reference voltage (V_{ref}).

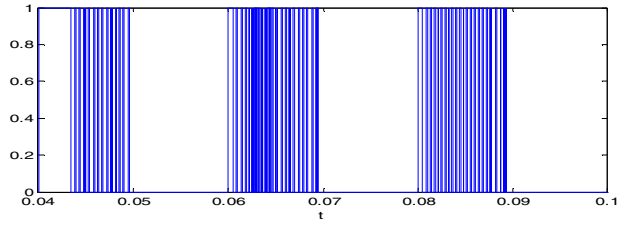


Fig. 11 The switching pulses produced by Relay1.

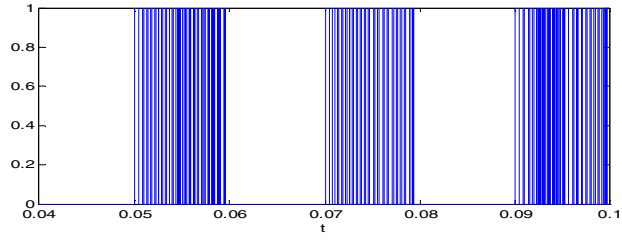


Fig. 12 The switching pulses produced by Relay2.

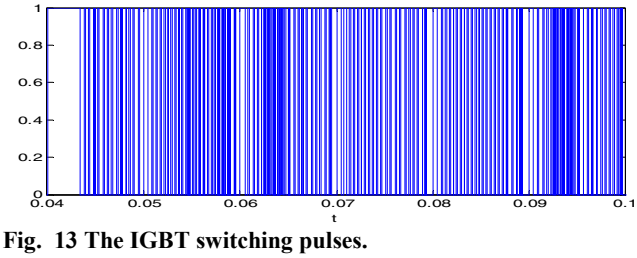


Fig. 13 The IGBT switching pulses.

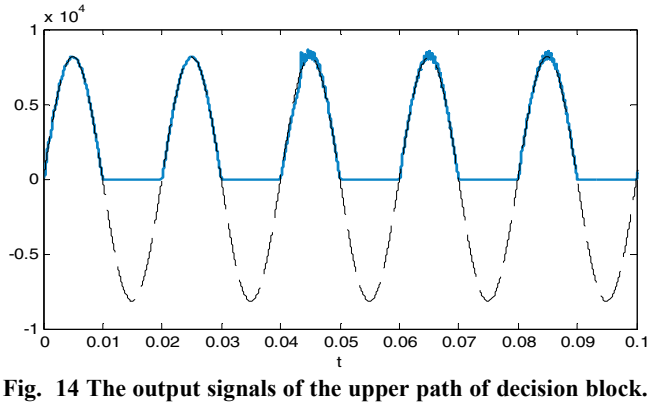


Fig. 14 The output signals of the upper path of decision block. Dashed line is V_{ref} and solid line is V_{actual}.

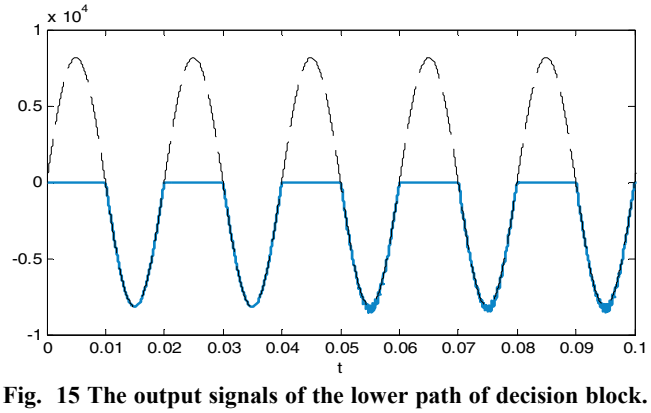


Fig. 15 The output signals of the lower path of decision block. Dashed line is V_{ref} and solid line is V_{actual}.

VI. Simulation Results

To verify the ability of proposed method in compensating voltage sag, different simulations have been carried out by MATLAB/Simulink software. In these simulations, the load of system is an R-L load. Voltage sag events with depth of 10% to 50% have been simulated. In normal conditions, the IGBT switch is off and the control system transfers the power to the load through thyristor (bypass switch). When voltage sag occurs, the control system of compensator makes the thyristor turn off and commands the IGBT to turn on, and power flows through IGBT and autotransformer. The input signals of control block are V_{input}, V_{load} and V_{load}. Considering the load voltage fed back to the control block, the hysteresis voltage controller switches the IGBT. Switching the IGBT will apply half of the desirable voltage to the primary of autotransformer. Considering that the turn ratio of autotransformer is 1:2, the output of autotransformer will be equal to the nominal voltage ($20/\sqrt{3}$ kv rms). Fig.14 and Fig.15 show the results of simulations for 20% voltage sag and 40% voltage sag.

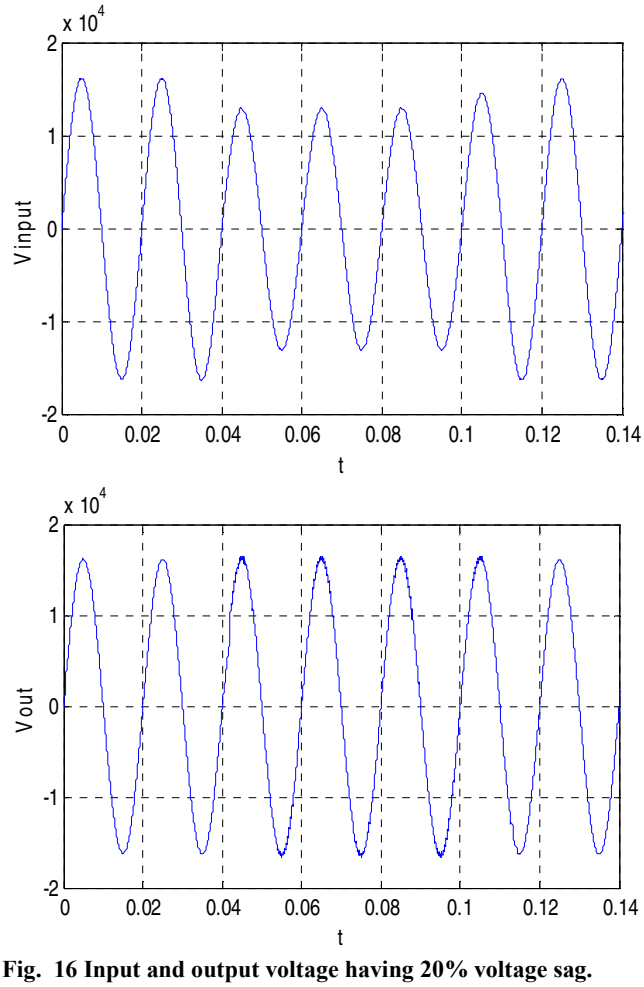


Fig. 16 Input and output voltage having 20% voltage sag.

Fig.16 and Fig.17 obviously confirm the ability of the proposed method to mitigate voltage sag events with different depth and different durations.

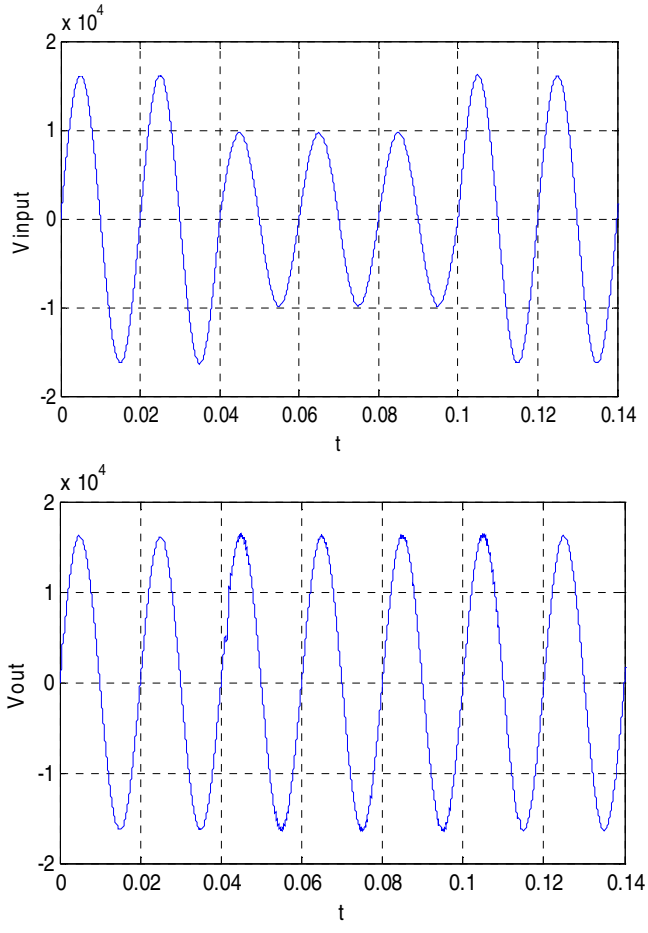


Fig. 17 Input and output voltage having 40% voltage sag.

VII. Conclusion

A new voltage sag compensator based on an autotransformer and an IGBT switched by hysteresis voltage control method has been presented in this paper. Hysteresis voltage control method suggested in this paper is a method in which the real voltage of system compares with a reference voltage through a hysteresis band, and hysteresis band produces switching signals of IGBT.

Different voltage sag events with depths between 10% to 50% and different duration have been simulated by MATLAB/Simulink software. The results of simulations verify the ability of proposed method to mitigate voltage sag events.

References

[1] Bimal K.Bose, *Modern Power Electronics and AC Drives*, Prentice-Hall Inc., 2002.

[2] M. H. J. Bollen, *Understanding power quality problems: voltage sags and interruptions*, lEEE Press, New York, 2000.

[3] I. Fatouh " A powerful and Efficient hysteresis PWM controlled inverter", EPE Journal, vol. 4, no.4, December 1994.

[4] H. Y. Chu, H. L. Jou, and C. L. Huang, "Transient response of a peak voltage detector for sinusoidal signals," IEEE Trans. Ind. Electron., vol. 39, no. 1, pp. 74–79, Feb. 1992.

[5] D. M. Lee, T. G. Habetler, R. G. Harley, J. R. Rostron, and T. L. Keister, "A Voltage Sag Supporter Utilizing a PWM-Switched Autotransformer," in Proc. IEEE 2004, Aachen, Germany,pp. 957–962, Nov. 2–6, 2004.

[6] K. M. Rahman, M. R. Khan, M. A. Choudhury, "Variable band hysteresis current controllers for PWM voltage source inverters," IEEE Trans. on Power Electronics , vol. 12, no. 6, pp. 964-970, November1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Analysis of X-Cut Lithium Niobate Electrooptic Modulators with Backside Slots

M. Khaled Hassan[1] *and M. Shah Alam*[2]

[1]Department of Electrical and Computer Engineering
Presidency University, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology (BUET)
Dhaka 1000, Bangladesh
E-mail: shalam@eee.buet.ac.bd

Abstract - A detailed microwave analysis of X-cut $LiNbO_3$ (LN) electrooptical (EO) modulators is presented in this paper by using the finite element method (FEM). The two-step back-slot structure considered here satisfies the velocity matching condition and eliminates the necessity of buffer layer of silicon dioxide (SiO_2) which is necessary for the ridge type Z-cut $LiNbO_3$ modulators. The optical 3-dB bandwidth and the effect of microwave loss on the bandwidth are shown in this paper. It has been shown that, an optical 3-dB bandwidth of more than 140 GHz can be achieved with X-cut $LiNbO_3$ substrate, when back side slots are incorporated in the structure.

I. Introduction

The demand for high capacity long-haul telecommunication systems has been increasing at a steady rate, and is expected to accelerate in the next decade. So, a broad-band optical modulator with a travelling wave (TW) electrode is essential for future optical communication systems [1]. External optical modulators made from the titanium diffused lithium niobate (Ti:$LiNbO_3$) in a Mach-Zehnder Interferometer (MZI) structure with a coplanar waveguide (CPW) type travelling wave electrode are most promising owing to their large modulation bandwidth, lower cost, large electrooptic coefficients, low bias voltage, well defined chirp characteristics, better coupling efficiency, and higher extinction ratio, which have warranted their wide applications in high speed optical networks and microwave photonic systems [2]-[10].

For designing an efficient system, the accurate bandwidth characteristics of the modulator should be determined and hence their analysis is very important. In this regard, microwave characterization to know the properties of travelling wave electrodes is a must. The operation bandwidth of the modulator is restricted mainly by the microwave characteristics of the modulator, especially microwave losses (dielectric and electrode losses) and by the mismatch in velocity between the modulating microwave signal and the modulated optical waves [2]-[5], [11].

It can be seen from the previous works that several methods have been used so far for the analysis of guided-wave electrooptic modulators using various crystal materials or semiconductors [1]-[5], [9]-[11]. The FEM in general, is very well-suited to the problem within limited regions defined by closed boundaries. It is a powerful and efficient tool for most general waveguiding problems and has been widely used for modeling and optimization of the TW electrode [1], [5]-[8].

In this work, the analysis of traveling-wave electrooptic modulator on X-cut $LiNbO_3$ substrate is carried out by using the finite element method based on a quasi-TEM analysis. By discretizing the modulator cross section with many triangular elements in adaptive meshing and solving the Laplace's equation in the *partial differential equation toolbox* (*pdetool*) environment of MATLAB, the potential distribution over the cross section is obtained. Then the capacitance per unit length and hence the microwave effective index (N_m), the characteristic impedance (Z_c), and the losses are calculated using the modulating electric field. The optical frequency response is also calculated and as a consequence, the bandwidth is estimated for velocity matching condition.

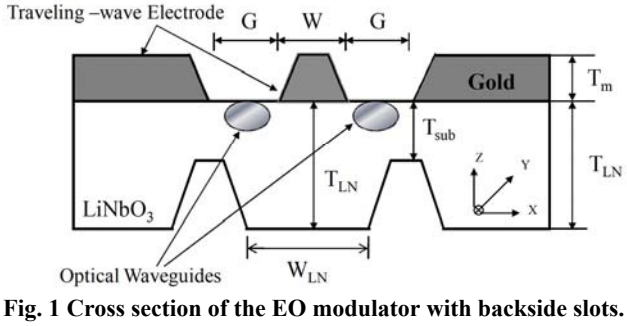


Fig. 1 Cross section of the EO modulator with backside slots.

II. Device Architecture & MZI

The cross section of the external EO modulator based on the Mach-Zehnder configuration using X-cut LN substrate is shown in Fig. 1. The two optical waveguides are placed underneath the electrode gaps between the central conductor and the grounded electrodes. The conventional Z-cut LN modulators must have a buffer layer of silicon dioxide (SiO_2) in order to satisfy the velocity matching condition and eliminate the optical insertion loss, but the buffer layer causes the problem of dc-drift phenomenon [6]-[7]. By introducing the two step backside slots, the necessity of SiO_2 buffer layer has been eliminated as this structure allows the microwave to

efficiently leak into the air around not only the CPW electrode but also the back slots and the velocity matching condition between lightwave and microwave can be achieved easily [6]-[7]. The T_{sub} as shown in Fig. 1 must be controlled within (10±2) μm to satisfy the velocity matching condition [6]. Fig. 2 shows the schematic diagram of a MZI modulator. An MZI modulator is composed of two single mode waveguides, usually one arm is under no electric field, and the other arm is under an electric field. In this device one optical wave is first split into two waves and then synthesized into one wave again to induce optical interference. A refractive index change results when the electrical field is applied to the $LiNbO_3$ modulator because of the electrooptic effect. The electric field is often applied in the transverse direction as because much higher magnitude of the longitudinal electric field is usually required to achieve the switching operation [10].

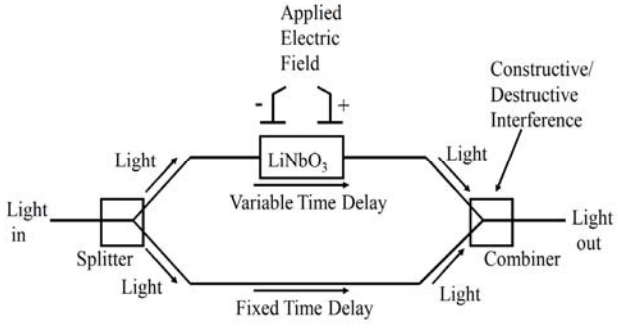


Fig. 2 Schematic diagram of $LiNbO_3$ Mach-Zehnder optical modulators.

III. Theory

Applied microwave signal (modulating signal) on CPW electrode induces an electric field in the $LiNbO_3$ substrate. The static electric field is obtained by solving the Laplace equation as follows [1], [12]:

$$\vec{\nabla}.(\varepsilon\vec{E}) = 0 \quad , \tag{1}$$

where ε is the dielectric constant of the material. The induced electric field has scalar potential Φ over the cross section. The linear triangular elements are used for discretizing the waveguide cross section neglecting the electrode region and Dirichlet type boundary conditions are imposed on the shielding wall and on the conductor surface. Once the potential is known, both the horizontal and vertical modulating electric field can be calculated.

A. Capacitance per unit length:

To calculate the capacitance per unit length of the CPW, C, the energy expression is employed as [13]

$$C = \frac{2W}{V_o^{\,2}} = \frac{1}{V_o^{\,2}} \iint_S [\varepsilon] \left|\nabla\Phi\right|^2 dS \quad , \tag{2}$$

where $[\varepsilon]$ is the permittivity tensor for anisotropic materials, V_0 is the potential difference between the electrodes, W is the total energy per unit length, S is the cross sectional domain of the structure. By replacing the dielectric materials by free-space (air), C_0, the capacitance of free space can be obtained using (2).

B. Microwave effective index and characteristic impedance:

From C and C_0, the microwave effective index (N_m) and the characteristic impedance (Z_c) of the electrode can be calculated [1]-[5], [13].

$$N_m = \sqrt{\frac{C}{C_0}} \tag{3}$$

$$Z_C = \frac{1}{c\sqrt{CC_0}} \quad , \tag{4}$$

where, $c=3\text{x}10^8$ m/s, is the velocity of light in free space.

C. Microwave loss coefficients:

The microwave loss that is due to the imperfect conductors can be calculated using incremental inductance formula [1], [5]

$$\alpha_C = \frac{R_S}{2Z_0 Z_C} . \frac{\partial Z_{C0}}{\partial n} \quad , \tag{5}$$

where Z_{C0} is the free space characteristic impedance of the electrode and $\partial Z_{C0} / \partial n$ denotes the derivative of Z_{C0} with respect to the incremental recession of the electrode surfaces. R_S is the surface resistance and $Z_0=(\mu_0/\varepsilon_0)^{1/2}$ is the impedance of free space. The microwave loss that is due to the lossy dielectrics can be calculated with the perturbation theory as [5], [13]

$$\alpha_d = \frac{P_d}{2P_o} \quad , \tag{6}$$

where P_d and P_o are the time averaged powers dissipated in the dielectrics and the total power, respectively. P_d can be calculated using the formula [13]

$$P_d = \omega\varepsilon \tan\delta \int_{S_{diel}} \left|\overrightarrow{E_0}\right|^2 dS \tag{7}$$

and the average propagating power P_o can be calculated using pointing theorem as

$$P_0 = \mathrm{Re}\int_S (\overrightarrow{E_0} \times \overrightarrow{H_0^*}).\hat{z}dS \quad , \tag{8}$$

where *tan* δ represents the loss tangent of the dielectric, $\omega = 2\pi f$ is the angular frequency, and S_{diel} is the cross-section covered by the dielectric material.

D. Optical Response:

The optical frequency response $m(f)$ of a travelling-wave EO modulator in presence of microwave attenuation can be calculated using [1]

$$m(f) = \left[\frac{1 - 2e^{-\alpha L}\cos 2u + e^{-2\alpha L}}{(\alpha L)^2 + (2u)^2}\right]^{1/2} \quad , \tag{9}$$

where

$$u = \frac{\pi f L(N_m - N_0)}{c} \text{ and } \alpha = \alpha_c \sqrt{f} + \alpha_d f ,$$

when microwave loss includes both the conductor loss and dielectric loss. Here, N_0 is the effective index of the optical wave, L is the length of the electrode in active region in cm, the frequency f is in GHz when α_c and α_d are normalized at 1 GHz. The optical 3-dB modulation bandwidth Δf can then be obtained from optical response such that $20\log_{10}[m(\Delta f)] = -3$ dB. However, when only conductor loss is considered, for perfect velocity matching condition, the bandwidth Δf can be estimated by [4], [5]

$$\Delta f = \left(\frac{6.84}{\alpha_C L}\right)^{1/2} . \qquad (10)$$

IV. Results and Discussion

To analyze the structure here, first, we calculate the microwave characteristics of CPW. In the microwave frequency range, the relative permittivities of the X-cut $LiNbO_3$ substrate are 28 and 43 in the perpendicular and parallel directions to the substrate surface, respectively, and the conductivity of the gold electrode is 4.1×10^7 S/m. In this work, the loss tangent value for $LiNbO_3$ was taken as 0.004. The thickness of electrodes T_{LN} is fixed at 15 μm and the width of the central electrode is set to 30 μm to decrease the electrode conductor loss. The metallization part of hot and cold electrodes of CPW are neglected while meshing the cross sectional domain. Fig. 3 shows the potential distribution, where the potential values are the finite element solutions of Laplace's equation. It can be seen that the potential field surrounds the hot conductor and is maximum on it. In Fig. 4, the electric field distribution is shown. The arrows show the relative amplitudes of the field strength.

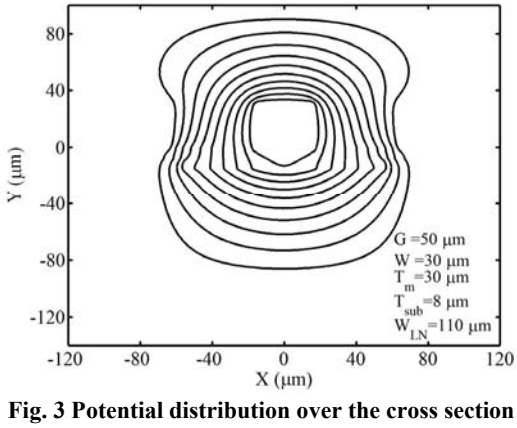


Fig. 3 Potential distribution over the cross section of back-slot EO modulator.

In the design of the optical modulator, the two key microwave parameters N_m and Z_c have significant impact in determining the bandwidth of the modulator. Fig. 5 and 6 show the variations of the N_m and Z_c for the EO modulator with the substrate thickness. It is seen that both the results agree very well with the available results of the experiment. The effective index increases with the increase in substrate thickness. On the other hand, impedance decreases with the increase in substrate thickness. By controlling T_{sub}, the modulation condition of N_m=2.2±0.05 can be achieved [6].

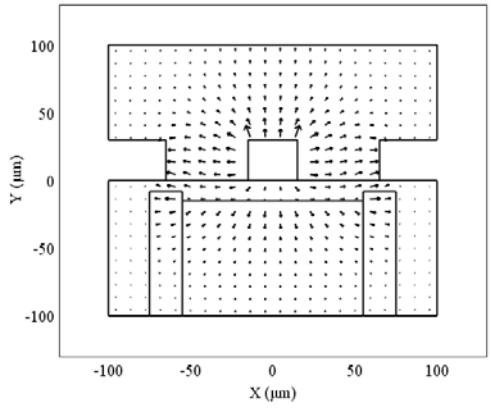


Fig. 4 Arrow plot of the Electric field over the cross section of back-slot EO modulator.

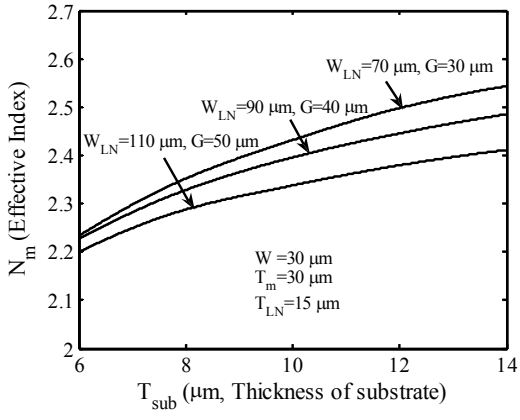


Fig. 5 Microwave effective versus substrate thickness.

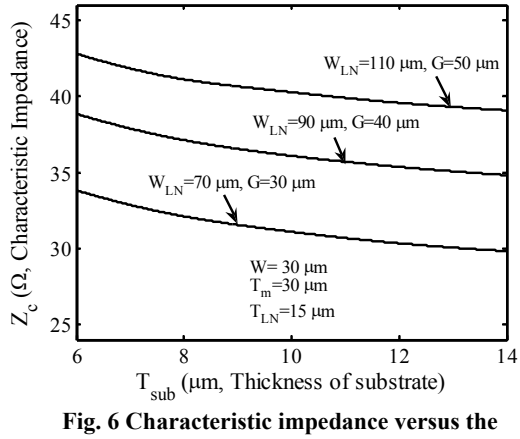


Fig. 6 Characteristic impedance versus the substrate thickness.

Microwave losses also play very important role in the determination of the maximum bandwidth, and, for high-speed modulators when the phase velocities are matched; the bandwidth primarily depends on the conductor loss, α_c. However, at higher operating frequencies, the dielectric loss cannot be ignored. The Fig. 7 shows the dependencies of the conductor loss, dielectric loss, and the total microwave losses on the operating frequencies. Since, the dielectric loss is proportional to the operating

frequency *f*, compared to the conductor loss, which is proportional to the $\sqrt{f}$, the dielectric loss cannot be neglected for the ultrahigh-speed modulators. It can be noted from the figure that at 100 GHz, the dielectric loss could be 25 % of the total loss, and this loss must be taken under consideration in the calculation of the modulator bandwidth.

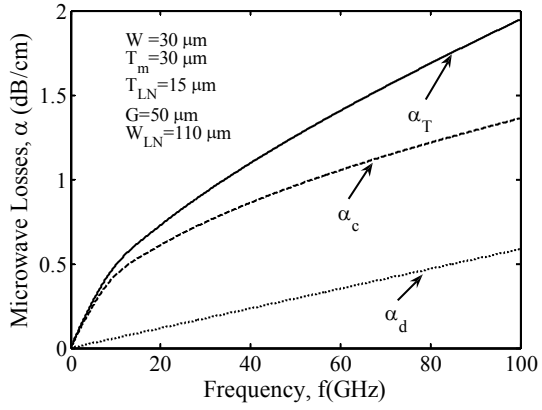


Fig. 7 Variations of conductor loss α_c, dielectric loss α_d, and the total loss α_T with the operating frequency f.

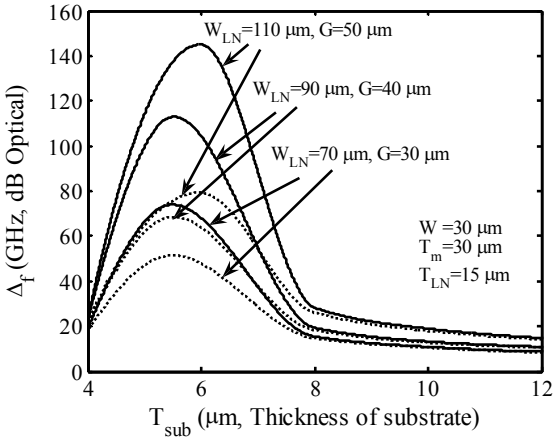


Fig. 8 Variation of the 3-dB optical bandwidth Δ*f* with the substrate thickness for different electrode gaps.

Next, Fig. 8 shows the bandwidth versus substrate thickness. The solid lines shows the curves when only conductor loss is considered. The dashed lines show the bandwidth when both conductor loss and the dielectric loss are taken into account. In this case, the values of N_0 and L are taken as 2.2 and 4 cm, respectively. The highest value as much as 140 GHz can be achieved with this modulator, when the dielectric material is lossless. But the maximum attainable bandwidth depends on the microwave loss. An assumption of lossless dielectric, however, may lead to an erroneous over estimation of bandwidth as because in practical cases the dielectric losses can not be simply neglected, especially when the operating frequency is higher.

V. Conclusion

The microwave analysis of X-cut $LiNbO_3$ optical modulator with back-slots is presented here. The microwave effective index, the characteristic impedance, the losses of CPW, and the effects of various waveguide parameters on microwave properties are investigated thoroughly. The optical 3-dB bandwidth is also determined, which takes into account the microwave losses. It is seen that about 140GHz bandwidth can be achieved with lossless dielectric material when a perfect velocity matching condition is achieved. But it is seen that the highest value of the bandwidth is significantly affected by the dielectric loss.

References

[1] M. Koshiba, Y. Tsuji, and M. Nishio, "Finite-element modeling, of broad-band traveling-wave optical modulators," *IEEE Trans. Microwave Theory Tech.* Vol. 47, no. 9, pp. 1627-1633, 1999.

[2] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A Review of Lithium Niobate Modulators for Fiber-Optic Communication Systems," *IEEE J. Selected Topics in Quantum Electron.*, vol. 6, no. 1, pp. 69-82, Jan./Feb. 2000.

[3] N. Dagli, "Wide-Bandwidth Lasers and Modulators for RF Photonics," *IEEE Trans. on Microwave Theory and Tech*., vol. 47, no. 7, pp. 1151-1171, July 1999.

[4] M. Minakata, "Recent progress of 40 GHz high-speed $LiNbO_3$ optical modulator," *Proceedings of SPIE*, pp. 21-24, Denver, Colorado, USA, Aug. 2001.

[5] S. Haxha, B.M.A. Rahman, and K.T.V. Grattan, "Bandwidth estimation for ultra-high-speed lithium niobate modulators," *Applied Optics*, vol. 42, no. 15, pp. 2674-2682, May 2003.

[6] J. Kondo, A. Kondo, K. Aoki, T. Mori, Y. Mizuno, S. Takatsuji, Y. Kozuka, O. Mitomi, and M. Minakata, "40-Gb/s X-Cut $LiNbO_3$ Optical Modulator With Two-Step Back-Slot Structure" *J. Lightwave Technol*., vol. 20, no. 12, pp. 2110-2114, Dec. 2002.

[7] J. Kondo *et. al*., "High-speed and low-driving voltage X-cut $LiNbO_3$ optical modulator with two step backside slot," *Electron. Lett*., vol. 38, no. 10, pp. 472—473, May 2002.

[8] Satoshi Oikawa , Futoshi Yamamoto, Junichiro Ichikawa , Sunao Kurimura and Kenji Kitamura, "Zero-Chirp Broadband Z-Cut Ti: $LiNbO_3$ Optical Modulator Using Polarization Reversal and Branch Electrode", *IEEE J. Lightwave Technol.*, Vol. 23, no. 9, pp. 2756-2760, Sep. 2005.

[9] Yongqiang Shi, "Micromachined Wide-Band Lithium-Niobate Electrooptic Modulators," *IEEE Trans. on Microwave Theory and Tech*., vol. 54, no. 2, pp. 810—815, Feb. 2006.

[10] D.-Yuan Chen and J. D. Phillips, "Analysis and Design Optimization of Electrooptic Interferometric Modulators for Microphotonics Applications," *J. Lightwave Technol*., vol. 24, no. 6, pp. 2340-2346, June 2006.

[11] T. Kitazawa, D. Polifko, and H. Ogawa, "Analysis of CPW for $LiNbO_3$ optical modulator by extended spectral-domain approach, " Microwave *and Guided Wave Lett*., vol. 2, pp. 313–315, 1992.

[12] Y.-K. Wu, W.-S Wang, "Design and Fabrication of Sidewalls-Extended Configuration for Ridged Lithium Niobate Electrooptical Modulator " *J. Lightwave Technol.*, vol. 26, no. 2, pp. 286-290, Jan. 2008.

[13] Z. Pantic and R. Mittra, "Quasi-TEM Analysis of Microwave Transmission Lines by the Finite-Element Method," *IEEE Trans. on Microwave Theory and Technol*., vol. MTT-34, no. 11, pp. 1096-1103, Nov. 1986.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Very Deep Nanoscale Domain Inversion in $LiNbO_3$ for High-Power and High-Efficiency SHG Devices

M. S. Islam[1] and Makoto Minakata[2]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[2]Optoelectronics Laboratory, Research Institute of Electronics, Shizuoka University, 3-5-1 Johoku, Hamamatsu 432-8011, Japan

E-mail: islams@eee.buet.ac.bd

Abstract - Very deep nanoscale domain inversion in lithium niobate ($LiNbO_3$) substrate have been realized utilizing the proposed circular form full cover electrode (CF-FCE) method. Initially, we highlighted the theory of QPM-SHG and the method of domain inversion in $LiNbO_3$ (hereafter referred as LN). We also analyzed the advantages and drawbacks of various electrodes utilized for periodic domain inversion in LN. Theoretical calculation of electric field distribution for conventional FCE shows that this method is not suitable for fine domain inversion patterns. We proposed the CF-FCE method for nanoscale domain inversion and the electric field distribution was calculated for this method. The calculated result shows that the CF-FCE is better than that of the conventional FCE for fine domain inversion patterns. Using the proposed CF-FCE, we successfully fabricated 2 μm periodic nanoscale domain patterns in a 500-μm-thick congruent LN (C-LN) crystal. We obtained very deep nanoscale domain inversion using this technique. Such a domain inversion technology is very important for next generation high-power and high-efficiency second harmonic generation (SHG) devices.

I. Introduction

Recently, in the field of opto-electronics, the coherent blue light source is demanded for the application of high density optical disc memory. One of the fabrication methods for the blue light source is second harmonic generation (SHG) device [1,2] using the single crystal ferroelectric materials, such as $LiNbO_3$ and $LiTaO_3$ (hereafter referred as LT). The LN is widely used because of its excellent acousto-optical, electro-optical and nonlinear optical properties. The wavelength of incident infrared light is halved to generate blue light in SHG device. In order to get a high converted power of SHG, the phase matching between the fundamental wave and SHG must be satisfied. There are many kinds of phase matching, quasi phase matching (QPM) is widely studied since it has the highest efficiency [3,4]. Quasi-phase matched second harmonic generation (QPM-SHG) devices using LN and LT were studied in the last couple of years [5-9]. The key technology in fabricating a high-power and high-efficiency QPM-SHG device is the formation of a very-deep fundamental periodic domain inversion. Using existing techniques, fabrication of smaller domain inversion structures (<3μm period) are difficult. Therefore, alternative methods should be investigated and developed. To date, there have been many reports of domain inversion formation [5-9]. The fabrication of a comb electrode by the lift-off method is the most useful one. However, the lift-off electrode fabrication for domain inversion becomes more difficult as the period is decreased to less than 2 μm. Shur, *et al*. demonstrated promising possibilities of domain engineering based on backswitched poling in C-LN with submicron period [10]. Restoin, *et al*. fabricated sub-μm scale 1.6 μm periodic inverted gratings on C-LN for photonic band-gap devices by electron beam bombardment [11]. But they did not report the domain inversion depths of such gratings.

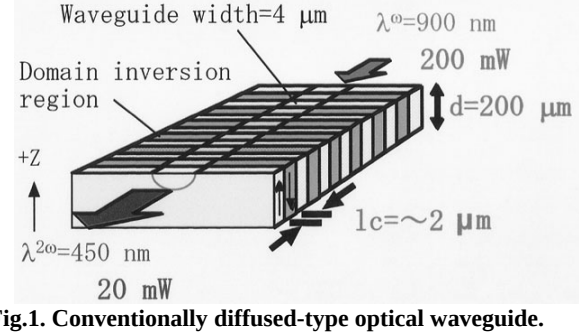


Fig.1. Conventionally diffused-type optical waveguide.

Using the conventionally diffused-type optical waveguide device (Fig. 1), output power of 20 mW and output frequency of 450 nm are obtained. The arrows in the crystal indication the directions of polarizations, l_c is the coherent length, the direction of polarization is periodically inverted by a period of $2l_c$, λ^{ω} and $\lambda^{2\omega}$ are the wavelengths of the fundamental wave and SHG respectively. When the fundamental wave propagates along the waveguide, part of it is converted into SHG effectively. In order to realize the rewritable DVD (digital versatile disc), more output power (30 mW) is required to the blue light source. Moreover, large output power bulk type blue QPM-SHG devices beyond 10W require very-

deep periodic domain inversion patterns. The next generation DBR (distributed Bragg reflector) waveguide type QPM-SHG device demands nanometer range periodic domain patterns (Fig. 2). Therefore, new domain inversion control technology for making the conventional aspect ratio (ratio of domain depth to domain inversion width) from 100 to 1000 has become necessary for high-power and high-efficiency QPM-SHG devices [12]. In this paper, initially we highlighted the theory of QPM-SHG and the method of domain inversion in LN. We also analyzed the advantages and drawbacks of various electrodes utilized for periodic domain inversion in LN. Finally we proposed the CF-FCE for nanoscale domain inversion in LN. Using the proposed CF-FCE, we fabricated very-deep nanoscale domain inversion in congruent LN (C-LN).

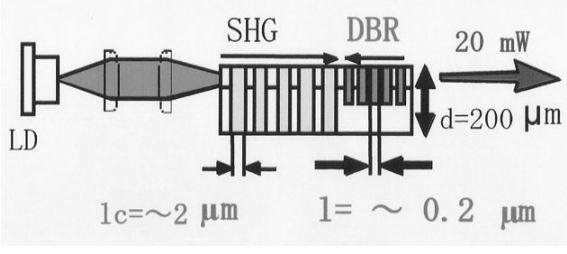


Fig.2. DBR waveguide-type QPM-SHG device.

II. The Theory of QPM-SHG and Method of Domain Inversion

A. The Theory of QPM-SHG

The conversion efficiency of QPM-SHG, η_{SHG}, is defined as the ratio of $P^{2\omega}$ to P^{ω} and is expressed by

$$\eta_{SHG} = \frac{P^{2\omega}}{P^{\omega}} = 2\left(\frac{\mu_0}{\varepsilon_0}\right)^{\frac{3}{2}} \frac{\omega^2 d^2 L^2}{n^3} \frac{P^{\omega}}{A} \frac{\sin^2(\Delta kL/2)}{(\Delta KL/2)^2} \quad (1)$$

where ε_0 and μ_0 are the electric and magnetic permeabilities of vacuum, respectively; ω is the angular frequency of the fundamental wave; d is the effective nonlinear optical coefficient; n is the refractive index; L is the interaction length; and A is the cross sectional area of the waveguide, P^{ω} and $P^{2\omega}$ are the power of the fundamental wave and the SHG respectively. For example, the maximum nonlinear optical coefficient d_{33} are 34 pm/V and 18 pm/V for LN and LT, respectively. The value of n is 2.2 for both materials. Therefore, LN is the appropriate material for the device. The difference in the propagation constant between the fundamental (input) wave k^{ω} and generated (output) wave $k^{2\omega}$, Δk is expressed as

$$\Delta k = k^{2\omega} - k^{\omega}. \quad (2)$$

The term $\frac{\sin^2(\Delta kL/2)}{(\Delta KL/2)^2}$ in the square of the sinc function. In the QPM method, the value of sinc function can be treated as unity, a maximum value, since the value of Δk is regarded as zero. The coherent length, l_c, is expressed as

$$l_c = \frac{\lambda^{\omega}}{4\left(n^{2\omega} - n^{\omega}\right)}, \quad (3)$$

where n^{ω} and $n^{2\omega}$ are the refraction index of the fundamental and SHG wave, respectively. In the case of LN, the coherent length ranges from 1.4 μm to 2.0 μm for generating light of 410 nm to 450 nm [4].

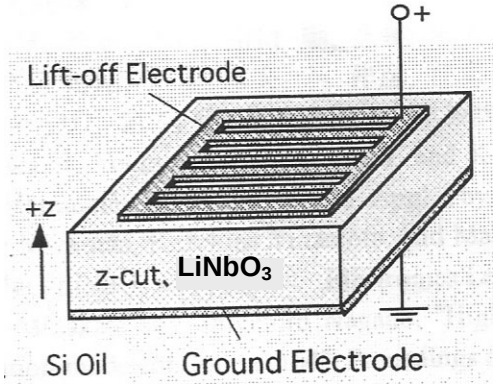


Fig.3. Lift-off electrode for domain inversion.

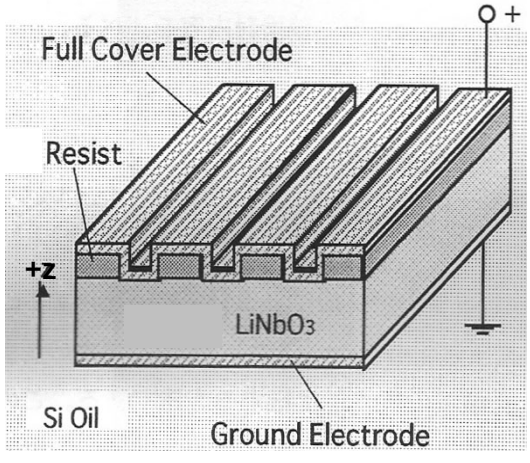


Fig.4. Rectangular full-cover electrode for domain inversion.

B. The Method of Domain Inversion in LN

The LN crystal has 180^0 domain inversion where only lithium atom changes its position. Other atoms remain their original positions. There are many reports of domain inversion formation [1,5-9,13]. Among others, domain inversion by using applied electric field can provide a structure with straight domain boundary perpendicular to the crystal surface and high uniformity in the thickness direction [5,13]. In this method, high voltage is applied between the patterned electrode on the +z-face and the uniform electrode on the –z-face (Fig.3). As a result, lithium atom changes its position and polarization/domain inversion occurs. This method has been examined extensively for the implementation of high-efficiency SHG device. One of the most important processes of this method is to fabricate the periodic electrode. The periodic electrode is fabricated using lift-off process. The lift-off process becomes more difficult as the periodicity is decreased. Moreover, the lift-off electrode results in insufficient poling current. To overcome the problems of lift-off electrode, Nagano *et al.* proposed [14] the rectangular full-cover electrode (FCE) (hereafter referred as conventional FCE) method (Fig.4) to fabricate small size periodic electrode. They found that the conventional

FCE method is advantageous compared to the lift-off method for small periodic domain inversion.

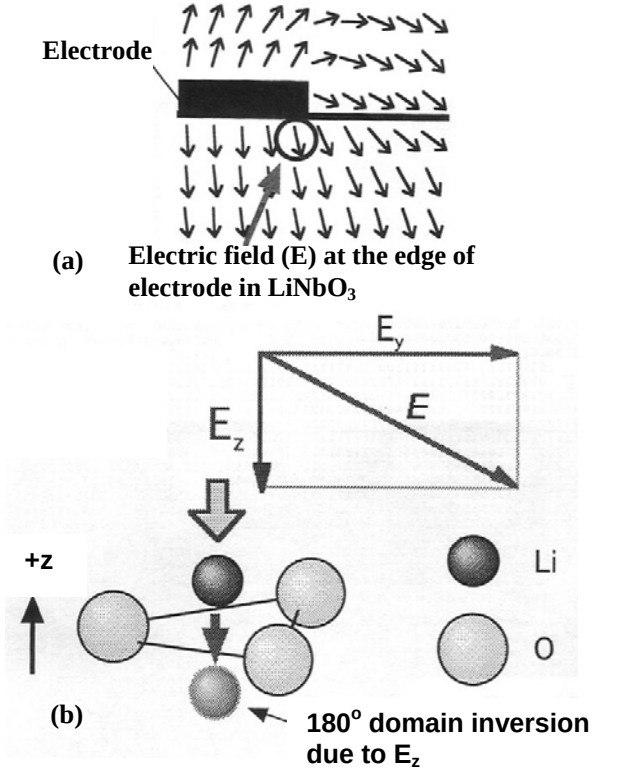


Fig. 5. (a) Electric field distribution and (b) 180^0 domain inversion in LN for rectangular electrode.

The electric field distribution in the crystal for both the lift-off electrode and the conventional FCE are calculated using successive over relaxation (SOR) method [15]. Figure 5 shows y-component and the z-component of the electric field distribution in the LN crystal that is calculated by the SOR method. The arrow shows the direction of electric field at each point. The scalar value $\left(|E| = \sqrt{E_y^2 + E_z^2}\right)$ of the electric field vector E and the direction can be calculated by the SOR method. A shown in Fig.5, the scalar value and the direction at each point on the grid near the edge of electrode are precipitously changing. On the other hand, it is found that the electric field distribution is almost parallel to the z-component and comparatively uniform near the central part of the electrode. The domain inversion is induced by the z-component of the electric field, because the LN crystal have a 180^0 domain structure (Fig.5(b)). The domain broadening is induced by the y-component of electric field particularly near the edge of the electrode. Since the two maximum electric fields exist directly under electrode extreme from electric field calculation in the conventional FCE (Fig.6 (a)), it generates nucleation near electrode extreme. Moreover, conventional FCE shows significant tangential electric field component (y-component), domain inversion region spreads more than electrode width (Fig.6(a)), production of small periodic domain inversion is difficult [14].

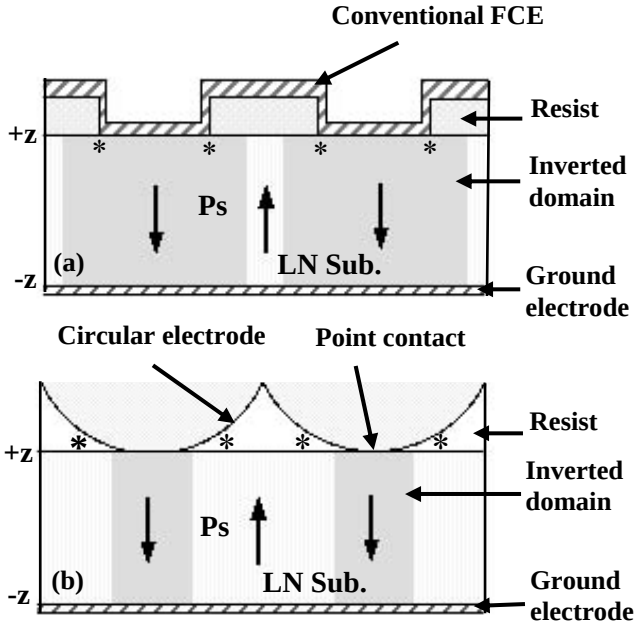


Fig. 6 (a) Cross-sectional view of conventional FCE and (b) New concept of point contact with circular electrode. Where * indicates the position of maximum electric field.

III. Proposal for CF-FCE

From the observations in section 2.B for conventional FCE, we thought that the electrode form which has E/E_R (E: electric field intensity below the electrode center and E_R: electric field intensity below the resist center) value beyond the lift-off electrode value (5.22 for LN crystal) and electric field concentrates directly under the electrode central part was important as a design indicator for producing small periodic domain inversion structures. The electrode form in which electric field concentrates on one place in the substrate serves as a point contact and can be considered circular or triangular. The circular electrode is easier to fabricate compared to the triangular one. Hence, we considered circular electrode in this paper. Since electric field concentrates at one point on the substrate (Fig.6(b)), tangential electric field component will be significantly reduced. Consequently, broadening of domain inversion region will be much smaller compared to the conventional FCE. In this case, nucleation would occur from one point where electric field concentration exists on the substrate. We calculated the electric field distribution of circular electrode. As a calculation model for the proposed CF-FCE, the conductive circle was put in order periodically at intervals of 2 μm on the substrate and the diameter of the circle was taken as the parameter (Fig.7). We calculated the electric field distribution in z-direction for 1V applied between circular and ground electrodes using successive over relaxation (SOR) method [15]. The calculated result of electric field distribution in the crystal and resist part (ε_r=3.0) is shown in Fig.8. Gray scale shows electric field intensity in crystal and resist. According to the numerical value of the electric field calculation, the E/E_R are 23.2

for 1 μm and 16.2 for 2 μm diameter circular electrodes. The E/E_R value for 2 μm diameter circular electrode is much higher than that of the lift-off electrode value (5.22) of LN crystal. We see that the maximum electric field exists inside the resist and the electric field is concentrated on one place to which the circle touches the crystal surface. In addition, as a result of calculation, electric field distribution depending on the diameter of a circle was obtained. It became clear from Fig. 8 that E/E_R increased as diameter decreased. Therefore, the production conditions of small periodic domain inversion were fulfilled by the circular form FCE method in which point contacts were obtained.

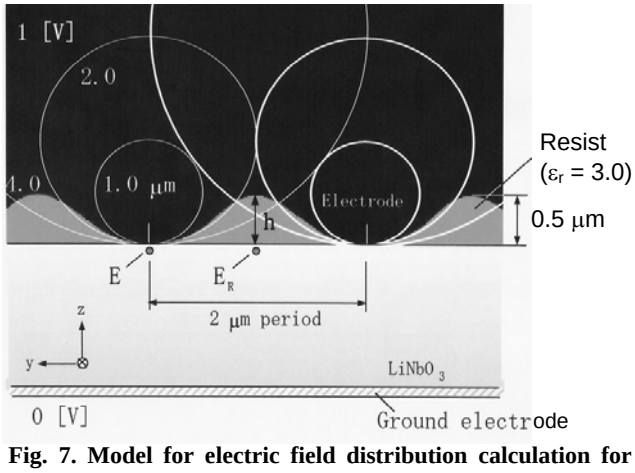


Fig. 7. Model for electric field distribution calculation for circular electrode by the SOR method.

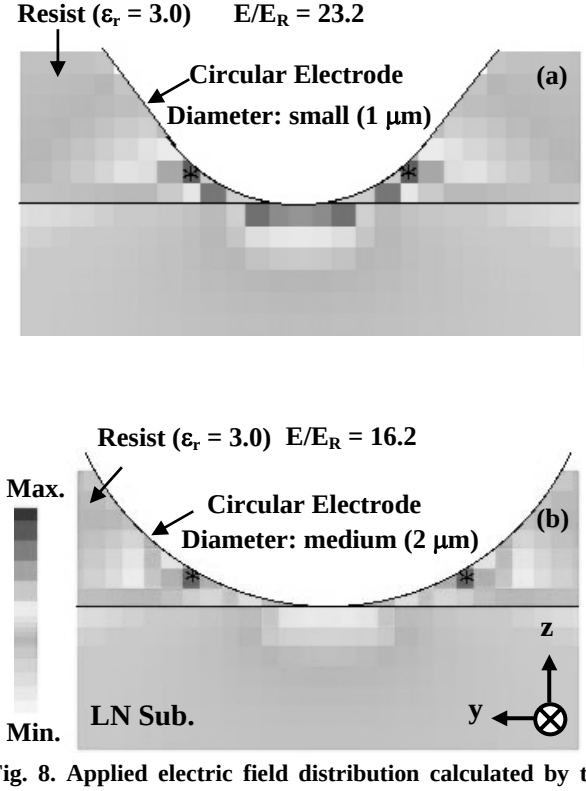


Fig. 8. Applied electric field distribution calculated by the SOR method [15] in the LN crystal induced by circular electrode, where * indicates the position of maximum electric field concentration: (a) 1 μm∅ and (b) 2 μm∅.

IV. Experiments, Results and Discussions

The z-cut C-LN crystal was used as the substrate. The thickness of the crystal was 500 μm. A 1.3 μm-thick photoresist (OFPR 800) was spin-coated on the +z surface and the periodic resist pattern was produced on +z surface by 2-beam laser interference exposure method (Fig.9). Ar^+ laser (488 nm, 1 mmø) was used for this interference exposure experiment. The incidence angle of the Ar^+ laser beams was set at 14^0 so that the period of the resist pattern became 2 μm after development. Initially, the laser beam power and exposure time were optimized for better resist profiles using test samples. Since the laser beam is Gaussian, the resist in the central part of the beam will be exposed most and exposure will be lower towards the periphery as confirmed by laser optical microscope (Fig.10). Therefore, we expect different types of resist profiles after exposure and development.

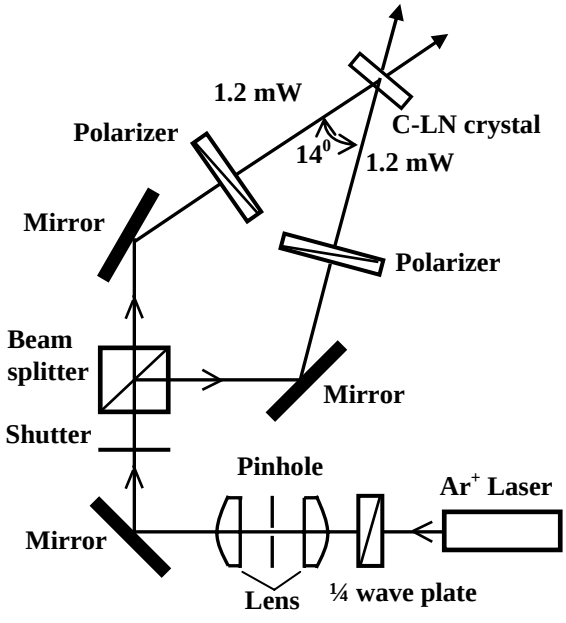


Fig. 9. Optical setup for 2-beam laser interference exposure experiment.

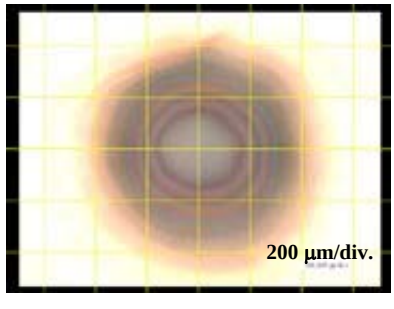


Fig. 10. Laser optical microscope observation of photoresist after 2-beam laser interference exposure and development.

Aluminum electrodes were produced on the +z and –z surfaces of the samples by thermal vacuum evaporation method. A high voltage pulse (Fig.11(a)) of 1 ms duration (11.4 kV, 1ms x 1pulse) for domain inversion was applied between electrodes at room temperature with samples immersed into Si-oil (Fig.11(b)). In one sample, the photoresist was removed by acetone. The residual Aluminum was etched by wet etching. The domain

inversion patterns were revealed by wet etching using a mixed-solution of HNO_3 and HF. The formed domain inversion pattern (from exposed center towards up) was observed by laser optical microscope (laser OM) and FE-SEM and are shown in Fig. 12. Around the exposed center (area A_1), we found that almost whole area of the C-LN has inverted domain (Fig.12(b)). Fine domain inversion pattern was observed in area A_2 (Fig.12(c)) and this may be due to point contact of electrode. The period of the domain inversion pattern is 2 μm. The domain inversion width for line-2 is 640 nm (Fig. 12(c)). Therefore, 2-μm periodic nanometer size domain inversion in a 500-μm-thick C-LN crystal is possible using this technique. The domain inversion width of upper line decreases as we move towards left (Fig. 12(c)). Domain inversion width as low as 350 nm were obtained for this line.

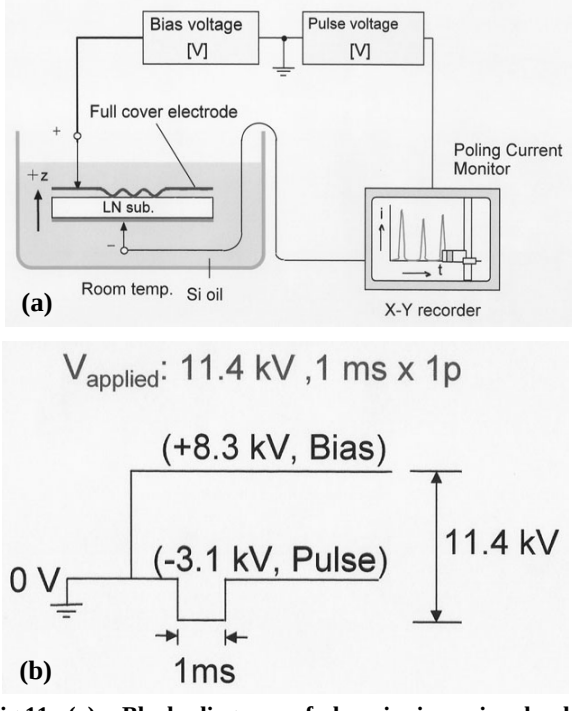


Fig.11. (a) Block diagram of domain inversion by high voltage pulser, and (b) Diagram of applied high voltage 11.4 kV (H.V. Pulser: -3.1 kV, DC bias voltage:+8.3 kV).

Then we tried to verify the exact resist profile types for which such smaller domain inversion widths occur. For this purpose, after high voltage application to another sample, the Al electrodes were etched by wet etching. A 56 nm Au layer were sputtered over the photoresist. Again photoresist was spincoated over the Au sputtered layer and the sample was postbaked at 80 ^{0}C for 30 min. The sample was cut by an automatic dicing machine (Disco Co. Ltd., Japan) and a marker was placed in the substrate near the center of exposed pattern. The cross-sectional views of the resist pattern were mapped by FE-SEM. The photoresist was removed by acetone. The residual Au was etched by a mixture of HNO_3:HCl (1:3 by vol.%) solution. Finally, domain inversion pattern was revealed by wet etching using a mixed-solution of HNO_3 and HF. The position of marker and domain inversion pattern on +z and y surface of the C-LN are shown in Fig. 13. The formed etched patterns from the marker were mapped by the same FE-SEM. The mapped etched patterns were then superimposed over the mapped resist patterns in order to determine the exact resist profile types for domain inversion regions (Fig. 14). The lower part of electrode adjacent to C-LN substrate can be represented by a circle of suitable diameter. This part of electrode is very important as maximum electric field and electric field concentration on C-LN substrate occur in this region as described in section III (Fig. 8). From Fig.14, we see that the electrode does not touch the substrate rather very small (~30 nm) residual resist exists between them. The smallest domain inversion width of 240 nm (line-1) is obtained for a residual resist height of 33 nm. Domain inversion widths of 410 nm (line-2) and 575 nm (line-3) are obtained for a residual resist height of 29 nm. The period of the domain inversion pattern is 2 μm. These results indicate that the domain pattern cannot be controlled well, although production of nanoscale domain inversion is possible using this technique. Therefore, rigid control of residual resist height is necessary for the effective fabrication of periodic domain patterns in the nanometer region using the proposed CF-FCE.

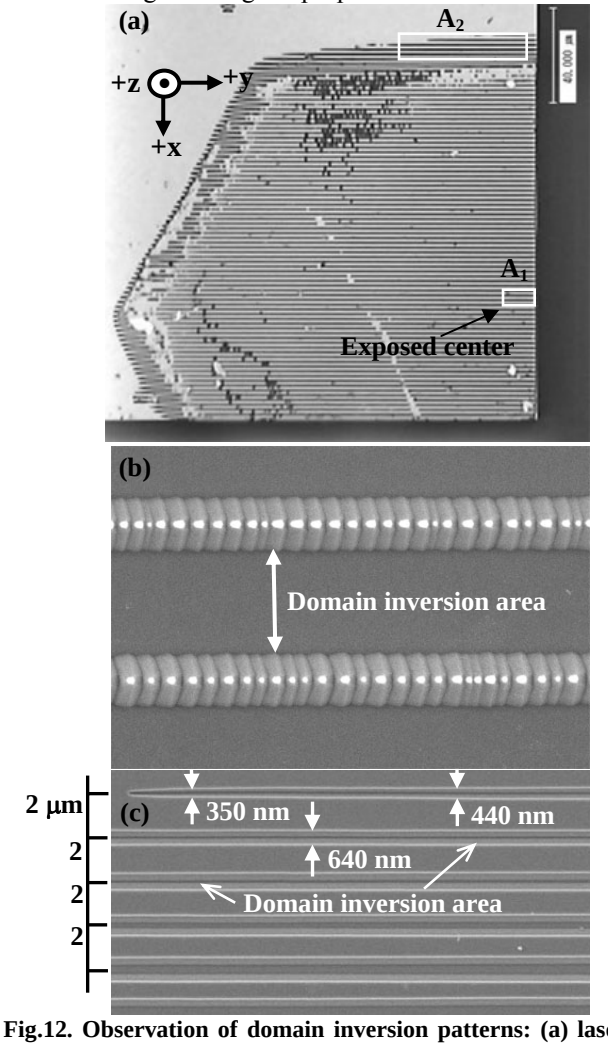


Fig.12. Observation of domain inversion patterns: (a) laser OM image, (b) FE-SEM image of area A_1, and (c) FE-SEM image of area A_2.

For high-efficient SHG devices, we need very deep periodic domain inversion pattern. For this reason, we determined the depths of domain inversion regions in y-

surface of Fig.13(b) by observing the cross-sectional view using FE-SEM. For line-1, domain inversion width is 240 nm (Fig. 14) and domain inversion depth is 146 μm. The calculated aspect ratio for this line is 608. For line-3, domain inversion width is 575 nm (Fig. 14) and domain inversion depth is 500 μm (i.e. whole substrate under line-3 is inverter). The calculated aspect ratio for line-3 is 870. Therefore, very-deep nanometer size domain inversion is possible using this technique which is highly demanded for high efficiency SHG devices as mentioned earlier. Once the resist profile types for particular domain inversion widths were determined, this technique could be extended for the fabrication of large-area, nanometer-size periodic domain inversion in LN substrates using a single laser beam scanning method which is under investigation.

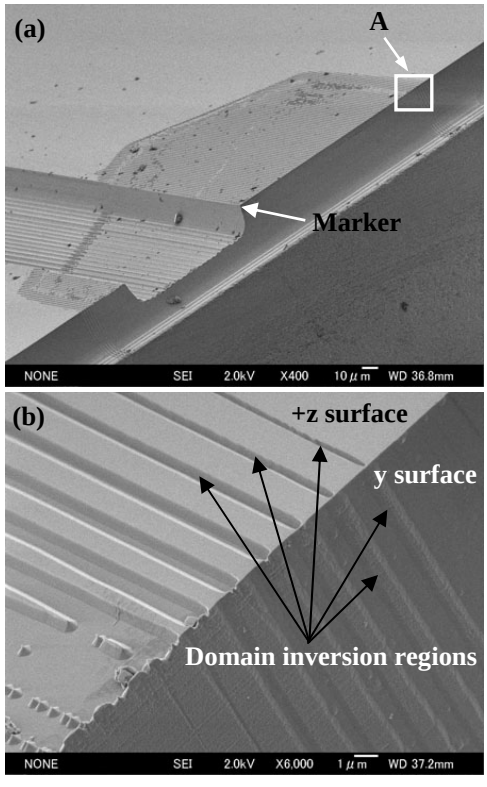


Fig. 13. FE-SEM images: (a) showing the positions of marker and fine domain inversion patterns and (b) zoom-in view of area A.

V. Conclusions

In this paper, a new method named "circular form FCE method" is proposed for nanoscale domain inversion in LN crystal which is the most important process for realizing high-power and high-efficiency QPM-SHG devices. The validity and limitation of the method is examined by the calculation of electric field distribution and experiments. Using the proposed method we successfully fabricated nanometer-size domain inversion. From the calculation of electric field distribution in the CF-FCE, it was shown clearly that the electric field was concentrated on one place of the crystal surface and the E/E_R became large. Rigid control of residual resist height is necessary for the effective fabrication of domain patterns in the nanometer region using the proposed CF-FCE. We verified the exact resist profile types for fine domain inversion patterns. Domain inversion width as low as 240 nm was obtained for a residual resist height of 33 nm. The results indicated that the domain inversion grew from the central part of electrodes and the tangential component of electric field is significantly reduced. The calculated aspect ratio for a 575 nm domain inversion width is 870. The high aspect ratio is strongly demanded for high efficiency QPM-SHG devices. Once the resist profile types for fine domain inversion patterns are known, this technique could be extended for the fabrication of large-area, nanometer-size domain inversion in LN substrates using a single laser beam scanning method, which is under investigation.

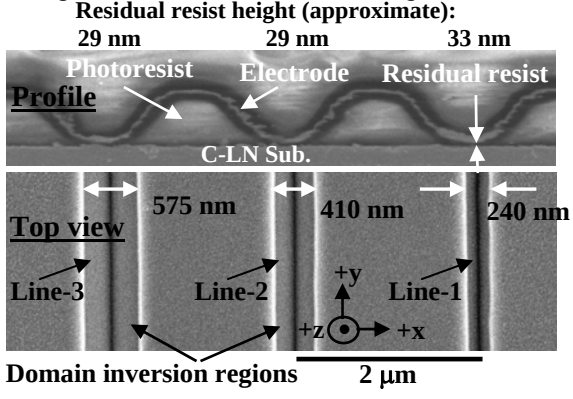


Fig. 14. Cross-sectional view of resist profile and top view of domain inversion regions on z-surface of C-LN substrate observed by FE-SEM.

References

[1] T. Suhara and H. Nishihara, J. Lightwave Technol., vol. 21, no.11, pp. 1097-1105, 1993.

[2] K. Yamamoto, J. Lightwave Technol., vol. 21, no. 11, pp. 1089-1096, 1993.

[3] T. Tanimura, M. Satho, and H. Itho, Oyo Buturi, vol. 67, no. 9, pp. 1046-1050, 1998.

[4] M. Minakata, M. Konishi, and T. Shiomi, in Proc. JICAST,98, pp. 165-168, 1998.

[5] M. Yamada, N. Noda, M. Saitoh, and K. Watanabe, Appl. Phys. Lett., vol.62, pp. 435-436, 1993.

[6] K. Mizuuchi, K. Yamamoto, and M. Kato, Appl. Phys. Lett., vol. 70, pp. 1201-1203, 1997.

[7] L. E. Myers, R. C. Eckardt, M. M. Fejer, R. L. Byer, W. R. Bosenberg, and J. W. Pierce, J. Opt. Soc. Am., vol. B12, pp. 2102-2116, 1995.

[8] K. Kintaka, M. Fujimura, T. Suhara, and H.Nishihara, IEEE J. Lightwave Technol., vol. 14, pp. 462-468, 1996.

[9] M. Minakata, M. Masahiro, T. Shiomi, and J. Zhong, Bull. Res. Inst. Electron. Shizuoka Univ., vol.33, pp. 77-84, 1998.

[10] V. Y. Shur, E. L. Rumyantsev, E. V. Nikolaeva, E. I. Shishkin, D. V. Fursov, R. G.. Batchko. L. A. Eyres, M. M. Fejer, and R. L. Byer, Appl. Phys. Lett., vol. 76, p. 143, 2000.

[11] C. Restoin and S. Massy: Opt. Mat., 22, pp.193-9. 2003.

[12] M. Minakata, S. Nagano, S. Yoneyama, and Y. Nakada, Optics Japan 2003, vol. 8pF11, pp. 194-195, 2003.

[13] K. Kintaka, M. Fujimura, T. Suhara, and H. Hishihara, Trans. Inst. Electron. Inform. Commun. Eng., vol. J78-C-1, pp. 238-245, 1995.

[14] S. Nagano, M. Konishi, T. Shiomi, and M.Minakata, Jpn. J. Appl. Phys., vol. 42, pp. 4334-4339, 2003.

[15] M. Minakata and M. Goto, Bull. Res. Inst. Electronics Shizuoka Univ., vol. 31, pp. 53-59, 1996.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Thermal Stress Effects on Higher Order Modes in Highly Elliptical Core Optical Fibers

Rahat M. Anwar and M. Shah Alam

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology (BUET)
Dhaka – 1000, Bangladesh
Email: sarkar.rahat@gmail.com, shalam@eee.buet.ac.bd

Abstract - The effect of thermally induced stress on the fundamental and higher order modes of highly elliptical core optical fibers is studied in this work using a finite element method. The stress analysis is carried out and anisotropic refractive index change is calculated using the plain strain approximation. After considering the stress optical effect, the modal analysis of the fiber is performed to obtain the modal solutions. The modal effective index, modal birefringence, and field distributions of different modes of the fiber are presented. The modal birefringence of the fundamental and higher order modes as a function of ellipticity of such fibers are also presented. It is seen that change in ellipticity causes significant change in birefringence for fundamental and higher order mode of the fiber.

I. Introduction

High birefringence fibers (Hi-Bi fiber) or polarization maintaining fibers (PMF) have been widely used for polarization control in fiber optic sensors, precision optical instruments, and optical communication systems [1]-[5]. Different structured fibers like bow-tie fibers, PANDA fibers, side-pit fibers, side-tunnel fibers have been analyzed in depth previously [3]-[7]. These types of fibers are designed to introduce a differential stress in the core so that the state of polarization of the guided light wave can be maintained by means of stress-optic effect [3]-[7]. Another fiber is designed to have an asymmetric core shape, *i.e.*, elliptical core to introduce birefringence in order to maintain the polarization [1]-[9].

In the elliptical core fibers, the birefringence primarily comes from the core ellipticity. But, like the other fibers, there is temperature induced birefringence also. Both the effects should be considered very carefully to find the correct modal characteristics. In many cases, they have been considered, but only for fundamental modes [1]-[5], [7], [10]-[11]. In [8]-[9], modal analysis is carried out, but stress effect is not considered. As for analysis methods, many different methods are used for analyzing PMFs [6]-[13]. Among them the finite element method (FEM) is suitable for producing accurate modal solutions of different types of fibers [7]-[9], [11], [14].

In this work, a simple structure with highly elliptical core optical fiber with a circular cladding is analyzed by using the FEM [14]. The plain strain approximation is used to include the stress-optic effect and a modal solver with the electric field as the working variable is used to obtain the modal solutions. The effect of thermally induced stress on the fundamental and higher order modes of highly elliptical core optical fibers is thus studied. It is seen that higher ellipticity of the core results much higher thermal stress in the fiber core, which in turn produces higher birefringence to produce non-degenerate modes in the fiber.

II. Theory

There are two sources of birefringence in an optical fiber [6]. One of them is the structure of the fiber. This type of structural birefringence has a relatively small value (at the range of 10^{-6}). The other source of birefringence is the thermally induced stress. During the manufacturing process, due to different thermal expansion coefficients of core and cladding material, the fiber becomes stressed. This stress causes the indices to become anisotropic and increases the birefringence of orthogonal polarized wave in the fiber. This birefringence has a relatively higher value (at the range of 10^{-4}).

In an optical fiber, a high value of birefringence is usually desired for single mode operation. To understand the overall birefringence of a fiber with elliptical core, it is necessary to carry out the stress analysis first to find the stress distribution and hence the change in refractive index due to stress-optic effect. Next, the optical analysis is to be performed with the new refractive index to obtain the modal solutions and the birefringence.

A. Stress Analysis

In the stress analysis, the effect of thermally induced stress on the refractive index in different domains in the cross section of the fiber is determined. Using tensor notation, the general linear stress-optical relation can be written as [14]

$$\Delta n_{ij} = -C_{ijkl}\sigma_{kl} \qquad (1)$$

where $\Delta n_{ij} = n_{ij} - n_0 I_{ij}$, n_{ij} is the refractive index tensor, n_0 is the refractive index for a stress-free material, I_{ij} is the identity tensor, C_{ijkl} is the stress-optical tensor, and σ_{kl} is the stress tensor. This constitutive relation can be characterized by some independent parameters in the

stress optical tensor. As a result of the symmetry, the number of these independent parameters can be reduced. Since n_{ij} and σ_{kl} are both symmetric, one gets $C_{ijkl} = C_{jikl}$ and $C_{ijkl} = C_{ijlk}$. The number of independent parameters can be further reduced and in the simplest stress analysis only two independent parameters, C_1 and C_2 can be used. So, with the reduction of the number of parameters, the stress-optical relation is simplified to

$$\begin{bmatrix} \Delta n_x \\ \Delta n_y \\ \Delta n_z \end{bmatrix} = - \begin{bmatrix} C_2 & C_1 & C_1 \\ C_1 & C_2 & C_1 \\ C_1 & C_1 & C_2 \end{bmatrix} \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \end{bmatrix}, \qquad (2)$$

where $n_x = n_{11}$, $n_y = n_{22}$, $n_z = n_{33}$, $\sigma_x = \sigma_{11}$, $\sigma_y = \sigma_{22}$, and $\sigma_z = \sigma_{33}$. Thus, one can write (2) in the following form

$$\begin{aligned} n_x &= n_0 - C_2\sigma_x - C_1(\sigma_y + \sigma_z) \\ n_y &= n_0 - C_2\sigma_y - C_1(\sigma_z + \sigma_x) \cdot \\ n_z &= n_0 - C_2\sigma_z - C_1(\sigma_x + \sigma_y) \end{aligned} \qquad (3)$$

It is assumed by using the two parameters C_1 and C_2, that the non-diagonal parts of n_{ij} and σ_{kl} are negligible. This means that the shear stress corresponding to $\sigma_{12} = \tau_{xy}$ is neglected. In addition, using the plane strain approximation, one can neglect the shear stresses corresponding to $\sigma_{13} = \tau_{xz}$ and $\sigma_{23} = \tau_{yz}$. The plane strain approximation can be applied in the case where the strain in z-direction is assumed to be zero and the structure is free in the x- and y-direction. This deformation state is not correct if the structure is free also in the z-direction, which is not included in this work.

B. Optical Analysis

In the optical analysis, the resulting anisotropic refractive indices as calculated by (3) are used. The modal analysis is carried out assuming that the wave propagates along the z-direction and the electric field of the wave has the form [14]

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(x, y) \exp[j(\omega t - \beta z)], \qquad (4)$$

where ω is the angular frequency and β is the propagation constant. An eigenvalue equation in terms of the electric field can be obtained from the Helmholtz equation

$$\nabla \times (n^{-2} \nabla \times \mathbf{E}) - k_0^2 \mathbf{E} = \mathbf{0}, \qquad (5)$$

and is solved for modal effective index, $n_{\text{eff}} = \beta / k_0$, as the eigenvalue. The boundary condition for electric field at the outside of the cladding boundary was set to zero. In the FEMLAB [14], however, a module based on the perpendicular hybrid mode wave using transversal fields is used for finding the modal solutions.

III. Results and Discussion

The cross section of the optical fiber with elliptical core is shown in Fig. 1, where the core dimensions in x-direction (major axis) and y-direction (minor axis) are $2a$ and $2b$, respectively. The cladding is circular with radius, r=25 μm. The refractive indices of core and cladding are n_1=1.523 and n_2=1.518, respectively. For analysis with the FEMLAB [14], the cross section of the fiber is meshed with quadratic triangular elements. For stress analysis, the boundary conditions are taken in such a way that along x-axis through the center of the core, no displacement is allowed in y-direction, and along y-axis through the center of the core, no displacement is allowed in x-direction, and in other regions, no restrictions are applied.

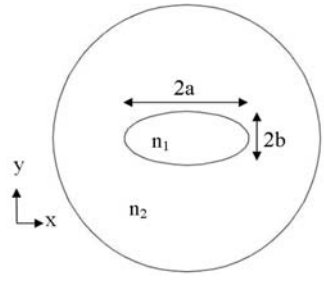
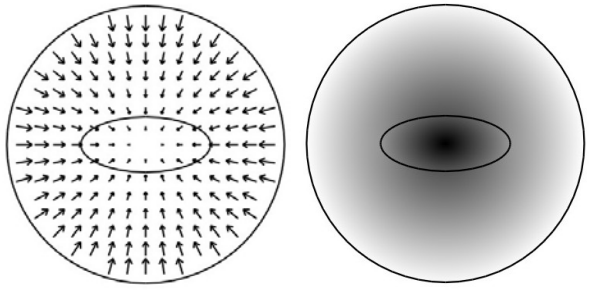


Fig. 1: Cross section of the optical fiber with elliptical core.

The stress-optic coefficients, C_1=4.184×10^{-12} m^2/N and C_2=7.5714×10^{-13} m^2/N. The thermal expansion coefficients of core and cladding are, α_{core} = 2×10^{-6} K^{-1} and α_{clad} = 1×10^{-6} K^{-1}. The Young's modulus of the fiber material, E = 72.324×10^9 Pa and the Poisson's ratio, υ= 0.186. The initial temperature of the fiber is 1020°C and the final temperature when the fiber is cooled down is 20°C. So the temperature difference is 1000°C, which causes the strain over the fiber cross section. As the core is not circular and have different thermal expansion



(a) (b)

Fig. 2: (a) Vector displacement, and (b) Stress distribution, over the cross section of the fiber.

coefficient than the cladding, the stress distribution will not be equal in all the directions. Fig. 2 shows the vector displacement and stress distribution over the cross section of the fiber that occur because of different thermal expansion coefficients. The length of the arrow in Fig. 2(a) is proportional to the amount of displacement and the direction of the arrow is the direction of the displacement. It can be seen easily that the displacement is towards the center of the fiber and lateral contraction occurs. From Fig. 2(b), it can be seen that the stress is more at the core than at the cladding region. This is expected as the displacement is towards the center, more stress develops at the center of the fiber.

Due to the stress-optic effect, the refractive index of both core and cladding changes. This is shown in Fig. 3. With the development of stress, it can be seen that in the core region, the refractive index is decreased from 1.523 to 1.52275, and in the cladding, the refractive index has a minimum value of 1.5176 at the interface of core and cladding, it increases towards the edge of the fiber to 1.5179, which is a closer value to the original value of 1.518.

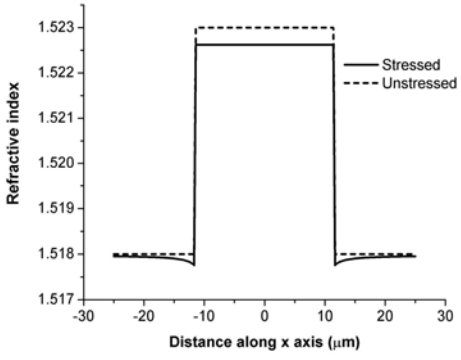
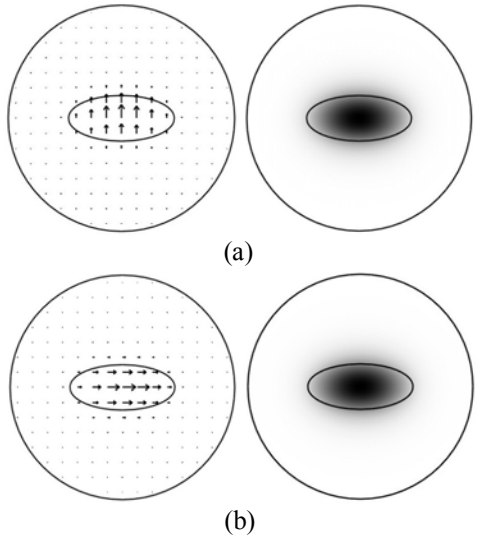


Fig. 3: Variation of refractive index along *x*-axis.

The anisotropic change in refractive index due to stress-optic effect will cause a higher value of birefringence in the fiber and the birefringence will cause the fundamental and higher order modes to split, which are usually two-fold degenerate modes in unstressed fiber. The electric field intensity distribution of different modes over the cross section of the stressed fiber are shown in Fig. 4, where a=11.67 μm, b=5 μm, and ellipticity, ξ=0.4 are taken for the calculation. Here, the ellipticity is defined by

$$\xi = \frac{(a-b)}{(a+b)}.$$

In Fig. 4, fundamental and higher order modal solutions are shown, where part (a) and (b) show *y*-polarized and *x*-polarized fundamental HE_{11} modes, respectively, part (c) and (d) show *y*-polarized TE_{01} mode and *x*-polarized TM_{01} modes, respectively, part (e) and (f) show *y*-polarized HE_{21} mode and *x*-polarized HE_{21} modes, respectively.



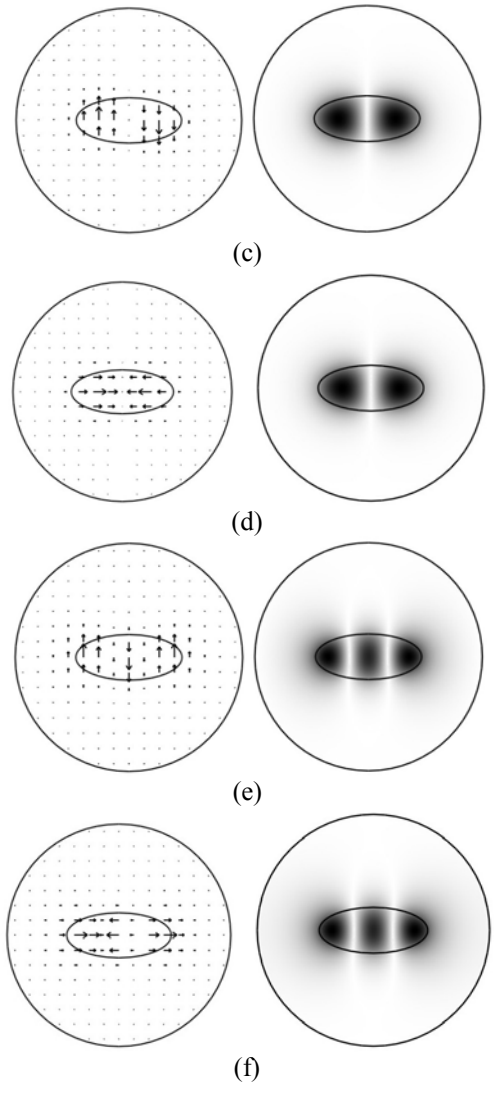


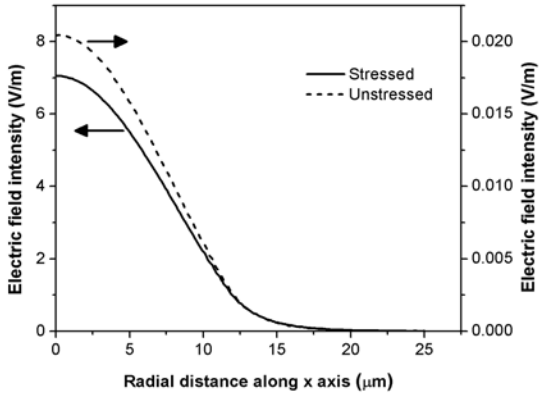
Fig. 4: Electric field intensity distribution of different modes of the fiber, (a) HE_{11}, (b) HE_{11}, (c)TE_{01}, (d) TM_{01}, (e) HE_{21}, (f) HE_{21}.

Next, in Fig. 5, the dominant electric field intensity of the thermally stressed and unstressed fibers are compared for different modes. The electric field value is taken over the radial line along the *x*-axis from the center of the fiber. From these curves, it can be seen that for the modes considered here, the stressed fiber has higher electric field intensity than the unstressed fiber. Though the shape of the field intensity distribution is similar, the electric field intensity of stressed fiber is about 300-500 times higher than the electric field intensity of the unstressed fiber for each of the modes considered here.

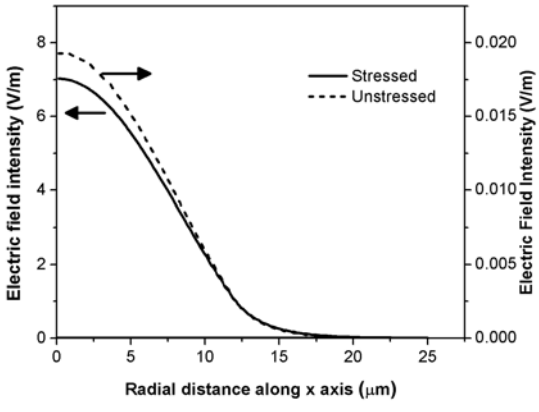
Finally, in Fig. 6, the birefringence of optical fiber for fundamental and higher order modes are shown as a function of ellipticity. The solid line, dashed line, and dotted line show the birefringence of fundamental HE modes, TE-TM modes, and higher order HE modes, designated by B_{11}, B_{01}, and B_{21}, respectively.

It can be seen that the birefringence of the fundamental HE modes, B_{11}, reaches at a peak value at the ellipticity of 0.7. It can also be seen that higher the mode, smaller the

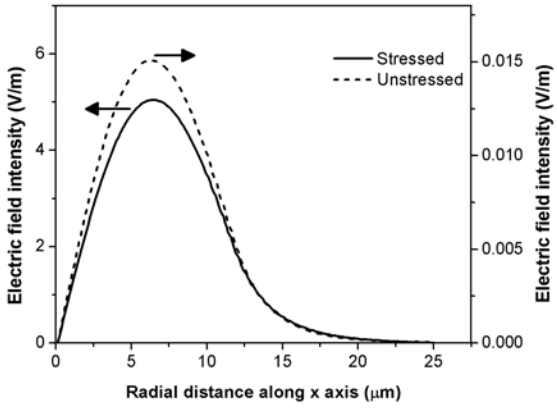
value of birefringence and smaller the ellipticity in which the peak occurs. Furthermore, the birefringence of the higher order modes, B_{01} and B_{21} decreases after reaching the peak at the ellipticity of approximately 0.6 and 0.65, respectively. However, it can also be noticed that the stressed fiber has higher value of birefringence for both fundamental and higher order modes than the unstressed fiber.



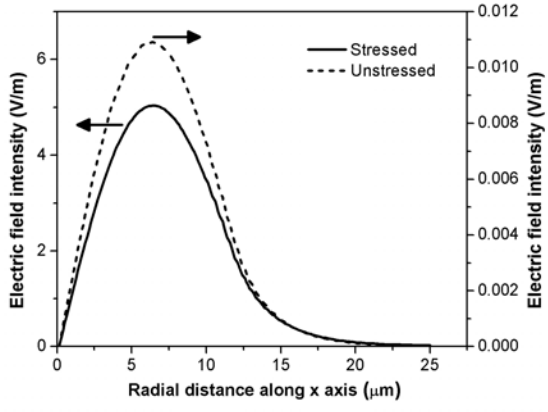
(a)



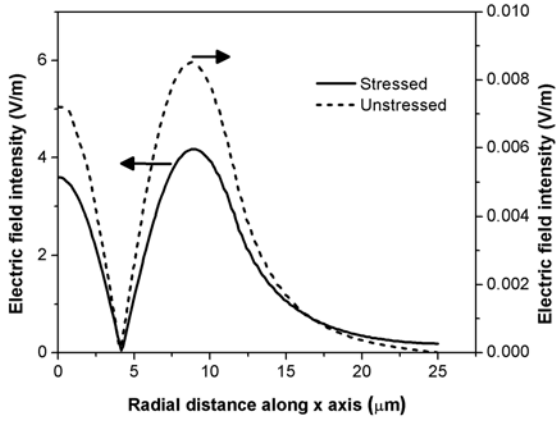
(b)



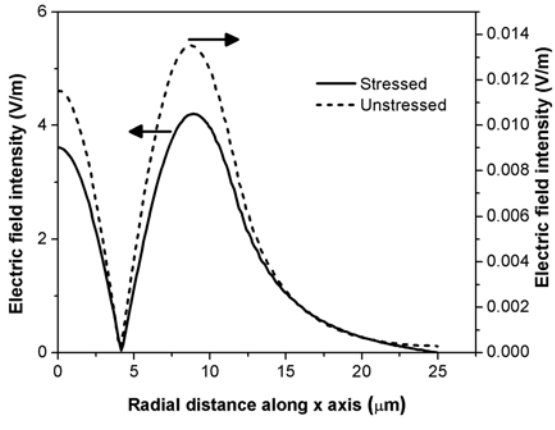
(c)



(d)



(e)



(f)

Fig. 5: Electric field intensity over the radial distance along *x*-axis, (a) HE_{11}, (b) HE_{11},(c) TE_{01}, (d) TM_{01}, (e) HE_{21}, and (f) HE_{21} mode.

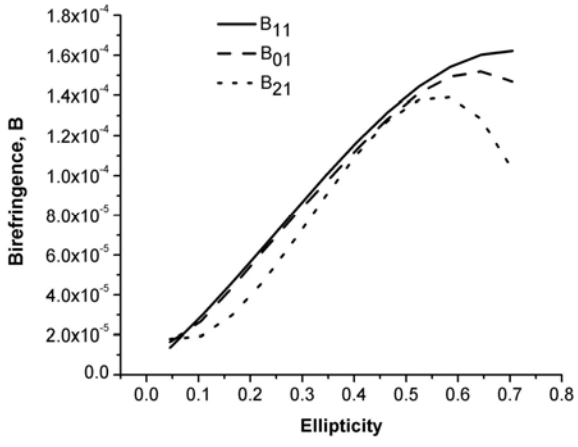


Fig. 6: Variation of birefringence with ellipticity.

IV. Conclusion

The effect of thermally induced stress on fundamental and higher order modes of a highly elliptical core optical fiber is studied. In an unstressed fiber, the birefringence of different modes have lower values than those of stressed fibers. The stress developed in a fiber during manufacturing process, causes fundamental and higher order modes of the fiber to have higher value of birefringence, which could be highly advantageous in practical usage. Also, the value of birefringence and the ellipticity of the fiber at which the maximum value of birefringence occurs decreases with the increase in the order of the modes. All these can be used to obtain non-degenerate higher order modes in fiber applications.

References

[1] W. Urbanczyk, T. Martynkien, and W. J. Bock, "Dispersion effects in elliptical-core highly birefringent fibers," *Appl. Optics.*, vol. 40, no. 12, pp. 1911-1920, April 2001.

[2] I. –K. Hwang, Y. –H. Lee, K. Oh, and D. N. Payne, "High birefringence in elliptical hollow optical fiber," *Optics Express*, vol. 12, no. 9, pp. 1916-1923, May 2004.

[3] J. Noda, K. Okamoto, and Y. Sasaki, "Polarization-Maintaining Fibers and Their Applications," *IEEE J. Lightwave Technol.*, vol. LT-4. no. 8, pp. 1071-1089, Aug. 1986.

[4] Y. Liu, B.M.A. Rahman, and K.T.V. Grattan, "Analysis of the Birefringence Properties of Optical Fibers Made by a Preform Deformation Technique," *IEEE J. Lightwave Technol.*, vol. 13, no. 2, pp. 142-147, Feb. 1995.

[5] Y. Liu, B. M. A Rahman, and K.T.V. Grattan, "Thermal-stress-induced birefringence in bow-tie optical fibers," *Appl. Opt.*, vol. 33, no. 24, pp. 5611-5616, Aug. 1994.

[6] K. Okamoto, *Fundamentals of Optical Waveguides*, Academic Press, 2000.

[7] K. Okamoto, T. Hosaka, and T. Edahiro, "Stress analysis of optical fibers by a finite element method," *IEEE J. Quantum Electron.*, vol. QE-17, pp. 2123-2129, Oct. 1981.

[8] M. Eguchi and M. Koshiba, "Accurate Finite-Element Analysis of Dual-Mode Highly Elliptical-Core Fibers," *IEEE J. Lightwave Technol.*, vol. 12. no. 4, pp. 607-613, Apr. 1994.

[9] M. Eguchi and M. Koshiba, "Behavior of the First Higher-Order Modes of a Circular Core Optical Fiber Whose Core Cross-Section Changes into an Ellipse," *IEEE J. Lightwave Technol.*, vol. 13. no. 2, pp. 127-136, Feb. 1995.

[10] T. Schreiber, H. Schultz, O. Schmidt, F. Roser, J. Limpert, and A. Tunnermann, "Stress-induced birefringence in large-mode-area micro-structured optical fibers," *Optics Express*, vol. 13, no. 10, pp. 3637-3645, May 2005.

[11] M. S. Alam, N. Somasiri, B. M. A. Rahman, and K. T. V. Grattan, "Effects of High External Pressure on Photonic Crystal Fiber," *Proceedings of the Third International Conference on Electrical and Computer Engineering,* ICECE 2004*,* Dhaka, Bangladesh, pp. 245-248, Dec. 2004.

[12] H. Shu and M. Bass, "Calculating the Guided Modes in Optical Fibers and Waveguides," *IEEE J. Lightwave Technol.*, vol. 25, no. 9, pp. 2693-2699, Sept. 2007.

[13] Y. Zhu, X. Chen, Y. Xu, and Y. Xia, "Propagation Properties of Single-Mode Liquid-Core Optical Fibers With Subwavelength Diameter," *IEEE J. Lightwave Technol.*, vol. 25, no. 10, pp. 3051-3056, Oct. 2007.

[14] COMSOL -Electromagnetics Module, -Structural Mechanics Module, version 3.2, Sept. 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Equivalent Circuit Model for Dual-Cavity QW-VCSELs

F. Emami, A. H. Jafari

Optoelectronic Research Centre of Shiraz University of Technology, Shiraz, IRAN
E-mail: emami@sutech.ac.ir a.h.jafari@sutech.ac.ir

Abstract - In this paper we propose a new circuit model for a double cavity quantum well vertical cavity surface emitting laser. The detail of our model is based on independent rate equations. The carrier density, optical modes and their powers are described independently. The proposed equivalent circuit is simulated by a simulator and the results are compared with a model that the rate equations are solved numerically.

I. Introduction

Vertical cavity surface emitting lasers (VCSEL) are one of the most important sources in the optical communication systems due to their high quality characteristics, low distortion and wide band modulation responses [1], [2]. To improve the performance of these lasers and tune ability considerations it is possible to combine a VCSEL with the coupled cavity (CC) mechanism (CC-VCSEL) [3], which contains two VCSEL cavities with a common mirror between them [4]. Specifications of these structures are described by Pellandini [5] for the first time. These lasers can operate in two different modes [6] and they can pumped optically [5] or electrically [7]. There are many works on CC-VCSEL in the literature (such as operation characteristics and the current variations of each cavity on the threshold conditions of the other cavity [8]). When we consider several transverse modes in the CC-VCSEL, we must find a circuit model [9] and analyse the structure considering the important effects such as hole burning, carrier diffusion and small signal analysis. In fact, the physical model for these lasers is usually very complicated and so we need a long time to analysing them. Hence, finding an equivalent circuit model, and then studying it by a circuit simulator, is a good idea.
Using the rate equations, it is possible to describe the modulation behaviour of a CC-VCSEL numerically [10]. Indeed, the temperature characteristics of VCSELs are survived by using a circuit simulator [11].

II. Theory

In electrically pumped CC-VCSEL lasers, the cavities are activated by two driving currents which depend to the current levels and their structures, as shown in Fig. 1. Two cavities of a CC-VCSEL are corresponded to two wavelengths λ_L and λ_S, containing an active layer with *8nm* InGaAs quantum wells (QW). The *'S'* and *'L'* indices are related to the short (upper cavity) and long (lower cavity) sections of the structure. There is a semi-transparent mirror to separate the upper and lower cavities and preparing mutual coupling between them. In other hand, there are mirrors at the both sides of each cavity with large reflectivity (typically 99.8% and 99.6%). In these structures, there are three independent contacts. Each cavity is excited by a current I_i ($i=1, 2$) [3].

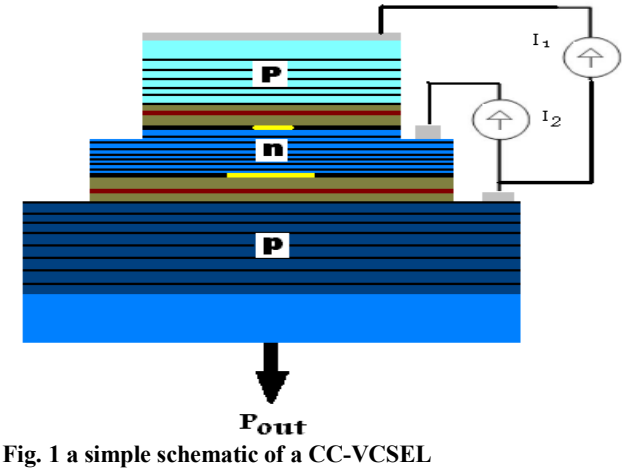


Fig. 1 a simple schematic of a CC-VCSEL

To understand the equivalent circuit of the structure, consider the time dependent concentrations and powers of the cavities, or the rate equations used in the VCSELs as follows [3]:

$$\frac{dN_1}{dt} = \frac{\eta_1 I_1}{q} - \frac{N_1}{\tau_e} - C_1^S g_1{}^s v(\frac{l_1}{L}) P^S - C_1^L g_1{}^L v(\frac{l_1}{L}) P^L \quad (1)$$

$$\frac{dN_2}{dt} = \frac{\eta_2 I_2}{q} - \frac{N_2}{\tau_e} - C_2^S g_2{}^s v(\frac{l_2}{L}) P^S - C_2^L g_2{}^L v(\frac{l_2}{L}) P^L \quad (2)$$

$$\frac{dP^L}{dt} = C_1^L g_1{}^L v(\frac{l_1}{L}) P^L + C_2^L g_2{}^L v(\frac{l_2}{L}) P^L - \frac{P^L}{\tau_p{}^L} - R_{sp}{}^L \quad (3)$$

$$\frac{dP^s}{dt} = C_1^S g_1{}^s v(\frac{l_1}{L}) P^S + C_2^S g_2{}^s v(\frac{l_2}{L}) P^S - \frac{P^S}{\tau_p{}^s} - R_{sp}{}^s \quad (4)$$

where N_i shows the carrier numbers of each cavity, $P^{L,S}$ are the photon numbers of *'Long'* and *'Short'* modes, v is group velocity, η_1 and η_2 are the fractions of the injected

currents enter to the active regions (less than 1 in sub-threshold and nearly 1 at threshold and above it). In the above equations, it is assumed that, L is the effective cavity length and l_i is the total length of the QWs in the i-th cavity. The $C_i^{L,s}$ factors are the structure constants [10]. R_{sp} describes the spontaneous emission mechanism and τ_e and τ_p are the carrier and the photon lifetimes respectively. We can define the gain parameters versus carrier densities as:

$$g_1{}^S = G_0{}^S Ln\frac{n_1 + n_0{}^S}{n_{tr}{}^S + n_0{}^S} \quad (5)$$

$$g_2{}^S = G_0{}^S Ln\frac{n_2 + n_0{}^S}{n_{tr}{}^S + n_0{}^S} \quad (6)$$

$$g_1{}^L = G_0{}^L Ln\frac{n_1 + n_0{}^L}{n_{tr}{}^L + n_0{}^L} \quad (7)$$

$$g_2{}^L = G_0{}^L Ln\frac{n_2 + n_0{}^L}{n_{tr}{}^L + n_0{}^L} \quad (8)$$

where n_{tr} is the transparency carrier density, n_i is the carrier density and $n_0^{L,S}$ denotes the QW absorptions at the corresponding wavelengths. The gain could be used by a linear approximation too.

The equations (1)-(4), (which are like the KCLs) show a system of first order coupled differential equations and can be solved by a simple circuit simulator. We can consider the carrier densities and the photon numbers (N_i and $P^{L,S}$) as the voltages on the resistors and the capacitors respectively and hence the time variations of them will be proportional the capacitor currents. Based on this idea, we can model the equations (1)-(4) in the form of a circuit shown in Fig. 2. The typical parameter values are listed in Table I. The circuit contains a capacitor and some dependent current sources (whose their currents are depend on the node voltages or equivalently N_i and $P^{L,S}$), which define the right hand sides of the coupled equations. The circuit input signals are I_1 and I_2 which denote by current sources as $\eta_i I$ in the circuit. Now, we can analyze the equivalent circuit and find the $V_{L,S}$ and V_{Ni} in each mode. To have a better convergence, we should scale the related equations using proper scaling factors.

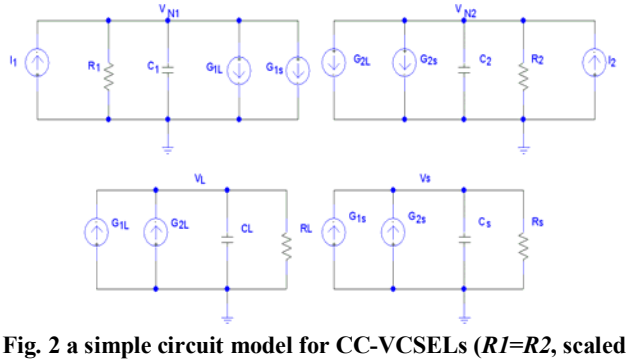


Fig. 2 a simple circuit model for CC-VCSELs (*R1=R2*, scaled for τ_e)

Now, consider the steady state case such that *d/dt=0* in the equations (1)-(4). When two modes are in the sub-threshold region both P^L and P^S are nearly zero and hence we can simplify the circuit model shown in Fig. 2 as considered in Fig. 3. In this case, at each time, we consider a fixed value of I_2 and I_1 varies from *0* to *0.38mA*. In the next step, we change the value of I_1 from *0.4mA* until *0.75mA*. For the steady state case, we have only the resistors and sources.

Table 1 Simulation Parameters.

Parameter	Value
τ_e	1ns
Γ_1	0.24
Γ_2	0.76
ζ_1	1.63
ζ_2	1.92
τ_p^S	0.289ps
τ_p^L	0.257ps

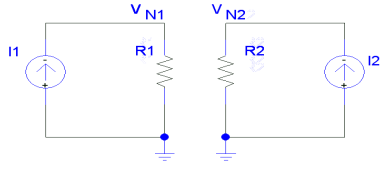


Fig. 3 simplified model in sub-threshold region for two modes

The lasing action is occurred for this region and we can reconstruct the equivalent circuit of Fig. 2 as Fig. 4.

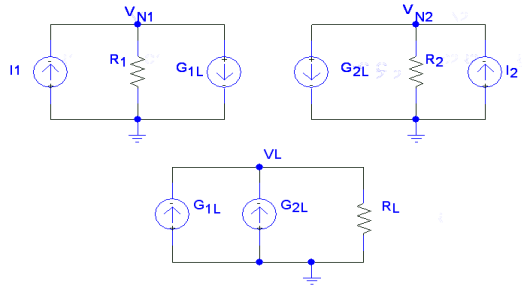


Fig. 4 simplified model of a CC-VCSEL, when the upper cavity is above the threshold and the lower cavity is under the threshold

The lasing *'L'* mode can be expressed by the following equations:

$$\frac{\eta_1 I_1}{q} - \frac{N_1}{\tau_e} - c^L{}_1 g_1{}^L v\frac{l_1}{L} p^L = 0 \quad (9)$$

$$\frac{\eta_2 I_2}{q} - \frac{N_2}{\tau_e} - c^L{}_2 g_2{}^L v\frac{l_2}{L} p^L = 0 \quad (10)$$

$$(c^L{}_1 g_1{}^L v\frac{l_1}{L} + c^L{}_2 g_2{}^L v\frac{l_2}{L})p^L - \frac{p^L}{\tau_p{}^L} = 0 \quad (11)$$

Now, if I_1 exceeds than *0.75mA* (but less than *1.25mA*) *'S'* mode will start to lase too. In this case both modes are above the threshold. There are two input sources I_1 and I_2 and two constants V_{N1} and V_{N2} for this case. If V_L decreases, V_S increases and lasing will be start. In all cases I_2=*1.5mA*.

Finally, for I_1>*1.25mA*, the *'L'* mode can not lase or

equivalently the carrier density of the lower cavity lies in sub-threshold ($P_L \approx 0$). In this case, we have (1)-(4) as:

$$\frac{\eta_1 I_1}{q} - \frac{N_1}{\tau_e} - c^s{}_1 g_1{}^s v \frac{l_1}{L} p^s = 0 \quad (12)$$

$$\frac{\eta_2 I_2}{q} - \frac{N_2}{\tau_e} - c^s{}_2 g_2{}^s v \frac{l_2}{L} p^s = 0 \quad (13)$$

$$(c^s{}_1 g_1{}^s v \frac{l_1}{L} + c^s{}_2 g_2{}^s v \frac{l_2}{L}) p^s - \frac{p^s}{\tau_p{}^s} = 0 \quad (14)$$

A proposed equivalent circuit of these equations is in the form of Fig. 5. In this case, we have only the *'S'* mode.

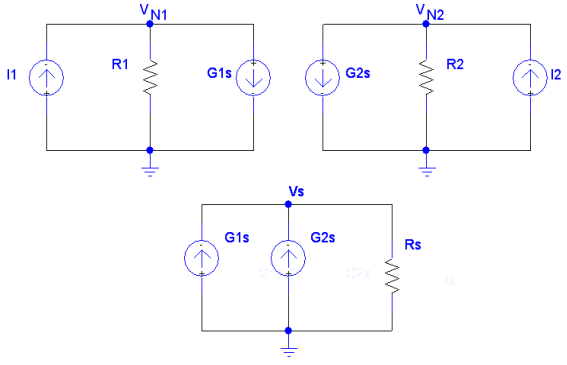


Fig. 5 simplified model of a CC-VCSEL, when the upper cavity is in the sub-threshold and the lower cavity is in the above threshold

III. Simulation results and discussion

In this section the above circuits are used and their results will compare with the previous reports [3]. To investigate the characteristics of CC-VCSEL, we can increase the first cavity current (short cavity) from zero when the longer cavity is fed by a fixed current. For each step, we compute the output photons in the corresponding wavelengths and carrier densities in each cavity. At first, when $I_1=0.0mA$ both cavities are in the sub-threshold region. At the same time I_2 is a fixed value so, there is not any generated photons. From the previous equivalent circuit and considering the fact that the equivalent resistor is a constant, the node voltage V_{N2} and hence N_2 is fixed. Note that, in sub-threshold regime for any increment in I_1, there is an increment in V_{N1} (Fig. 3). In this case, $V_L=0$ and there is not any emitted photons.

If the current in the first cavity, I_1, is increased the equivalent circuit should be in the form of Fig. 4. For this case any increment in I_1 is equivalent to a growth in V_{N1}. On the other hands, since G_{1L} (scaled to the related gains and photon numbers) depends on V_{N1}, the value of this parameter and hence the amount of V_L ($=I_L\ R_L$) is increased, with a constant R_L. This is valid for G_{2L}, since it is related to V_L. Because of a constant I_2 the current through R_2 decreases, and hence V_{N1} or equivalently N_1 is diminished. Increasing I_1 causes both cavities lie above the threshold. In this case, the carrier density of each cavity is constant and by increasing I_1, the photon density of long mode decreases and the photon density of short mode increases. The results of the carriers and output power are shown in Fig. 6 and 7 respectively. As shown, the laser contains two wavelengths λ_1 and λ_2. By increasing I_1 the output power in the longer wavelength (λ_2) decreases and it goes into the sub-threshold regime. So, the first cavity can generate power (in λ_1) as shown in Fig. 7.

IV. Conclusion

We have proposed a simple circuit model of CC-VCSEL characteristics which uses a set of rate equations, under some reasonable assumptions. In this equivalent circuit, the number of carriers and photons are replaced by scaled voltage nodes. Also the products of nonlinear terms which contain gain and photon density are replaced by dependent current sources. Therefore, the general electrical circuit simulation technique is obtained that can be used to predict the device behaviors by this simple model. As we have seen, the model exhibits good compatibility between numerical analysis and reported empirical results, suggesting its usefulness for describing a variety of CC-VCSELs.

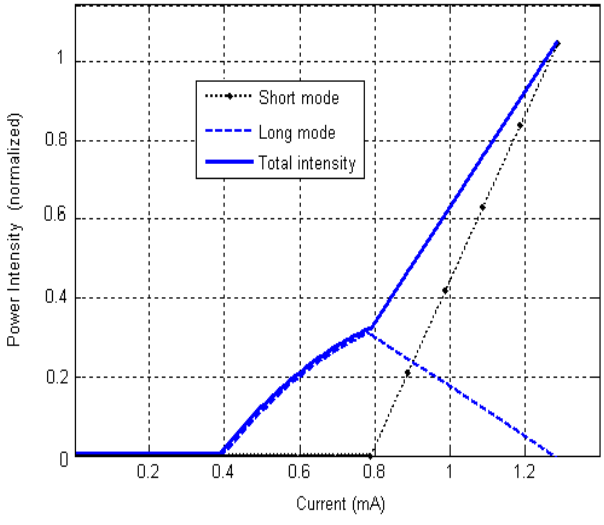


Fig. 6 Photon density of CC-VCSEL

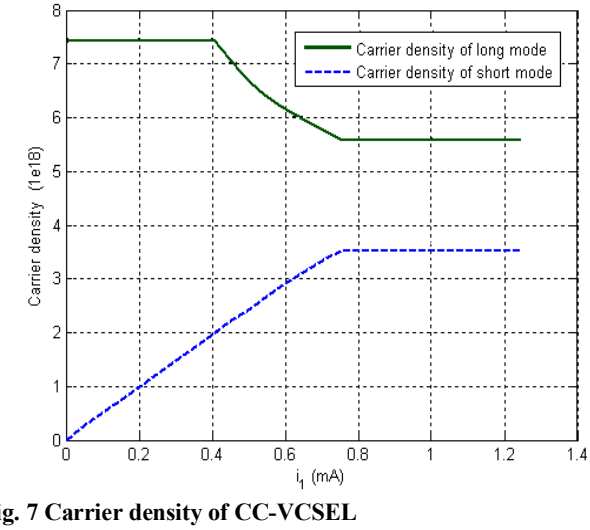


Fig. 7 Carrier density of CC-VCSEL

References

[1] T. Marozsak et al., "Direct modulation in radio over fiber applications," in Tech. Dig. Int. Topical Modelling microwave photonics. 2002, pp 129-132.

[2] C. Carlos et al, "Analog modulation properties of oxide confined VCSEL's at microwave frequency", *J. Lightwave Technol.* vol. 20, pp. 1740-1749, 2002.

[3] V. Badilita, J. Francois, M. Llegems, K. Panajotov, " Rate equation model for coupled-cavity surface emitting lasers," *IEEE J. Quantum Electron*., vol. 40, pp.1664-1656, 2004.

[4] R. P. Stanley, R. Hudre, U. Oesterle, M. Llegems and C. Weisbuch, "Coupled semiconductor micro-cavities", *Appl. Phys. Lett*., vol. 65, pp. 2093-2095, 1994.

[5] P. Pilandini, R. P. Stanley, R. Hudre, U. Oesterle, M. Llegems and C. Wcisbuch, "Dual-wavelength emission from coupled semiconductor micro-cavity," , *Appl. Phys. Lett*., vol. 71, pp. 864-866,1997.

[6] M. Brunner and K. Culden, R. Hvel, J. F. carlin, M. Llegems, "Continuous wave dual wavelength lasing in two section vertical cavity laser," *IEEE Photon. Tech. Lett.,* vol. 75, pp. 3020-3022, 1999.

[7] J. F. Carlin, R. P. Stanley, P. Pellandini, U. Oestle and M. Llegems, "The dual wavelength bi-vertical cavity surface emitting lasers," *App. Phys. Lett.,* vol. 75, pp. 908-910, 2000.

[8] D. M. Grasso and K. D. Choquette, "Threshold and modal characteristics of composite-resonator vertical cavity lasers," *IEEE J. Quantum Electron.,* vol. 39, pp. 1526-1530, Dec. 2003.

[9] J. Marozsak, "Circuit model for multiple transfer model vertical surface-emitting Lasers," *IEEE J. Lightwave Technol*. vol. 21, No. 12, pp. 2977-2982, 2003.

[10] D. M. Grasso, D. K. Serkland, G. M. Peake, "Direct modulation characteristics of composite resonator Vertical-cavity lasers," *IEEE J. Quantum Electron*., vol. 42, pp. 1248-1254, 2006.

[11] P. V. Mena, J. Marikuni, S. M. kang fellow, A. V. Harton K. W. Wyatt, " a comprehensive circuit level model vertical cavity surface emitting laser," *IEEE Lightwave Technol.* vol. 12, pp. 2612-2632,1999.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Optimum Design of a Dispersion Managed Photonic Crystal Fiber for Nonlinear Optics Applications in Telecom Systems

S. M. Abdur Razzak[1], Muhammad Abdul Goffar Khan[2], Yoshinori Namihira[1], and Md. Yeakub Hussain[3]

[1]Department of Electrical and Electronics Engineering, University of the Ryukyus
1 Senbaru, Nishihara, Okinawa 903-0213, Japan
E-mail: razzak91@yahoo.com

[2]Department of Electrical and Electronic Engineering, Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
E-mail: qmagk@yahoo.com

[3]Department of Electronics and Communication Engineering, Manarat International University
Dhaka, Bangladesh
E-mail: yeakub@manarat.ac.bd

Abstract – This paper presents an optimum design for highly nonlinear dispersion managed photonic crystal fibers. The APSS version 2.3 software is used as the simulation tool. According to simulation, an eight-ringed photonic crystal fiber can be designed with a high nonlinear coefficient at 1550 nm of the order 36.5 $W^{-1}km^{-1}$ with simultaneously flatter dispersion characteristics and low confinement losses. This fiber also assumes a high birefringence and has a modest number of design parameters.

I. Introduction

Photonic crystal fibers (PCFs) [1] have claddings that contain tiny air-holes in a pure silica background. Such a cladding concept provides flexibility to design a wide range of index contrast between the core and the cladding. This novel property of PCFs helps in tuning transmission characteristics namely dispersion, nonlinearity, and birefringence in smart ways. As a result PCFs are finding applications nowadays in different areas of communication systems [2]. Such fibers are therefore, suitable for both linear and nonlinear optics applications [2]. This is due to the fact that in these fibers nonlinearity can be controlled suitably with less design efforts to design very high nonlinear coefficient. PCFs having very high nonlinear coefficient are suitable for various applications in nonlinear optics for example, all-optical signal processing, wavelength conversion, ultra-short soliton pulse transmission, optical parametric amplification, and supercontinuum generation [3]. In such applications, control of chromatic dispersion keeping a low confinement loss is crucial for ensuring stable operation of the system [4, 5]. Many PCFs design exists in the literature with remarkable dispersion and leakage properties [4]-[6] but nonlinear coefficient of these PCFs are often less than a 16 $W^{-1}km^{-1}$. HNL-PCFs with identical air-holes [7, 8] have also been reported, however, such PCFs with small core and large air-holes tend to shift zero dispersion wavelengths towards shorter wavelength [9]. Therefore, identical air-holes PCFs are not suitable for designing highly nonlinear PCFs (HNL-PCFs) for the telecom window. Saitoh *et al.* [9] proposed a HNL fiber having nonlinear coefficient of order 30 $W^{-1}km^{-1}$, although it shows significant increment in the nonlinear coefficient, many design parameters i.e., ten rings, and five different air-hole diameters impose fabrication challenges.

We, in this paper, propose a PCF structure which is an optimized model of the HNL-PCF reported by Saitoh *et al.* [9]. We successfully have shown that a nonlinear coefficient of the order 36.5 $W^{-1}km^{-1}$ can be achieved from an eight-ring PCF with only three air-hole dimensions. It can operate effectively as a single mode fiber with ultra-flattened chromatic dispersion of 0 - 0.25 ps/(nm-km) in a 200 nm bandwidth around 1550 nm and low confinement losses less than 0.02 dB/km. Thus the proposed PCF is attractive for nonlinear regime applications because of both its novel properties and optimized design.

II. Design Methodology

Figure 1 shows geometry of the HNL-PCF. The air-hole diameters on the first ring is d_1, while air-hole diameters on the second to 4th ring is d_2, and diameter of air-holes on other rings are d_3. Dimension of the first ring is lowered down to flatter the dispersion characteristics. Although it is not impossible to design dispersion-flattened HNL-PCF with uniform air-hole sized cladding, however, uniform cladding PCFs with small core and large air-holes tend to shift the zero dispersion wavelengths toward shorter wavelengths. Because of this difficulty we have modulated air-holes so as to achieve flatter dispersion. We numerically show the possibility to design a relatively simple PCF structure with fewer parameters for the

telecom window without distorting the dispersion flatness, low confinement loss, and fiber nonlinearity.

In Fig. 1 the background material is pure silica, air-holes in the cladding are arranged in a hexagonal rotational symmetry. Air-hole to air-hole distance is the pitch which is denoted by the symbol Λ. The first ring i.e., the inner most ring contains six air-hole and the other rings contains integer multiple of six air-holes.

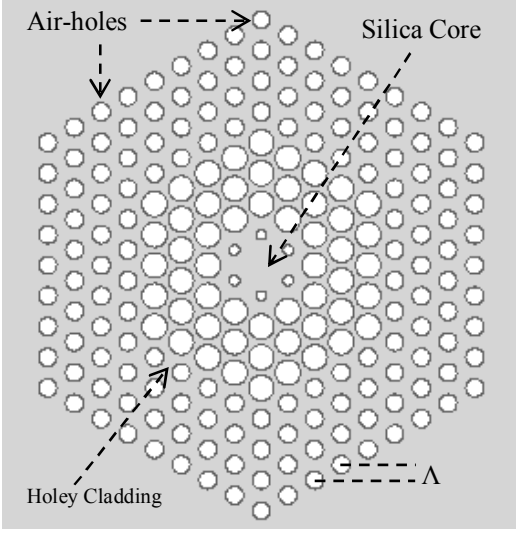


Fig. 1 Geometry of proposed HNL-PCF with 1st ring's air-hole diameter d_1, 2nd to 4th rings' air-hole diameters d_2, and diameter of air-holes on 5th to 8th rings' is d_3.

III. Simulation Methodology

The APSS software 2.3 version is used as a simulation tool. The computational window was surrounded by eight perfectly matched layers (PMLs) boundary which is considered the most efficient boundary conditions for the PCF simulation. Once the modal effective refractive index, n_{eff} is obtained by solving an eigenvalue problem drawn from Maxwell equations using the APSS 2.3, chromatic dispersion D can be obtained using the following relation [10]-

$$D = -\frac{\lambda}{c}\frac{d^2\,\mathrm{Re}[n_{eff}]}{d\lambda^2} \qquad (1)$$

where, Re[n_{eff}] is the real part of n_{eff}, λ is the wavelength, and c is the velocity of light in vacuum. The material dispersion given by Sellmeier formula is directly included in the calculation. Therefore, D in (1) corresponds to the total dispersion of the PCF.

Confinement losses of PCFs are often computed using the formula [10]-

$$Lc = 8.686 \times \mathrm{Im}[k_0 n_{eff}] \times 10^3 \text{ dB/km} \qquad (2)$$

where, Im[n_{eff}] is the imaginary part of n_{eff},, and k_0 is the free space wave number equal to $2\pi/\lambda$.

Finally the effective area A_{eff} and nonlinear coefficient are calculated by the following equations (3) and (4) respectively [11, 12]-

$$A_{eff} = \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|E|^2\,dxdy\right)^2 \Big/ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|E|^4\,dxdy \qquad (3)$$

$$\gamma = (2\pi n_2 / \lambda A_{eff}) \times 10^3 \quad W^{-1}km^{-1} \qquad (4)$$

where, E is the electric field derived by solving the Maxwell equations, λ is the wavelength, γ is the nonlinear coefficient, and n_2 is the nonlinear refractive index.

IV. IV Simulation Results

An optimum dispersion curve of the proposed HNL-PCF is shown in Fig. 2. Optimizing the parameters, ultra-flattened dispersion of 0 - 0.25 ps/(nm-km) is obtained in a 1400 to 1600 nm wavelength range for d_1/Λ = 0.34, d_2/Λ = 0.82, d_3/Λ = 0.58, Λ = 0.91 μm, and number of rings Nr = 8. Dispersion at 1550 μm is as low as 0.10 ps/(nm-km). Dispersion tolerance of the proposed HNL-PCF due to fiber's global diameter variation (d, Λ) of order ±1% is also investigated. Dispersion tolerance study is important because during fabrication process ±1% variation occurs in fiber global diameter [13]. It is found that the influence of fiber global diameter variation of order ±1% causes dispersion to change about ±2.0 ps/(nm-km) which is satisfactory [6].

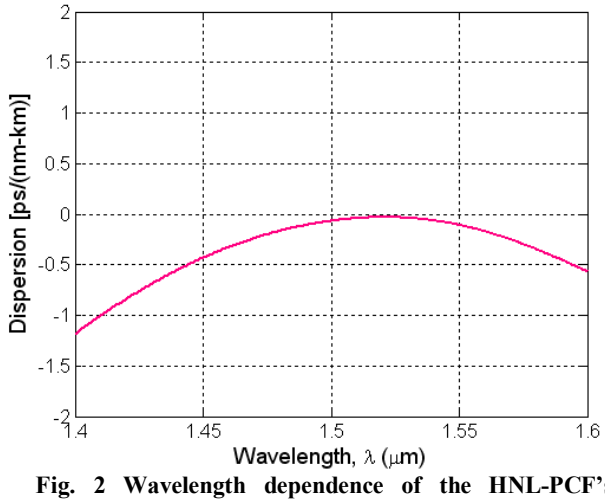


Fig. 2 Wavelength dependence of the HNL-PCF's chromatic dispersion for optimum parameters: d_1/Λ = 0.34, d_2/Λ = 0.82, d_3/Λ = 0.58, Λ = 0.91 μm, and Nr = 8.

Figure 3 shows wavelength dependence of confinement loss of the HNL-PCF for optimum parameters mentioned earlier. The confinement loss is found to be less than a 0.02 dB/km within the band of interest. This is due to the fact that dimension of the first ring is scaled down to shape dispersion characteristics while dimension of the other rings are kept larger to reduce confinement losses. Figure 3 also shows that effective area of the proposed fiber is 2.62 μm^2 at 1550 nm wavelength. The nonlinear co-efficient corresponding to the effective area 2.62 μm^2 is about 36.5 W^{-1}Km^{-1} as is shown in Fig. 4. It should be pointed out that in short wavelengths this fiber supports a second order mode. But confinement loss of the second order mode at 1550 nm wavelengths is more than a 16 dB/km. Therefore, the proposed fiber will effectively operate as a single mode

fiber within the entire band of interest. Figure 4 also shows that the fiber can assume a birefringence of the order 6.7×10^{-4} at 1550 nm which is higher than the conventional polarization maintaining (PM) fibers. This value of birefringence is useful for usage of nonlinear fibers in supercontinuum generation.

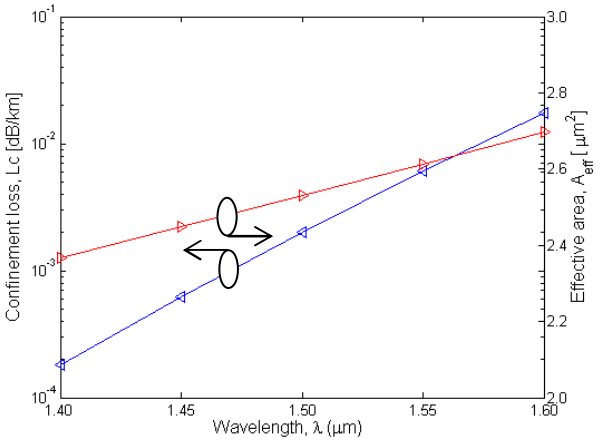


Fig. 3 Wavelength dependence of confinement loss and effective area of proposed HNL-PCF at 1550 nm wavelength for optimum design parameters: d_1/Λ = 0.34, d_2/Λ = 0.82, d_3/Λ = 0.58, Λ = 0.91 μm, and Nr = 8.

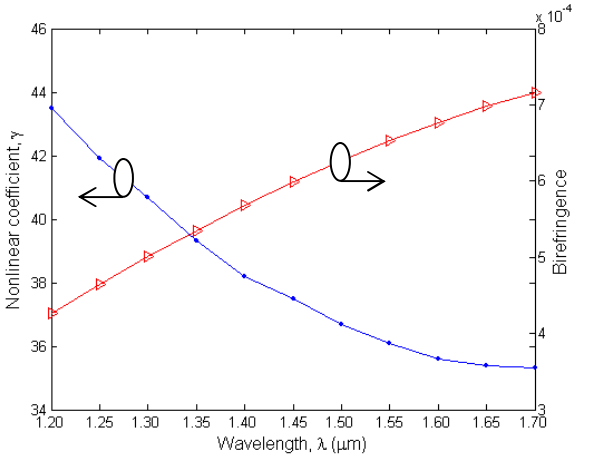


Fig.4. Nonlinear coefficient and birefringence of the HNL-PCF for the optimum design parameters: d_1/Λ = 0.34, d_2/Λ = 0.82, d_3/Λ = 0.58, Λ = 0.91 μm, and Nr = 8.

Figure 5 shows wavelength dependence of mode field diameter (MFD) and splice loss between this fiber and conventional single mode fibers (SMFs). The MFD of the SMF is considered 10.0 μm. The MFD is calculated by the well known Pitermann II formula [14], and the splice loss *Ls* is calculated by [15]-

$$L_s = -20\log_{10}\frac{2w_{SMF}w_{PCF}}{w_{SMF}^2 + w_{PCF}^2} \qquad (5)$$

in decibel. Where, w_{SMF} and w_{PCF} are the MFDs of the SMF and the HNL-PCF respectively.

Because of smaller MFD of proposed fiber, splice losses are generally higher. This is also true for all highly nonlinear fibers. This high splicing loss can be eliminated by the use of recent splice free interconnection technique between SMFs and PCFs [16].

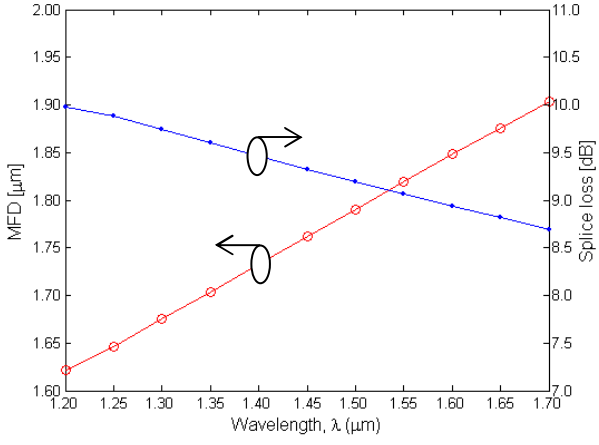
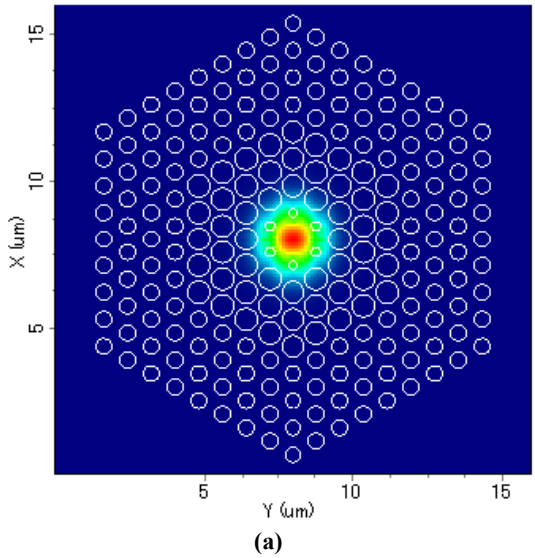
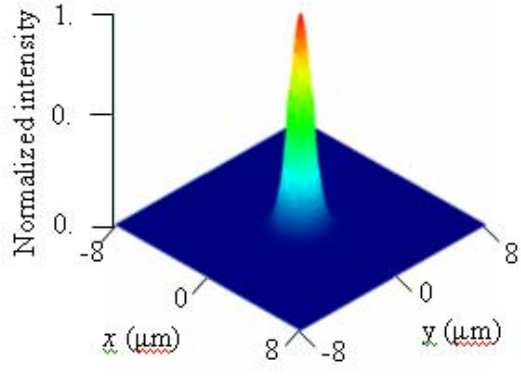


Fig.5. Splice loss and MFD of the HNL-PCF for the optimum design parameters: d_1/Λ = 0.34, d_2/Λ = 0.82, d_3/Λ = 0.58, Λ = 0.91 μm, and Nr = 8.



(a)



(b)

Fig. 6 Mode field pattern of the fundamental mode at 1550 nm; (a) 2D and (b) 3D distribution. Red color represents the highest intensity and blue the lowest.

Figure 6 shows mode field intensity distribution of the fiber at 1550 nm wavelength corresponding to the optimum design. Both the 2-D and 3-D plots justify that the field has confined well in the core as there are no evidences of light leakage into the cladding region beyond the first ring.

Therefore, in the light of above investigation two major advantages of the proposed design are apparent. Firstly, design parameters are reasonably a modest number i.e., three air-hole dimensions, eight rings, and a pitch and secondly, it has a birefringence higher than that of conventional PM fibers. Thus it can be concluded that the proposed HNL-PCF may be a suitable candidate for various applications in nonlinear optics because of the high nonlinear coefficient, dispersion-flat characteristics, and low confinement losses.

V. Discussion

One important issue to consider for the PCFs is the fabrication process. For highly nonlinear PCF high degree of fabrication accuracy is required because a small change in dimension may cause some important properties to drift from the designed value. Therefore, HNL-PCFs with standard uniformity have proven to be extremely difficult to manufacture. Although the well known stack and draw method can fabricate PCFs of almost any structure, increased design complexity limits the accuracy of such fabrication. The US Naval Research Laboratory has a new approach [17] to fabricating high-quality preforms used in PCFs production that can maintain uniformity, roundness and accuracy in air-hole dimensions. PCFs fabricated from such high-quality preforms have improved accuracy in comparison to PCFs fabricated using other conventional methods. We believe that using such a method it is possible to realize the HNL-PCF with standard accuracy.

VI. Conclusion

A relatively simple highly nonlinear PCF with flatter chromatic dispersion and low confinement loss has been reported. It has been shown through numerical simulation results that using an eight-ring PCF, ultra-flattened dispersion of 0 - 0.25 ps/(nm-km) can be obtained within a 1400 to 1600 nm wavelength range (200 nm band). The confinement loss is as low as a 0.02 dB/km within the entire band of interest.

References

[1] J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkins, "All-silica single-mode optical fiber with photonic crystal cladding," Opt. Lett., vol. 21, pp.1547-1549, 1996.

[2] A. Bjarklev, J. Broeng, and A. S. Bjarklev, Photonic Crystal Fibres, Kluwer Academic Publishers, 2003.

[3] S. M. A. Razzak and Y. Namihira, "Proposal for highly nonlinear dispersion-flattened octagonal photonic crystal fibers," IEEE Photonics Technol. Lett., vol. 20, no. 4, pp. 249-251, Feb. 2008.

[4] T-L. Wu, and C. H. Chao, "A Novel Ultraflattened Dispersion Photonic Crystal Fiber," IEEE. Photonics Technol. Lett., vol. 17, no. 1, pp.67-69, January 2005.

[5] A. Ferrando, E. Silvestre, J. J. Miret, and P. Andres, "Nearly zero ultraflattened dispersion in photonic crystal fibers," Opt. Lett. vol. 25, no.11, pp.790-792, June 2000.

[6] K. Saitoh, N. J. Florous, and M. Koshiba, "Ultra-flattened chromatic dispersion controllability using a defected-core photonic crystal fiber with low confinement losses," Opt. Express, vol. 13, no. 21, pp. 8365-8371, October 2005.

[7] V. Finazzi, T. M. Monro, and D. J. Richardson, "Small-core holey fibers: nonlinearity and confinement loss trade-offs," J. Opt. Soc. Am. B, vol. 20, pp.1427-1436, 2003.

[8] T. Yamamoto, H. Kubota, S. Kawanishi, M. Tanaka, and S. Yamaguchi, "Supercontinuum generation at 1.55 μm in a dispersion-flattened polarization-maintaining photonic crystal fiber," Opt. Express, vol. 11, pp. 1537-1540, 2003.

[9] K. Saitoh, M. Koshiba, T. Hasegawa, and E. Sasaoka, "Highly nonlinear dispersion-flattened photonic crystal fibers for supercontinuum generation in a telecommunication window," Opt. Express, vol. 11, pp.843-852, 2004.

[10] K. Saitoh, M. Koshiba, T. Hasegawa, E. Sasaoka, "Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion," Opt. Express, vol. 11, no. 8, pp.843-852, April 2003.

[11] S. M. A. Razzak, Y. Namihira, F. Begum, S. Kaijage, N. H. Hai, and N. Zou, "Design of a decagonal photonic crystal fiber with ultra-flattened chromatic dispersion," IEICE Trans. Electron., vol. E90-C, no. 11, pp. 2141-2145, Nov. 2007.

[12] S. M. A. Razzak and Y. Namihira, "Proposal for highly nonlinear dispersion-flattened octagonal photonic crystal fibers," IEEE Photonics Technol. Lett., vol. 20, no. 4, pp. 249-251, Feb. 2008.

[13] W. H. Reeves, J. C. Knight, and P.St.J. Russell: "Demnostration of ultra-flattened dispersion in photonic crystal fibers," Opt. Express, vol.10, no.14, pp. 609-613, July 2002

[14] K. Petermann, "Contrainsts for fundamental-mode spot size for broadband dispersion-compensated single-mode fibers," Electron. Lett., vol. 19, pp. 712-714, September 1983.

[15] S. M. A Razzak and Y. Namihira, "Tailoring dispersion and confinement losses of photonic crystal fibers using hybrid cladding," to be published in the IEEE/OSA J. of Lightwave Technol., 2008.

[16] S. G. Leon-Saval, T. A. Birks, N. Y. Joly, A. K. George, W. J. Wadsworth, G. Kakarantzas, and P. St. J. Russell, "Splice-free interfacing of photonic crystal fibers," *Optics Letters*, Vol. 30, No. 13, pp. 1629-1631, 2005.

[17] U. S. Naval Research Laboratory website- http://www.nrl.navy.mil/

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Dispersion Analysis of Photonic Crystal Fiber

M. Jalal Uddin[1], Iftekhar A. Khan[2], and M. Shah Alam[3]

[1, 3] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET)
Dhaka – 1000, Bangladesh
Email: jalaluddin@eee.buet.ac.bd, shalam@eee.buet.ac.bd

[2]Department of Computer Science and Engineering, International Islamic University Chittagong, Dhaka Campus
Dhaka – 1205, Bangladesh

Abstract – A detailed modal analysis of photonic crystal fiber (PCF) is carried out by using the finite element method (FEM) in FEMLAB environment. The effective mode index of the fundamental mode is found out and the effective mode area is calculated for different PCF structures. Finally, dispersion property of different PCF structures is numerically calculated and it is found that the dispersion profile of the PCF can be controlled by appropriately choosing diameter and pitch of air holes.

I. Introduction

Broadband optical transmission with wavelength division multiplexing technique is effective for large capacity networks, and optical fibers are widely used as optical transmission media. However chromatic and modal dispersion, nonlinearity and losses restrict the wavelength region available for these fibers.

Photonic crystal fibers (PCFs) recently attracted a great deal of interest because of their excellent propagation properties [1], [2]. Many research groups all over the world are making constant effort to establish the superiority of PCFs over conventional fibers because of its novel optical characteristics [3]–[10]. It has been reported that PCF can realize endlessly single-mode operation [3], flexible chromatic dispersion over a wide wavelength range [4]–[6], large effective area [6]–[7], controllable nonlinearity [7], ultralow loss [8], [10] and high group birefringence [9]. Basically PCFs are single-material fibers with an arrangement of air holes running along the length of the fiber, and they provide confinement and guidance of light in a defect region around the centre. As for the light confinement mechanism, index guiding PCFs rely on total internal reflection to confine light in the region of a missing air hole forming a central core.

Various methods have been used for analysis of PCFs [11], [12] each of which has its advantages and limitations. In this work, the FEM is used to find modal and dispersion characteristics of PCFs in FEMLAB environment. The effect of structural parameters on modal and dispersion properties is studied in detail. The results are compared with the existing published results and found very good agreement [4] – [6] and [8].

II. Analysis Method

The FEMLAB based finite element technique is used in this work for the analysis of dispersion properties of PCF. Fig. 1 shows the transverse cross section of a typical PCF consists of a central high index defect region or a missing hole in a regular triangular or hexagonal array of air holes. There are three design parameters: the air hole diameter d, the spacing between holes or pitch Λ, and the number of air hole rings N_r.

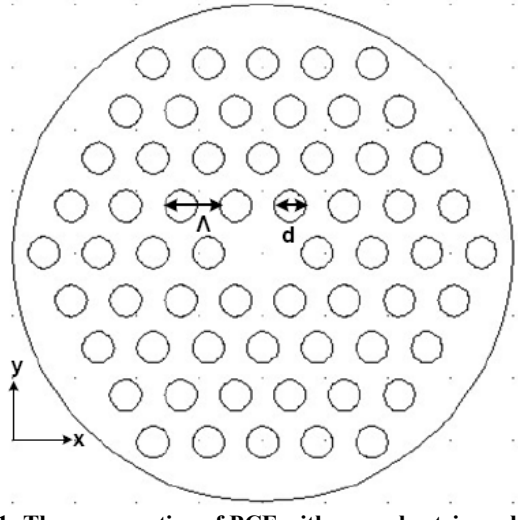


Fig. 1: The cross section of PCF with a regular triangular air hole array defined by the air hole diameter d and the pitch Λ.

The modal analysis is carried out first on the cross section in the x-y plane of the fiber, i.e., on the cross section of the fiber as shown in Fig. 1. It is assumed that the wave propagates in the z-direction and has the form,

$$\mathbf{H}(x, y, z, t) = \mathbf{H}(x, y, z)\exp[j(\omega t - \beta z)], \qquad (1)$$

where ω is the angular frequency and β is the propagation constant. An eigenvalue equation in terms of the magnetic field $\mathbf{H}$ is derived from the Helmholtz equation,

$$\nabla \times (n^{-2}\nabla \times \mathbf{H}) - k_0{}^2\mathbf{H} = 0, \qquad (2)$$

and is solved for the eigenvalue of $-\beta^2$.

As for the boundary condition along the outside of the cladding boundary, the magnetic field is set to zero considering perfect magnetic conduction (PMC) of the boundary where,

$$\mathbf{a_n} \times \mathbf{H} = 0 . \quad (3)$$

Here, $\mathbf{a_n}$ is the unit normal on the surface. The inner air hole and fiber material should maintain a boundary condition for continuity of field.

In FEMLAB, however, modal analysis of a perpendicular hybrid-mode wave using electromagnetics module is done. The eigenvalue equation is solved to give effective mode index,

$$n_{eff} = \frac{\beta}{k_0}, \quad (4)$$

of a guided mode for a given wavelength. Here, k_0 is the free space wavenumber. Once the modal solution is obtained, various post-processing data is readily available for visualization of the solution in different ways.

The effective mode index has both real and imaginary parts. The chromatic dispersion of PCF is calculated easily [4] from the real part of the effective mode index n_{eff}, values versus wavelength using,

$$D = -\frac{\lambda}{c}\frac{d^2 \,\mathrm{Re}[n_{eff}]}{\lambda^2}, \quad (5)$$

where, λ is the wavelength and c is the velocity of light. In order to check the accuracy of the results found here, we considered an index guiding PCF with a single core of six equally spaced air holes with hole diameter d = 5 μm, pitch Λ = 6.75 μm, the background refractive index n = 1.45, and the operating wavelength, λ =1.45 μm. The effective index found here, n_{eff} = 1.445395 + i 1.807358×10^{-19}. This result is in good agreement with that of the full vector FEM [4] except that of the imaginary part of the effective mode index. The limitation of finite element method in calculating the imaginary part of the effective index is stated in [13]. As dispersion depends only on the real part of the n_{eff}, we concentrate our focus on the dispersion properties of PCF, varying different structural parameters.

Another important property of the fiber, the effective mode area, A_{eff} can be calculated [4] using the electric field as,

$$A_{eff} = \frac{\left(\iint |E|^2 \, dxdy\right)^2}{\iint |E|^4 \, dxdy}. \quad (6)$$

The area integration is carried out over the cross section of the fiber and the electric field intensity E is given by,

$$E = \sqrt{E_x^2 + E_y^2}, \quad (7)$$

where, E_x is the x component of the electric field and E_y is the y component of the electric field.

III. Results and Discussion

The effective mode index is found out for each operating wavelength and for different PCF structures. Fig. 2 shows the plot of effective mode index versus wavelength for different structural parameters.

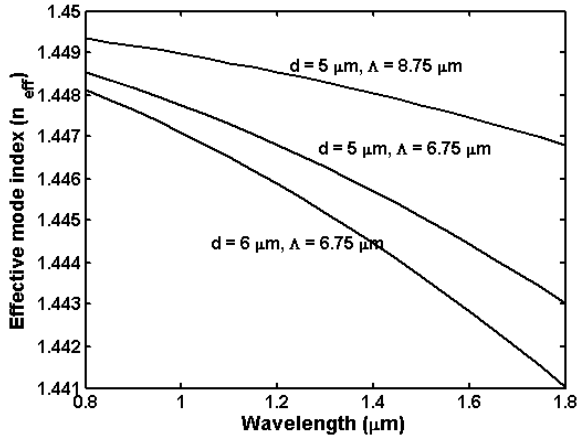


Fig. 2: Effective index as a function of wavelength. The variation of effective index with the air hole diameter, d and pitch, Λ is also shown in the figure. One air hole ring is use in each case.

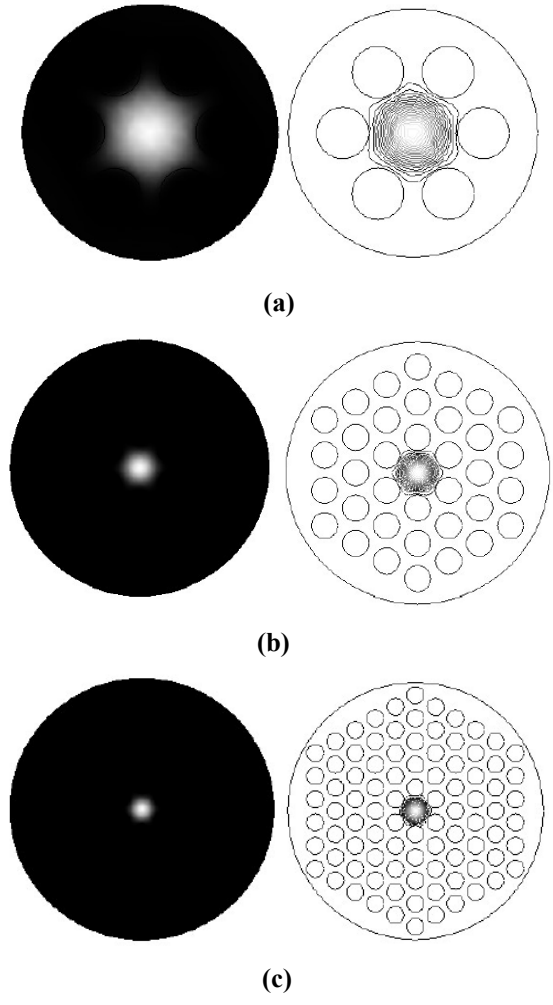


Fig. 3: Two dimensional surface field distribution and contour plot. (a) for one air hole ring, (b) for 3 air hole ring and (c) for 5 air hole rings.

Table 1: Change of effective mode index with wavelength, diameter and pitch.

λ (μm)	d (μm)	Λ (μm)	Re (n_{eff})
0.8	5	6.75	1.448546
1.2	5	6.75	1.446801
0.8	6	6.75	1.448122
0.8	5	8.75	1.449345

It is found that the effective mode index decreases with the increase in wavelength which is in good agreement with [6]. It is also found that the effective mode index increases with the decrease of air hole diameter and with the increase of pitch. Table 1 shows the relative change of effective mode index with the variation of structural parameters.

In Fig. 3, two dimensional surface plots and contour plots for the electric field distribution are presented for different number of air hole rings. Fig. 3(a) shows the plot for one surrounding air hole ring, Fig. 3(b) for three air hole rings and Fig. 3(c) for five air hole rings. It is clear from the plots that the propagating light becomes more concentrated in the centre, decreasing the spot size when the number of air hole rings is increased.

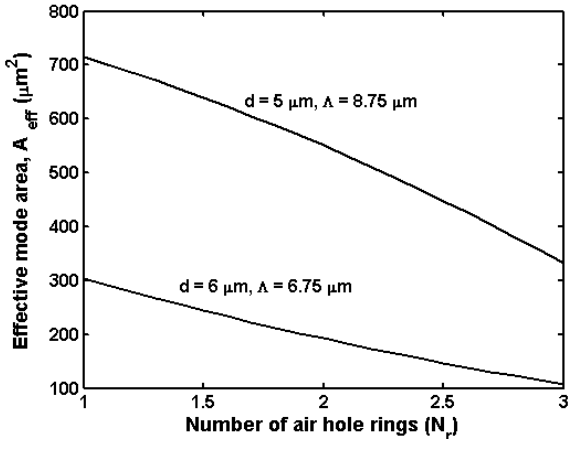


Fig. 4: Effective mode area as a function of number of air hole rings, Nr. The variation of effective mode area with the air hole diameter, d and air hole spacing, Λ is also shown in the plot.

Table 2: Change of effective mode area with diameter, pitch and number of air hole rings.

N_r	d (μm)	Λ (μm)	A_{eff}(μm^2)
1	6	6.75	300
3	6	6.75	110
3	5	8.75	330

Fig. 4 shows the variation of effective mode area against the number of air hole rings. It is clear that the PCF exhibits a very large effective mode area. It is found that the effective mode area decreases with the increased number of air hole rings. It is also found that when the diameter is decreased and pitch is increased, the effective mode area is increased. Table 2 shows the comparative change of effective mode area with the variation of structural parameters.

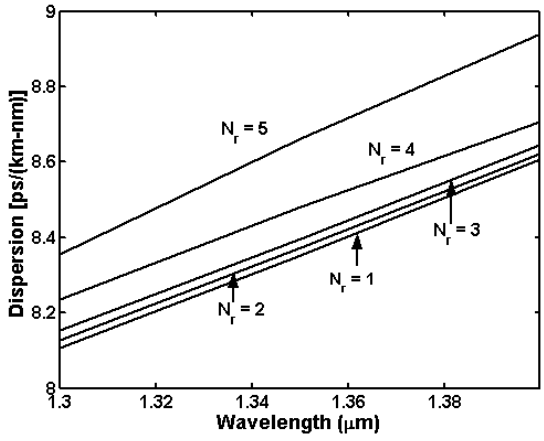


Fig. 5: Dispersion of photonic crystal fiber against wavelength for different number of air hole rings. Air hole diameter, d = 5 μm and hole spacing, Λ = 6.75 μm.

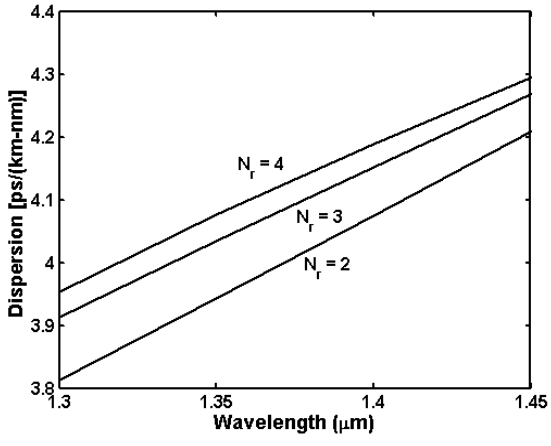


Fig. 6: Dispersion of photonic crystal fiber against wavelength for different number of air hole rings. Air hole diameter, d = 5 μm and hole spacing, Λ = 8.75 μm.

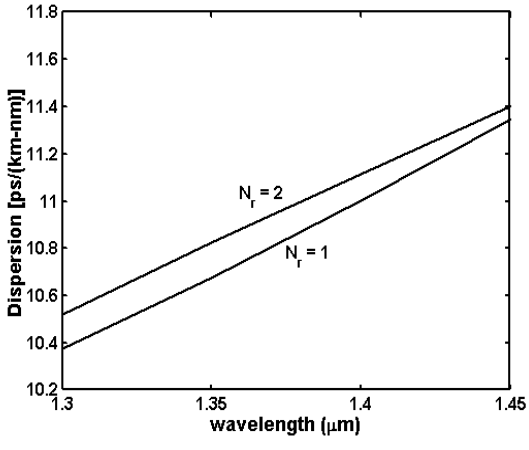


Fig. 7: Dispersion of photonic crystal fiber against wavelength for different number of air hole rings. Air hole diameter, d = 6 μm and hole spacing, Λ = 6.75 μm.

Figs. 5–7 depict the dispersion property of photonic crystal fiber against wavelength for different structural parameters. In each case it is found that the dispersion increases with the increase of operating wavelength which agrees with the results found in [4] and [8], and the

amount of increment is small compared to that of the conventional optical fiber, resulting almost flattened dispersion profile. It is also found that the dispersion increases with the increasing number of air hole rings.

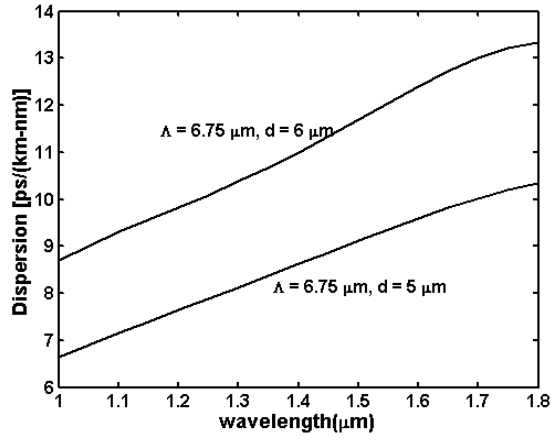


Fig. 8: Dispersion of photonic crystal fiber against wavelength for different air hole diameter. Air hole spacing, Λ and number of air hole rings, N_r is constant.

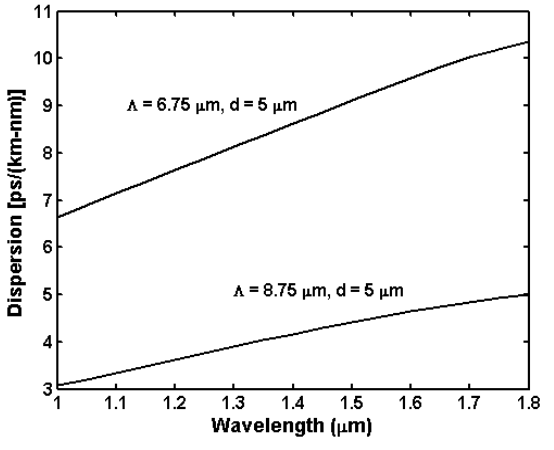


Fig. 9: Dispersion of photonic crystal fiber against wavelength for different air hole spacing. Air hole diameter, d and number of air hole rings, N_r is constant.

Table 3: Change of dispersion, D with wavelength, air hole diameter, pitch and number of air hole rings.

λ (μm)	d (μm)	Λ (μm)	N_r	D [ps/ km-nm)]
1.3	5	6.75	2	8.17
1.4	5	6.75	2	8.6
1.3	6	6.75	2	8.53
1.3	5	8.75	2	3.82
1.3	5	6.75	5	8.37

Fig. 8 depicts the dispersion property of photonic crystal fibers against wavelength for different air hole diameters. It is found that the dispersion increases with the increase in air hole diameter that agrees with the results found in [5]. Finally, Fig. 9 depicts the dispersion property of photonic crystal fiber against wavelength for different pitch. It is found that the dispersion decreases with the increase in pitch. Table 3 compares the dispersion property of PCFs for different structural parameters and wavelengths.

IV. Conclusion

An efficient approach is used for the analysis of PCF by using the finite element method in FEMLAB environment. It has been found that the effective mode index decreases with the increase of wavelength and air hole diameter and increases with the increase of pitch of PCF. It has been observed that the electric field distribution becomes more concentrated to the centre when the number of air hole rings is increased. The effective modal area of the PCF has been found to be large compared to that of the conventional optical fiber and it decreases with the increase in number of air hole rings. Finally, dispersion is calculated for different structural parameters. It has been found that dispersion increases with the increase in wavelength, number of air hole rings and air hole diameter respectively and that decreases with the increase in pitch length.

References

[1] P. Russel, "Photonic crystal fibers," *Science*, vol 299, pp. 358-362, 17 January, 2003.

[2] J. C. Knight, "Photonic crystal fibers," *Nature*, vol. 424, pp. 847-851, 14 August, 2003.

[3] T. Birks, J. Knight and P. Russel, "Endlessly single mode photonic crystal fiber," *Opt. Lett*., vol. 22, no. 13, pp. 961-963, July 1999.

[4] K. Saitoh, M. Koshiba, "Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion," *Opt. Express*, vol. 11, no. 8, pp. 843-852, 21 April, 2003.

[5] L. Shen, W. Huang and S. Jian, "Design of photonic crystal fibers for dispersion related applications," *J. Lightwave Technol.,* vol. 21, no. 7, pp. 1644-1651, 7 July, 2003.

[6] T. Matsui, J. Zhou, K. Nakajima and I. Sankawa, "Dispersion flattened photonic crystal fiber with large effective area and low confinement loss," *J. Lightwave Technol*., vol. 23, no. 12, pp. 4178-4183, 12 Decmber, 2005.

[7] J. Knight, T. Birks, R. Cregon, P. Russel and J. Sandro, "Large mode area photonic crystal fiber," *Electron. Lett.,* vol. 34, no. 13, pp. 1347-1348, June 1998.

[8] K. Tajima, J. Zhou, K. Nakajima and K. Sato, "Ultralow loss and long length photonic crystal fiber," *J. Lightwave Technol.,* vol. 22, no. 1, pp. 7-10, 1 January, 2004.

[9] M. S. Alam, K. Saitoh and M. Koshiba, "High group birefringence in air-core photonic crystal fibers," *Opt. Lett.,* vol. 30, no. 8, pp. 824-826, 15 April, 2005.

[10] V. Finazzi, T. M. Monro and D. J. Richardson, "Small-core silica holey fibers: nonlinearity and confinement loss trade-offs," *J. Opt. Soc. Am. B,* vol. 20, no. 7, pp. 1427-1436, July 2003.

[11] T. P. White *et al,* "Multipole method for microstructured optical fibers," J. Opt. Soc. Am. B, vol. 19, no. 10, pp. 2322-2330, October 2002.

[12] A. Ferrando et al, "Full vector analysis of realistic photonic crystal fiber," *Opt. Lett.,* vol. 24, no. 5, pp.276-278, 1 March, 1999.

[13] F. Zolla, G. Reneversez, A. Nicolet, B. Kuhlmey, S. Guenneau and D. Felbacq, "Foundation of Photonic Crystal Fibers," 1st edition, Imperial College Press, 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

$In_xGa_{1-x}N$ Based Multi Junction Concentrator Solar Cell

Md. Sherajul Islam, A.K.M. Zillur Rahman, Md. A. R. Chowdhury, Md. Rafiqul Islam and Ashraful G. Bhuiyan

Dept. of Electrical & Electronic Engineering, Khulna University of Engineering and Technology (KUET), Khulna-920300, Bangladesh
E-mail: sheraj_ruet @yahoo.com

Abstract - In this paper, the $In_xGa_{1-x}N$ based multi junction solar cell under concentrator has been designed theoretically and the performances are evaluated. The consequence of increasing the concentration ratio up to 500 suns and the number of junctions up to 8 are predicted. A relative comparison of efficiency with and without concentrator has also been studied. The efficiency of $In_xGa_{1-x}N$-based single junction, triple junction, and eight junction solar cells without employing concentrator were found to be 24.67, 37.57, and 46.15 % respectively, while those using concentrator have been improved to 41.07, 46.92 and 50.41% respectively.

I. Introduction

To meet the rising energy demand photovoltaic conversion of solar energy utilizing the semiconductor alloys can play an important role. The great challenges of that type of cells are their high cost per KWh of energy production resulting from costly semiconductor materials and low conversion efficiency. To be competitive with the conventional energy source the cost of the solar cells must be reduced and also the efficiency of that cell must be improved. The efficiency of solar cell has been improved by introducing multi junction schemes. However, still the efficiency is not good enough and the cost of solar cell materials has not been reduced significantly. The latest modification in photovoltaic incorporates the use of concentrators, promising for high power density and potentially reduces the cost of power generation. The concentrator decreases the area of solar cell materials being used in a system which are very expensive. Replacing semiconductor solar cell area with low cost plastic lenses leads to a system with lower overall cost. Also the efficiency of the solar cell can be improved by introducing concentrator.

In the present days, $In_xGa_{1-x}N$ has been proposed as a promising material for the fabrication of high efficiency solar cells. Because the band gap of $In_xGa_{1-x}N$ alloys can be varied continuously from 0.7 to 3.4eV [1], which provides an almost perfect fit to the full solar spectrum. The $In_xGa_{1-x}N$ films show an exceptionally strong and robust photoluminescence, although it grown on lattice mismatched substrates. It is also shown that the optical and electronic properties of the $In_xGa_{1-x}N$ alloys exhibit a much higher resistance to high-energy (2 MeV) photon irradiation than the currently used photovoltaic materials such as GaAs and GaInP, and therefore offer great potential for radiation-hard high-efficiency solar cell for space applications[2-4]. Since only single material system is required the $In_xGa_{1-x}N$-based MJ solar cell will be technologically significant. We have projected the performances of $In_xGa_{1-x}N$-based MJ solar cells. However, the efficiency of $In_xGa_{1-x}N$-based single junction, triple junction, and eight junction solar cells were limited to be 24.67, 37.57, and 46.15 %, respectively [5]. Therefore, in an effort to further increase the efficiency and to reduce the cost of the solar cells $In_xGa_{1-x}N$-based MJ concentrator solar cells have been studied in this work. This includes the theoretical design of $In_xGa_{1-x}N$-based MJ concentrator solar cells. The performances of the device have been evaluated varying the number of junctions up to 8 and concentration ratio up to 500.

II. Theoretical model

Light concentration is one of the important issues for the development of an advanced PV system. In order to attain high efficiency the performances of $In_xGa_{1-x}N$-based MJ solar cells under light concentration have studied. The schematic diagram for a concentrator PV system is shown in Fig. 1. In this work, the $In_xGa_{1-x}N$-based MJ concentrator solar cells have been designed and the performances of the system are investigated theoretically varying the concentration ratio up to 500 suns and number of junctions up to 8. All the sub cells are made with In_xGa_{1-x} N alloys with different composition to obtain the required band gap and arranged from bottom to top with lower to higher band gap. Detail modeling of $In_xGa_{1-x}N$-based solar cells is described in other paper [6]. The total thickness of each sub cell is considered as 2 μm. The model optimizes the efficiency of MJ solar cells for different junctions. The energy gaps of the $In_xGa_{1-x}N$ alloys that should be used for the tandem cells are optimized assuming a perfect quantum response of materials and equal photo current densities of each

junction. The indium fraction was calculated using the relation given in reference [7]. Unless specified all the cells are considered emitter as n and base as p, i.e., n on p. few properties of the identified $In_xGa_{1-x}N$ alloys are not available in literature; in such cases data of GaN is considered.

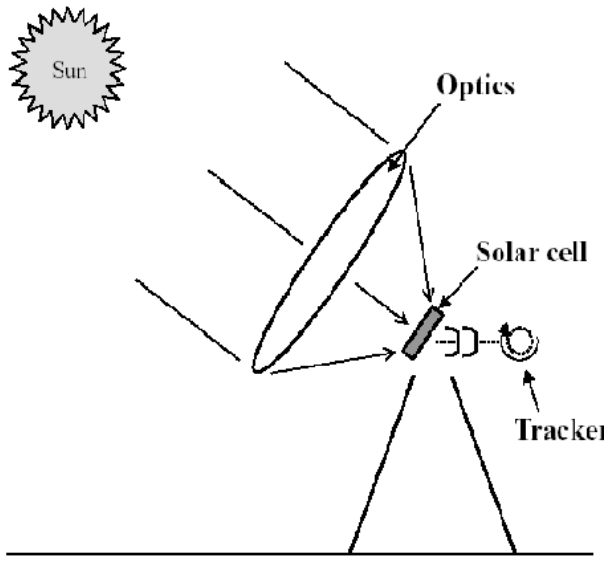


Fig. 1 System configuration of concentrator PV system

The absorption coefficients of the identified $In_xGa_{1-x}N$ alloys are taken from reference. The majority carrier concentration was taken equal to 10^{18} cm^{-3} on each side of the junction. The surface and the rear recombination velocities were taken to be equal to 10^3 cm/s. The carrier diffusion lengths and coefficients of the identified $In_xGa_{1-x}N$ alloys are assumed to be same in all junctions and are supposed to be equal to those of the GaN.

III. Results and discussions

For efficiency maximization, the short circuit current and open circuit voltages are calculated for different concentration ratio. Figure 2 shows the effect of concentrator on short-circuit current density for different number of junctions. The short circuit current density increases linearly with the increase of concentration ratio because of the increase in number of photons. On the other hand, short circuit current density decreases as the number of junction increase which results from increased path resistance. Varying the concentration ratio from 1 to 500 suns the short circuit current density, J_{SC}, is found to be varied from 31.15 to 15575 mA/cm^2, 16.10 to 7895 mA/cm^2, and 6.95 to 3475 mA/cm^2 for single, triple and eight junctions, respectively. Figure 3 shows the effect of concentrator on open- circuit voltages for different number of junctions. The open circuit voltage, $V_{OC,}$ is found to be varied from 0.8988 to 1.4441 V, 2.701 to 3.2541 V and 7.389 to 7.9421 V for single, triple and eight junctions, respectively, varying the concentration ratio from 1 to 500 suns.

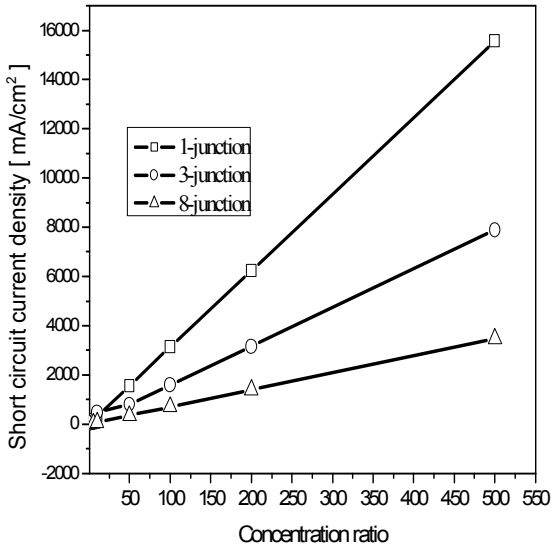


Fig. 2 Short circuit current as a function of concentration ratio.

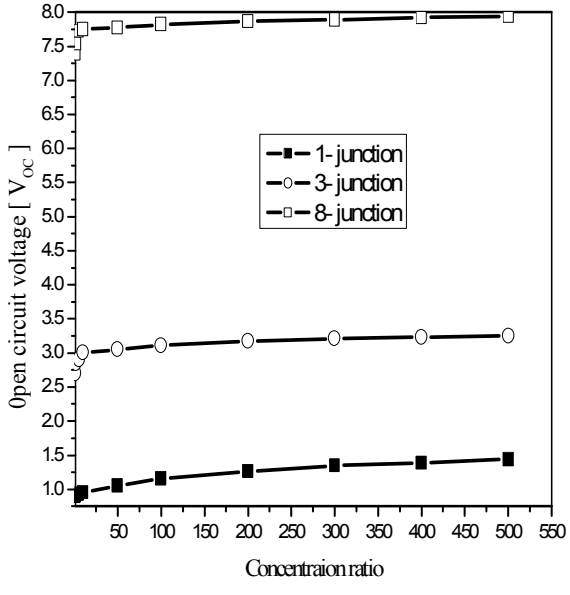


Fig. 3 Variation of open circuit voltage with concentration ratio.

The efficiency of the solar has been calculated under light concentration for different number of junctions. Figure 4 shows the variation of efficiency with concentration ratio for different number of junctions. The efficiency is found to be varied from 24.67 to 41.078 %, 37.57 to 46.92 % and 46.15 to 50.41% for single, triple and eight junctions, respectively, varying concentration ratio from 1 to 500 suns. The short circuit current density J_{SC} increases appreciably with the concentration ratio, on the other hand, the open circuit voltage V_{OC} increases appreciably with the junction.

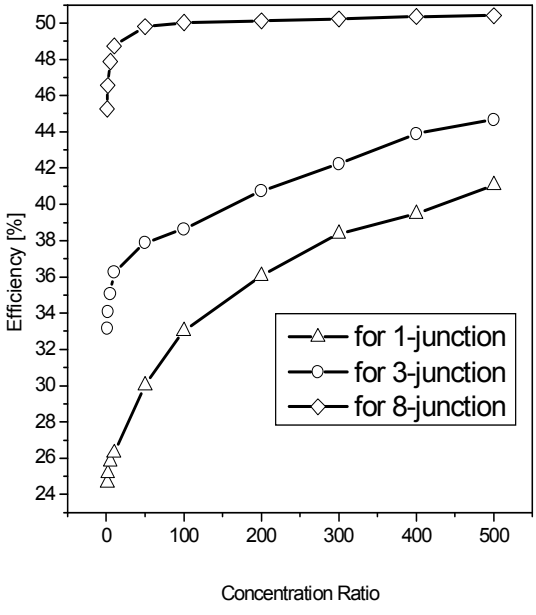


Fig. 4 Variation of efficiency with concentration ratio for different number of junctions

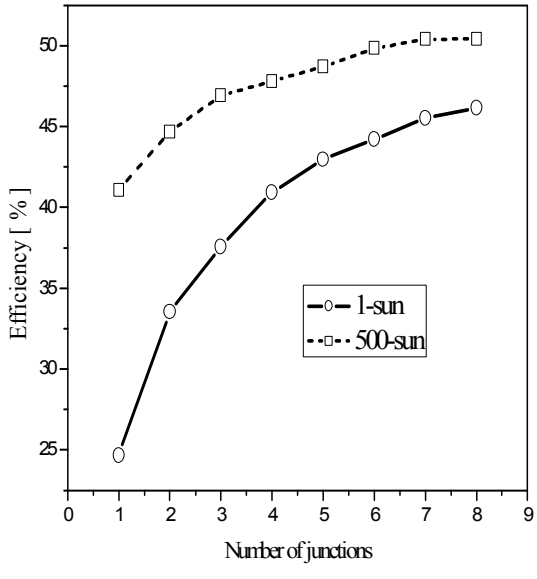


Fig. 5 Variation of efficiency with number of junctions using concentrator (500 suns) and without concentrator

Therefore, the overall efficiency of the solar cells increases by increasing the concentration ratio and junction number. However, the effect of concentration ratio reduces with increasing the number of junction as shown in the Fig. 5. Figure 5 shows the variation of efficiency with number of junctions using concentrator (500 suns) and without concentrator. Table 1 gives the details simulated results of the short circuit current densities, open circuit voltages and efficiencies of $In_xGa_{1-x}N$ based MJ solar cells with and without concentrator comprising single to eight junctions.

Table 1: Simulated results for short circuit current density, open circuit voltage and efficiency using concentration ratio of 500 suns and without concentrator solar cell comprising single to eight
junctions

No. of junction	Short circuit current J_{sc}(mA/cm^2)		Open circuit voltage V_{oc}(V)		Efficiency (%)	
	1-sun	500-sun	1-sun	1-sun	1-sun	500-sun
1	31.15	15575	.8988	1.4441	24.67	41.0784
2	20.47	10235	1.838	2.9301	33.55	44.6779
3	16.10	7895	2.701	3.2541	37.574	46.9215
4	12.18	6086	3.75	4.3051	40.93	47.8052
5	9.79	4895	4.897	5.4481	42.97	48.7064
6	9.50	4750	5.313	5.7451	44.21	49.8402
7	8.15	4070	6.346	6.7791	45.53	50.3912
8	6.95	3475	7.393	7.9421	46.15	50.4056

Simulation results show that for the eight junction's concentrator solar cell an efficiency of more than 50% with a short circuit current density of 3475mA/cm^2 and an open circuit voltage of 7.942V is found.

IV. Conclusion

In conclusion, we have studied $In_xGa_{1-x}N$ based MJ concentrator solar cell as a promising candidate for future high performance solar cells. The performances are evaluated varying the number of junctions up to 8 and concentration ratio up to 500 suns. varying the concentration ratio from 1 to 500 suns, the short circuit current density is found to be varied from 31.15 to 15575 mA/cm^2, 16.10 to 7895 mA/cm^2 and 6.95 to 3475 mA/cm^2 for single, triple, and eight junctions, respectively. The open circuit voltage is found to be varied from 0.8988 to 1.4441 V, 2.701 to 3.2541 V and 7.389 to 7.9421 V for single, triple and eight junctions, respectively, varying the concentration ratio from 1 to 500 suns. The efficiency of $In_xGa_{1-x}N$-based single junction, triple junction, and eight junction solar cells without employing concentrator were found to be 24.67, 37.57, and 46.15 % respectively, while those using concentrator have been improved to 41.07, 46.92 and 50.41% , respectively. The above results show that the $In_xGa_{1-x}N$

based MJ concentrator solar cells have very nice performances for tandem cells applications.

References

[1] J. Wu, W. Walukiewicz, K. M. Yu, W. Shan, J. W. Ager III, E. E. Haller, H. Lu, W. J. Schaff, W. K. Metzger and S. Kurtz, "Superior radiation resistance of $In_{1-x}Ga_xN$ alloys: Full-solar-spectrum photovoltaic material system," Journal of Applied Physics, 94, pp. 6477-6482, (2003).

[2] C.J Gelderloos, C. Assad, P.T. Balcewicz, A.V. Mason, J.S. Powe, T.J. Priest and J.A. Schwartz, "Characterization testing of hughes 702 solar array," Proc. 28th IEEE Photovoltaic Specialists Conf., pp.972-975, (2000).

[3] M.J O'Neill, A.J McDanal, M.F Piszczor, M.I Eskenazi, P.A Jones, C. Carrington, D.L Edwards and H.W Brandhorst, "The stretched lens ultralight concentrator array," Proc. 28th IEEE Photovoltaic Specialists Conf., pp.1135-1138, (2000). [

[4] *D.D Krut, G.S. Glenn, B. Bailor, M. Takahashi, R.A. Sherif, D.R. Lillington and N.H. Karam, "Wide acceptance angle, non-imaging, triple junction based, 10× composite space concentrator," Proc. 28th IEEE Photovoltaic Specialists Conf., pp.1165-1168, (2000).*

[5] Md. Rafiqul Islam, M. A. Rayhan, M. E. Hossain, Ashraful G. Bhuiyan, M. R. Islam and A. Yamamoto,"Projected Performance of $In_XGa_{1-X}N$-based Multijunction Solar Cells", 4th International Conference on Electrical & Computer Engineering (ICECE 2006), Dhaka, Bangladesh, IEEE catalogue number: 06EX1362, December 19-21, 2006, p. 241.

[6] Md. Rafiqul Islam, Md. Tanvir Hasan, Ashraful G. Bhuiyan, M. R. Islam, and A. Yamamoto, "Design and performance of $In_xGa_{1-x}N$-based MJ solar cells", In press, IETECH Journal of Electrical Analysis.

[7] J. Nelson K. Barnham, J. Connolly, G. Haarpaintner, C. Button, and J. Roberts, "Quantum Well Solar Cell Dark Currents", Proceedings of the 12th EC Photovoltaic Solar Energy Conference, Amsterdam, 1370, (1994)

[8] F.Bechstedt, J. Furthmueller, M. Ferhat, L.K. Teles, L.M.R. Scolfaro, J.R. Leite, V.Yu. Davydov, O. Ambacher, R. Goldhahn, Energy gap and optical properties of $In_xGa_{1-X}N$, Phys. Status. Solid. (a) 195 (3), 628-633, 2003.

[9] M. Hori, K.Kano, T. Yamaguchi, Y. Satto, T. Araki, Y. Nanishi, N. Teraguchi, and A. Suzuki, Optical Properties of $In_xGa_{1-x}N$ with entire alloy composition on InN buffer layer grown by RF-MBE, Phys. stat. sol. (b) 234 (3), 750-754, 2002.

[10] H.L. Cotal, D.R. Lillington, J.H. Ermer, R.R. King, N.H. Karam, "Triple junction solar cell efficiencies above 32%: the promise and systems challenges of their application in high-concentration-ratio pv systems," Proc. 28th IEEE Photovoltaic Specialists Conf., pp.955-960, (2000).

[11] A.W. Bett, F. Dimroth, G. Lange, M. Meusel, R. Beckert, M. Hein, S.V. Riesen and U. Schubert," 30 % Monolithic tandem concentrator solar cells for Concentrations exceeding 1000 suns," Proc. 28th IEEE Photovoltaic Specialists Conf., pp.961-964, (2000).

[12]http://sunbird.jrc.it/events/0511fullspectrum/2_1_Yamaguchi .pdf, 25/08/2006

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A new SS7-SIGTRAN protocol interchanger software operated with modified USB E1 for nationwide IP backbone – Necessity for BTCL PSTN

M. S. Munir[1], K. Ahmed[1], ASM Shihavuddin[1], Tahmid Latif[1] and Md. Saifur Rahman[2]
[1]Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT), Bangladesh.
[2]Electrical and Electronic Engineering Department, Bangladesh University of Engineering and Technology (BUET),
Mailing Address: 16/9 North Circular Road, Vuter Goli, Dhanmondi, Dhaka, Bangladesh.
E-mail: {smunir, kabir_89}@iut-dhaka.edu, {shihav407, tahmidlatif, saif672}@yahoo.com,

Abstract – This paper presents a new Signaling System no. 7 (SS7) and IP based Signaling Transport (SIGTRAN) protocol interchanger software and related hardware for Bangladesh Telecommunications Company Ltd. (BTCL) to expand its existing network using IP backbone. As the demand of wireline phone of BTCL is increasing day by day, it has become a necessity for BTCL to expand its network. At this point, BTCL can expand its network by establishing IP backbone and connecting it with the existing SS7 network using the new SS7-SIGTRAN interchanger software. This will be a good utilization of the newly arrived IP based transport and switching technologies which is reducing the cost of delivering different services. Again, compared to voice, packet data is becoming more significant proportion of traffic for BTCL. It has become essential to find ways of consolidating voice, data traffic, platforms and services to reduce the operational, maintenance and initial cost of BTCL. IP network is the best solution to this problem. This makes network expansion through IP backbone while keeping connection to the existing SS7 network a necessity for BTCL which is possible by using the proposed software and hardware.

I. Introduction

Considering the success of mobile communications, it may appear that wireline phones are going to be obsolete. But the global wireline operator revenues are estimated to grow at a compound annual rate of 4% between now and 2009[1]. The index of fixed telephone lines per 100 population for major world regions during 1994-2006 shows a significant growth of the total number of the fixed telephone connection (Table-1) [2].

Table 1: Index of fixed telephone line of the world per 100 population.

	1995	2000	2005	2006
Fixed telephone	0.3	0.4	0.8	0.88
Cellular subscribers	-	0.2	6.4	13.3
Mobile phone as share of total phone lines	-	36.2	89.4	94.4

Similar to the global demand for the fixed telephone line, BTCL also has a lot of pending demand which shows that its customer base is still to be grown in future (Table-2) [1][3].

Table 1: Pending demand of BTCL.

Region	Capacity	Connection	Pending demand
Dhaka	609,350	542,265	54,211
Chittagong	237,295	138,145	26,028
Khulna	129,005	75,168	2,560
Rajshahi	61,296	34,693	1,311
Rangpur	76,119	38,210	686
Sylhet	75,01	39,721	8,692
Country Total	**11,13,065**	**8,68,202**	**93488**

So, BTCL invariably will have to expand its network in future. As the competition in the telecommunication market is intense due to the presence of the mobile operators, the cost of network expansion of BTCL will play a significant role while taking the decision. Again the various value added services provided by the other operators will put pressure on BTCL to introduce those services too. Moreover data traffic is increasingly becoming a significant proportion of the total traffic [4]. All IP network seems to be a perfect solution of this problem. But the most cost effective and easiest way to meet these challenges will be to expand the network using IP backbone while keeping connection with the existing SS7 network. This situation makes it essential to expand the network using IP backbone and use SS7-IP protocol interchanger to connect it with the existing SS7 network. At this point we have proposed a new SS7 and IP based SIGTRAN protocol interchanger software and related software to perform the task.

II. Proposed System

Protocol can be defined as the rules governing the syntax and synchronization of communication. We have worked with two protocols: SS7 and SIGTRAN.

Signaling System no. 7 (SS7) is a set of telephony signaling protocols [5], which are used to set up the vast majority of the world's public switched telephone network telephone calls. SS7 provides a universal structure for telephony network signaling, messaging, interfacing and

network maintenance [6]. SIGTRAN is the name given to an IETF working group that produced specifications [7] for a family of protocols that provide reliable datagram service and user layer adaptations for SS7 and ISDN communications protocols. Our developed software along with our designed hardware is capable of taking signaling channel data from PSTN network and analyzes that according to SS7 and SIGTRAN protocol structure defined by ITU-T [5][6][8]. Performance evaluation of SIGTRAN-based Signaling Links Deployed in Mobile Networks [9] is very important and with our software, the incorrect data can also be detected.

Our developed software takes SS7 protocol data from SS7 network through designed USB E1 card and SIGTRAN data from SIGTRAN network through Ethernet port using "Data reader" window of the software. Then the data is processed according to the protocol stack selected by the "Stack selector" window of our software. Then the data of a selected protocol is exchanged with the other protocol using the "Protocol exchanger" window of our software. The working principle of our developed hardware and software is like,

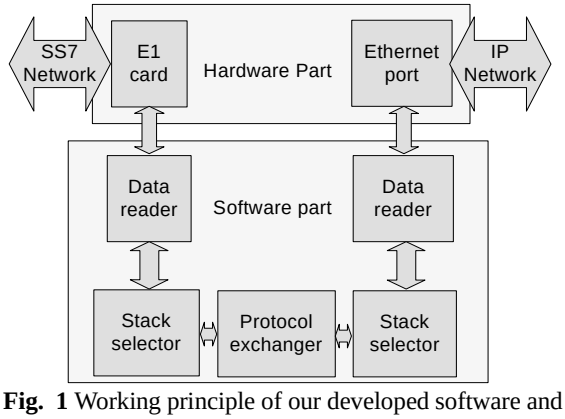


Fig. 1 Working principle of our developed software and hardware

III. Software Architecture

A. Protocol Stack

The first part of the development of our software was to develop the protocol stack of SS7 and SIGTRAN protocol.

A.1. SS7 Stack

The SS7 protocol stack borrows partially from the OSI Model of a packetized digital protocol stack. OSI layers 1 to 3 are provided by the Message Transfer Part (MTP) of the SS7 protocol; for circuit related signaling, such as the Telephone User Part (TUP) or the ISDN User Part (ISUP), the User Part provides layers 4 to 7, whereas for non-circuit related signaling the Signaling Connection and Control Part (SCCP) provides layer 4 capabilities to the SCCP user [10].

The MTP covers the transport protocols including network interface, information transfer, message handling and routing to the higher levels. SCCP is a sub-part of other L4 protocols, together with MTP 3 it can be called the Network Service Part (NSP), it provides end-to-end addressing and routing. TUP is a link-by-link signaling system used to connect calls. ISUP provides a circuit-based protocol to establish, maintain and end the connections for calls. TCAP is used to create database queries and invoke advanced network functionality mobile services (MAP), etc. [11]

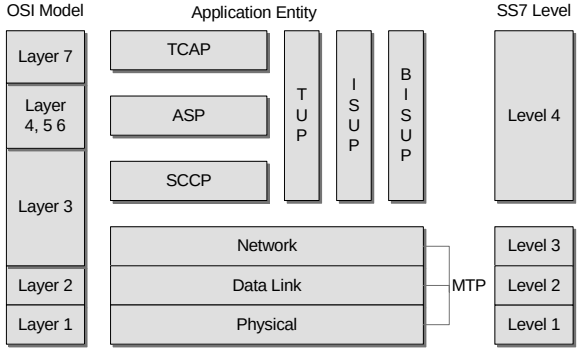


Fig. 2 SS7 Protocol Stack [12]

Each SS7 data is transmitted in MTP (Message Transfer Part). Each MTP contains different SS7 layers, which are well defined by ITU-T. The MTP structure for SS7 data used in our software is like,

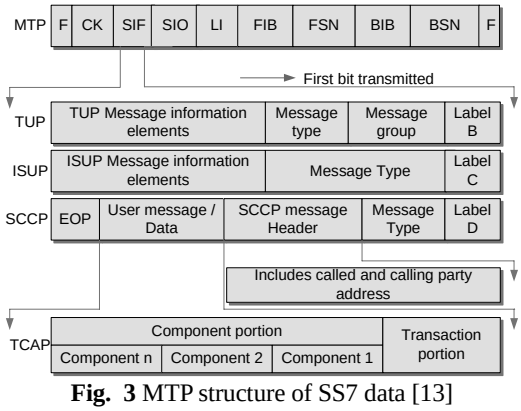


Fig. 3 MTP structure of SS7 data [13]

Using these structures we have developed the SS7 protocol stack of our software.

A.2. SIGTRAN Stack

The SIGTRAN protocols specify the means by which SS7 messages can be reliably transported over IP networks. The architecture identifies two components:

1. A common transport protocol for the SS7 protocol layer being carried and
2. An adaptation module to emulate lower layers of the protocol.

For example, if the native protocol is MTP (Message Transport Layer) Level 3, the sigtran protocols provide the equivalent functionality of MTP Level 2. If the native protocol is ISUP or SCCP, the sigtran protocols provide the same functionality as MTP Levels 2 and 3. If the native protocol is TCAP, the SIGTRAN protocols provide the functionality of SCCP (connectionless classes) & MTP Levels 2 & 3.

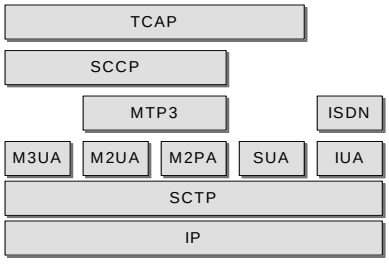


Fig. 4 SIGTRAN Protocol Stack [14]

B. SS7 and SIGTRAN Protocol Interchanger

The task of the data exchange between SS7 and SIGTRAN protocol is performed according to the following structure,

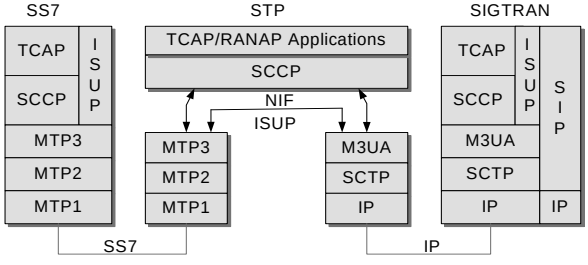


Fig. 5 SS7 and SIGTRAN protocol interconnection structure [15]

Our software exchanges SS7 protocol data via our designed USB E1 card and SIGTRAN protocol data from Ethernet port. This exchange of data between two different protocols takes place by analyzing the data [16] and maintaining the structure mentioned in figure 5. There are three part of the software which performs the data exchange operation. They are,

1. Data reader
2. Stack selector and
3. Data exchanger

Part 1: "Data Reader" window of our software is used to take SS7 data via E1 card or SIGTRAN data via Ethernet port. At first, the user has to specify the source to extract data for exchange. The selected source will be used as the base protocol which will be converted to the other protocol data. The software view for performing this task is shown bellow,

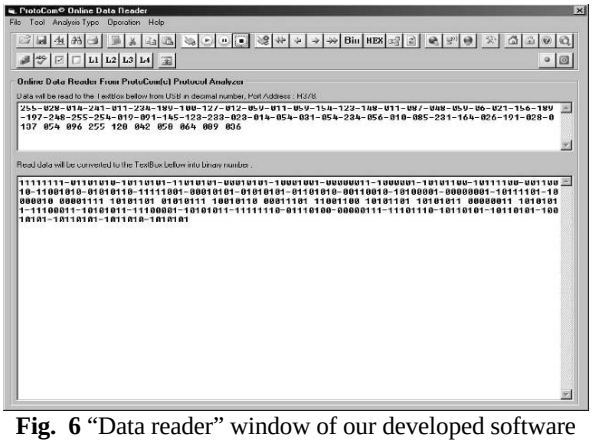


Fig. 6 "Data reader" window of our developed software

Part 2: "Stack Manager" window of our software lets the user to choose which message in the SS7 data is to be exchanged. It is very useful operation when the SS7 data length is very high. By choosing specific message, then user can reduce the number of exchanged messages so that only the desired message is transmitted. The is particularly important to maintain the various value added services easily. The software window of "Stack Manager" is shown bellow:

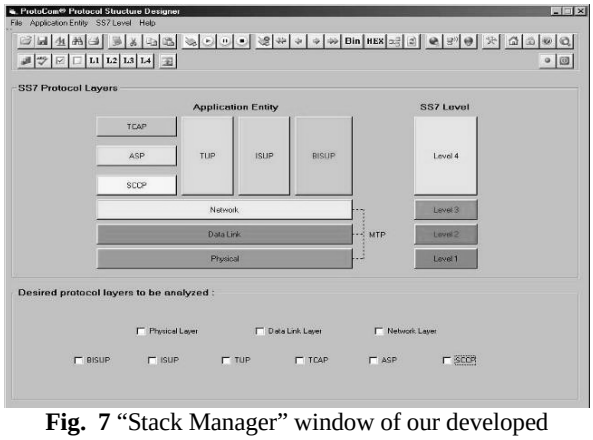


Fig. 7 "Stack Manager" window of our developed software

Part 3: "Data Exchanger" is the last part of our software with which the selected portion of the base protocol data is exchanged to the other protocol data. During this exchange process, the exchanged data can be saved for analysis. Our exchanger also has an analyzer mode with the following features which meets almost all the features of current available protocol analyzers:

- Summary View displays MTP2, MTP3 information and SS7 Message types, Called and Calling number, SCCP message type, SSN, INAP information, and more
- Hex/Binary view displays the frame information in HEX and ASCII format
- Exports detailed and summary information to a comma delimited file for subsequent import into a database or spreadsheet
- Supports filtering and search features based on Frame length, FSN, BSN, SSN and so on
- Hex Dump View displays the frame information in HEX and BINARY format.
- View of duration of completed call, OPC, DPC, CIC, Called and Calling Party Numbers, and more. [17]

The data read from the Ethernet port or E1 card using the selected source by the Data Reader window of the software and according to the selected stack using the Stack Manager window of the software. This data is then converted into the other protocol data and routed. The software window performing data exchanging from SS7 to SIGTRAN protocol is shown bellow,

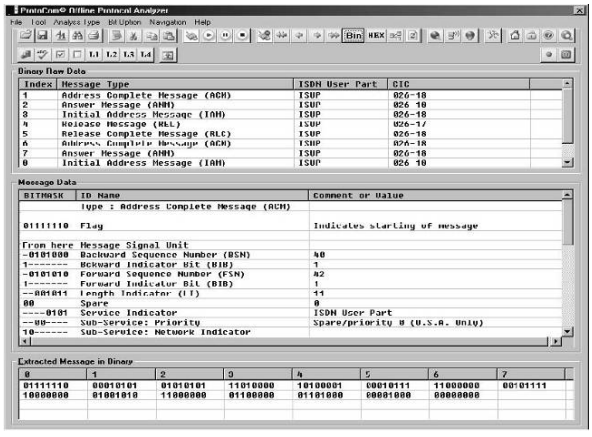


Fig. 8 "Data Exchanger" window of our developed software.

IV. Hardware Architecture

For reading data from SS7 link, we have designed an USB E1 card with which the data of signaling channel of SS7 link can be directly saved by our developed software. Our USB E1 card is an enhanced third-generation hardware that consolidates the essential pieces of industry-standard test equipment into a powerful, PC-Based USB E1 solution.

PEF2256: The FALC56 is the latest addition to Infineon's FALC® family of sophisticated E1/T1/J1 framer and Line Interface Unit (LIU) transceivers which supports all standard E1/T1/J1 functions. [18]

ATmega16: The ATmega16 is a low-power CMOS 8-bit microcontroller based on the AVR enhanced RISC architecture, which achieves throughputs approaching 1 MIPS per MHz allowing optimization of power consumption. [19]

FT245BM: The FT245BM is the 2nd generation of FTDI's popular Single Chip USB Parallel FIFO bi-directional Data Transfer I.C. The FT245BM provides an easy cost-effective method of transferring data to / from a peripheral and a host P.C. at up to 8 Million bits (1 Megabyte) per second. [20]

74HC244: These octal buffers and line drivers are designed specifically to improve both the performance and density of 3-state memory address drivers, clock drivers, and bus-oriented receivers and transmitters.

Using these ICs we have designed the USB E1 card. For avoiding complex PCB, we have designed the E1 card in two parts. They are,

1. Transceiver module: It includes
    - Transceiver IC
    - Synchronizing Clock
    - Impedance matching transformer
    - Transient voltage Suppressor
    - Connector for E1 link
2. Controlling module: It includes
    - Microcontroller
    - USB interface
    - Clock
    - Serial and Parallel interface
    - Power supply unit

The schematics of the controlling and transceiver module are shown bellow.

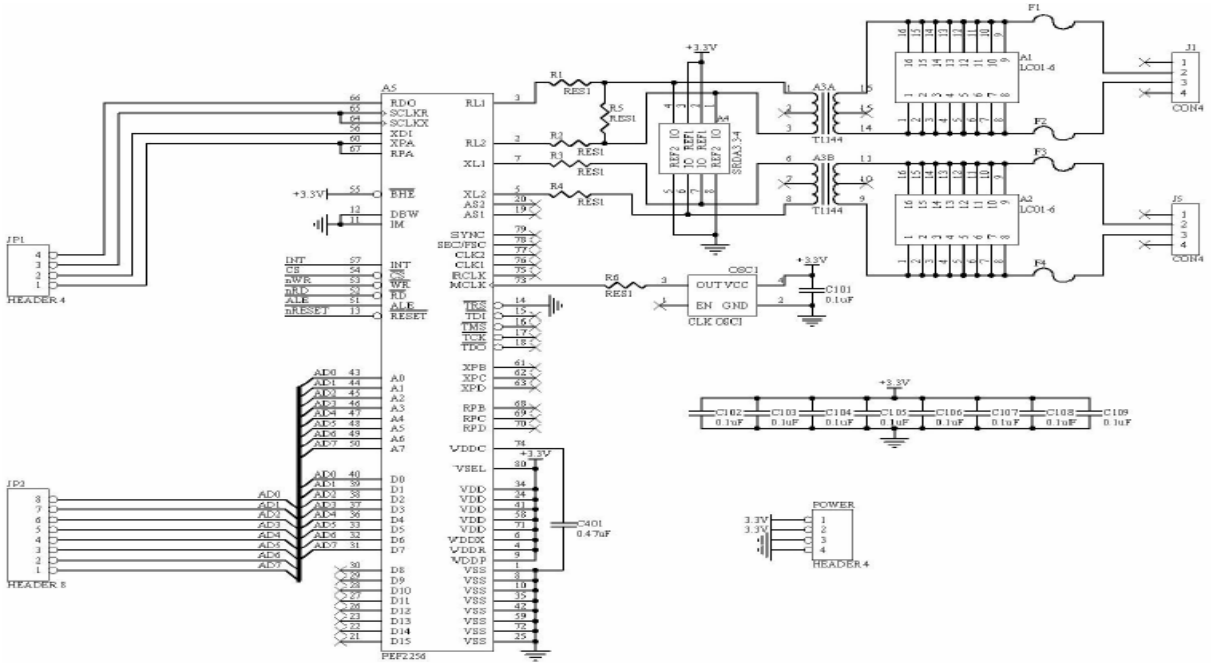


Fig. 9 Transceiver module of the E1 card.

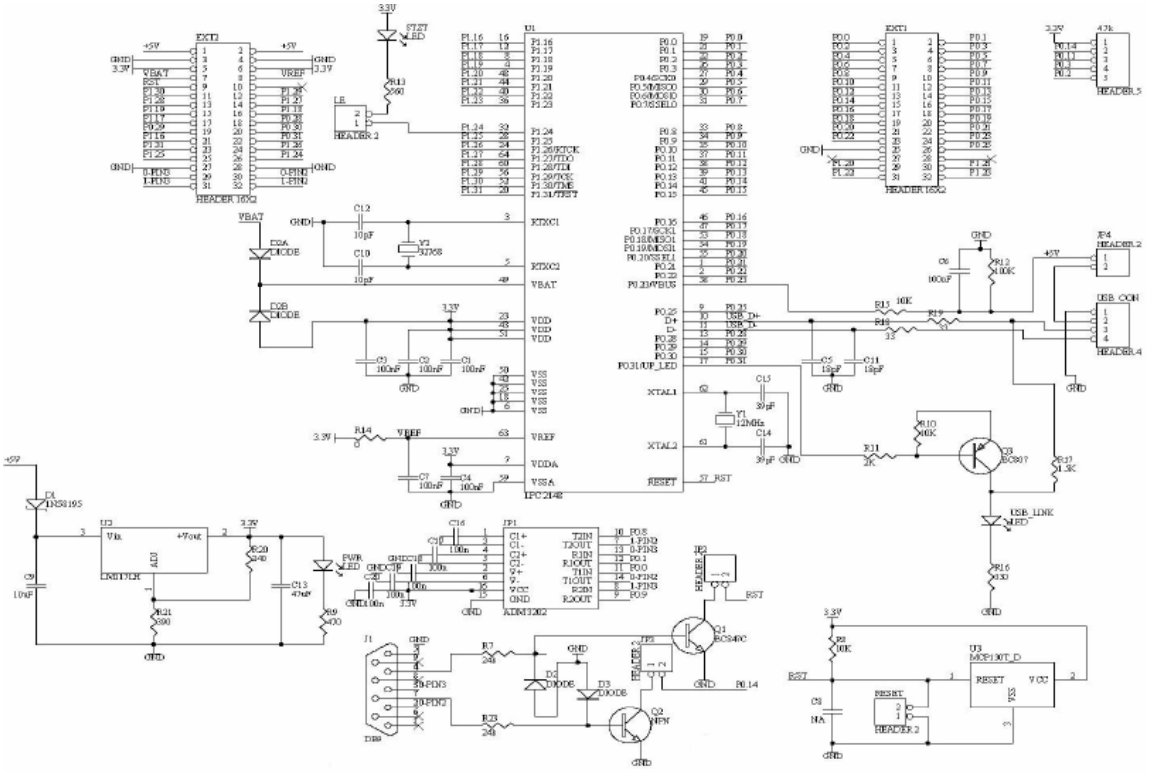
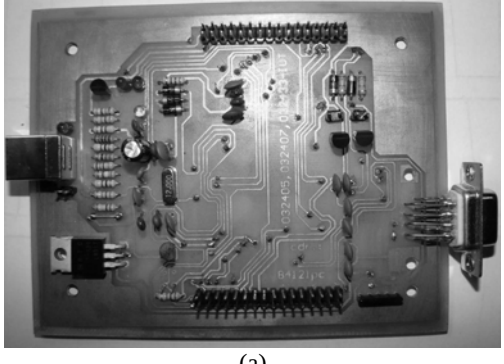


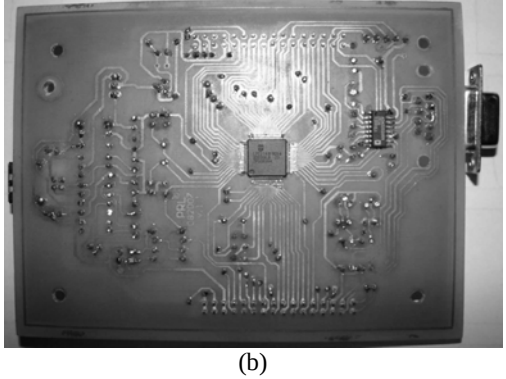
Fig. 10 Controlling part of the E1 card.

V. Result

The hardware designing part is completed and currently it's working soundly.



(a)



(b)

Fig. 11 (a) Top and **(b)** Bottom layer of the designed PCB of the E1 card

SS7 and SIGTRAN data can be read [Fig. 6] according to the selected structure [Fig. 7] and exchanged using our hardware and software [Fig. 8] successfully. One of the major achievements of our work is the tremendous reduction of the cost. The manufacturing cost of the hardware is very minute compared to the market price of similar range hardware. A PCI slot E1 card costs about 500USD in the current market whereas our hardware costs only 55USD.

VI. Conclusion

In this paper, we have presented a new SS7 and SIGTRAN protocol exchanger software and modified USB E1 card hardware. Protocol exchanger is going to be very necessary to consolidate voice, data traffic, platforms and services to reduce the operational, maintenance and initial cost of the network. This software will enable BTCL to expand the network using IP backbone which is very cheap compared to SS7 network. Whereas, existing SS7 network can be connected with the new IP network easily. Also the different value added services will be easier to provide through the SIGTRAN network. Finally, we believe that this software and hardware will be a feasible solution for BTCL to expand its network.

References

[1] Md. Anwar Hossain Masud. "Nationwide all IP network: necessity and possibility for BTTB PSTN". ICECE 2006.

[2] Global fixed and mobile connection growth status. http://www.unescap.org.

[3] BTCL yearly report from www.btcl.gov.bd.
[4] T. Yamanda. “A new public network including IP communications as a universal service”. 2000. Ritsumeikan University, Japan.
[5] ITU data from http://www.itu.int/ITU-D/
[6] ITU-T Recommendations, Q.700 to Q.716.
[7] Fred Halsal, Multimedia communication.
[8] IETF reference document “Framework Architecture for Signaling Transport”, RFC-2719, http://www.ietf.org/rfc/rfc2719.txt.
[9] Performance Technologies protocol standard document. www.pt.com.
[10] Chukarin, A. Pershakov, N. Samouylov. ‘Performance of Sigtran-based Signaling Links Deployed in Mobile Networks.’ ConTel 2007.
[11] Dawis, E.P. ‘Architecture of an SS7 protocol stack on a broadband switch platform using dualistic Petri nets.’ Communications, Computers and signal Processing, 2001. PACRIM.
[12] ITU-T reference document Q.701, Page 5.
[13] “Tutorial on Signaling System 7 (SS7)” from Performance Technologies, Page 6. www.pt.com.
[14] ITU-T reference document Q.700 page 14.
[15] SIGTRAN Protocol Suite by Jim Darroch, Page 8.
[16] Reference document on “SS7 over IP signaling transport and SCCP” by International Engineering Consortium, Page 16. www.iec.org.
[17] Signaling System # 7 protocol analyzer software. www.gl.com/ss7.html.
[18] GL communication protocol analyzer software standard www.gl.com/protocol_analysis.html.
[19] Falc2256 product brief of Infenion Technologies, Page 1. www.infenion.com.
[20] ATmega16L Datasheet from ATMEL Corporation. Page 1. www.atmel.com.
[21] FT245BM datasheet from Future technologies Ltd. Page 1. www.ftdichip.com

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Security Adaptive Protocol Suite: Ranked Neighbor Discovery (RND) and Security Adaptive AODV (SA-AODV)

Rasib Hassan Khan , K. M. Imtiaz-ud-Din , Abdullah Ali Faruq ,
Abu Raihan Mostofa Kamal , Prof. Dr. Abdul Mottalib

Department of Computer Science and Information Technology,
Islamic University of Technology (IUT),
Board Bazar, Gazipur, Dhaka

E-mail: rasib85@hotmail.com, imtu7986@hotmail.com, himel0000@yahoo.com,
raihanrcc@gmail.com, mottalib@iut-dhaka.edu

Abstract - ***With the increasing demand of mobility and ad hoc networking, vulnerability of wireless networks is also becoming a crucial issue. This dissertation sheds light on the security features of wireless communication, and proposes a model with an increased integration of security features. The proposed model, a Ranked Neighbor Discovery (RND), and a Security Adaptive Ad-Hoc On-Demand Distance Vector (SA-AODV) routing protocol suite lays out the percept of solution for the security issues, which includes the neighbor discovery, as well as the routing protocol for transmission. The neighbor discovery phase consists of the determination of trusted neighbors, based on distance metrics, leading to a process of trust ranking. The routing protocol, using the fact of the trusted neighbors, and the required security level, then sets up a security adapted route from the source to its destination. The main advantage of this protocol pair would be the achievement to obtain a route with a user-defined level of security for a specific application. The two protocols thus provide the anchor to a package for a total solution for a secured environment for wireless transmission with an increased integration of security features.***

I. Introduction

Wireless technologies have become increasingly popular in our everyday business and personal lives, and gadgets like laptops, personal digital assistants (PDA), cellular phones are common, allowing everyone access to the world of information. Thus, as the ultimate backbone of communication is turning out to be wireless, the security of the medium of transmission is becoming an increasingly important issue. In wireless networks, ad hoc networking is currently a very active area of research, and the development of the most optimum protocol for all terrain use is still an aim trying to be fulfilled. In an ad hoc network, wireless nodes cooperate to form a network, forwarding packets for each other to allow nodes not within direct wireless transmission range of each other to communicate. In contrast to traditional network routing protocols, for example for wired networks, the behavior of ad hoc networks can be quite dynamic due to factors such as node movement and variations in radio propagation condition, creating frequent changes in network topology, differing concentration in traffic load on the network, and other challenges to the operation of the network protocols, and thus must adapt more quickly.

Wireless networks implement two sets of protocols, for a secured networking environment. One is the initial phase of neighbor discovery [1], and next comes the play of the routing protocols [7, 8]. The neighbor discovery phase is very crucial, in the sense that, the routing protocols will use the information from this phase, for setting up a route from a specific source to its destination.

Extensive studies ensued in many protocols, and repeated improvements of the existing ones. Many of these applications may run in non-trusted environments and may therefore require the use of a secure routing protocol. The basic problem with all of these protocols still remains; the functionality to reduce packet drops during transmission, the mechanism to detect wormholes, tunnels, and ability to distinguish between adversary nodes and trusted nodes.

II. A Formal Definition of Security

A routing protocol is said to be (*statistically*) secure if, for any configuration and any real-world adversary, there exists an ideal-world adversary, such that the output of the real-world model is (*statistically*) indistinguishable from the output of the ideal-world model. [2]

III. The Wireless Environment

A. Securing the Process of Neighbor Discovery

Many wireless networking mechanisms, notably routing, require that wireless nodes be aware of their neighborhood. This means that the nodes must know which other nodes they can communicate with directly. The procedure used to acquire this knowledge is called *neighbor discovery* [1]. In mobile wireless networks, the neighbor relationships change dynamically, which makes neighbor discovery an important mechanism.

Neighbor discovery can be achieved through simple protocols, where a node that wants to determine who its neighbors are broadcasts a neighbor discovery request, and every node that receives this request responds with a neighbor discovery reply. Receiving a reply means that the requesting node and the responding node can hear each other's transmission, and can communicate with each other directly, and hence are neighbors. The neighbor discovery protocol is sometimes called "*hello protocol",* and the request and the reply are called "*hello messages"* [1].

An adversary can try to thwart the successful execution of the neighbor discovery protocol, for instance, by jamming the communication between two nodes, or by providing a node with false information regarding another node, which is not a direct neighbor in reality, but leading on to make the requesting node believe that the other node is indeed a direct neighbor. In this way, the adversary achieves that two nodes, which otherwise could communicate directly, cannot establish a neighbor relationship, or a relationship with faulty information [1, 3, 5]. Blocking the links between many pairs of nodes in this manner can have serious consequences to the connectivity of the network, and on the upper layer protocols, such as routing and transmission.

The different types of secured neighbor discovery protocols are the centralized approaches of Statistical Detection, Multi-Dimensional Scaling [1, 2]. The other approach, the decentralized mechanisms, [2] are those with Position Information of Anchors, Directional Antennas, and Distance Estimation Techniques, which include Geographical Leashing, Temporal Leashing, and Mutual Authenticated Distance-Bounding (MAD). We will however consider the Temporal Leashing with TESLA [9] Instant Key-Disclosure (TIK) [2], protocol as the base, with a modification, for the proposed RND protocol.

B. The Routing Protocols

The main divisions of protocol categorization are in terms of reactive and proactive. Proactive protocol can be termed as those which act on their own, which means they find routing paths independently of the usage of the paths. On the other hand, reactive routes are those which set up a route only when demanded by a source to communicate with a destination. In regular use, we find the reactive protocols in practice, namely Dynamic Source Routing (DSR) protocol [2, 7], and Ad-Hoc On-Demand Distance Vector (AODV) protocols [2, 8].

DSR is a routing protocol, similar to AODV in that it forms a route on-demand when a transmitting computer requests one. This protocol is truly based on source routing whereby all the routing information is maintained and continually updated at the mobile nodes. The main disadvantages in DSR are that, it does not locally repair a broken link; the stale route cache information could result in inconsistencies; the connection setup delay is higher; performance degrades rapidly with increasing mobility; and routing overhead is high.

When compared to the other protocol, AODV, which has been jointly developed in Nokia Research Center of University of California, Santa Barbara and University of Cincinnati, it has a lot of significant advantages. AODV has already been modified into the Secured AODV (SAODV), and is globally the most recognized and implemented protocol for wireless networks. The SAODV is an extension of the AODV routing protocol that can be used to protect the route discovery mechanism providing security features like integrity, authentication and non-repudiation. AODV, a distance vector protocol, remains silent until a request is generated, and then sets up the route on the basis of next hop addresses. With AODV, we can avail the advantages if it being capable of unicast and multicast routing, distance-vector routing, avoidance to the counting-to-infinity problem, setting up of the latest route to the destination, and the connection setup delay being less. We will later see why these features are important, and how these facts are used in the modified form in SA-AODV.

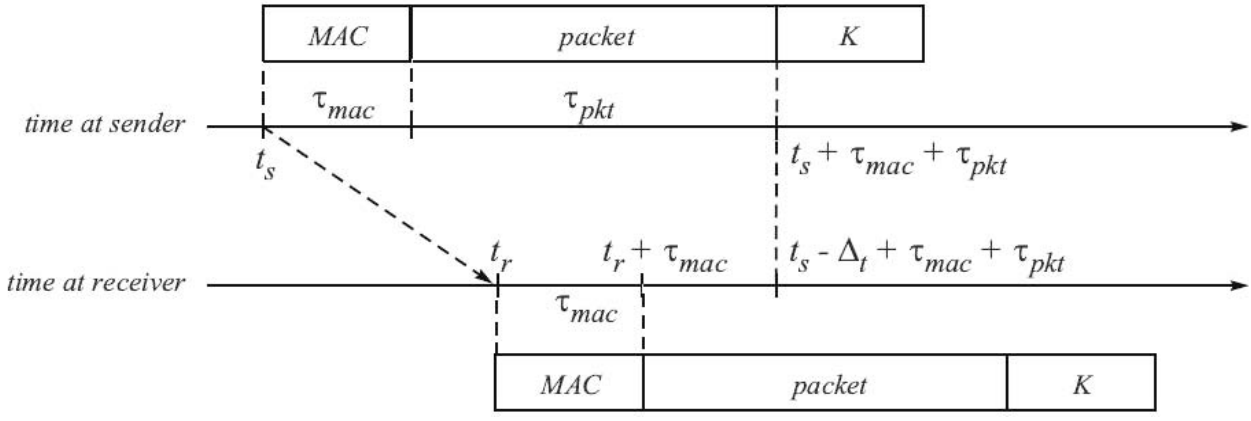


Figure 1: Timing diagram to show the sequence of information correspondence between the sender and the receiver for them to synchronize in TIC, before they start to transmit data [2].

IV. A Security Adaptive Protocol Suite

The proposed process is a technique for neighbor discovery, combined with a routing protocol, the whole of which is based on a rank-based system. In the phase of neighbor discovery, the neighboring nodes will be assigned a rank, with which we will be able to judge the trustworthiness of a neighboring node. With the information we have collected with the rank-based system of trust for the neighboring nodes, we will then proceed onto a routing protocol. Talking in shallow terms, this protocol will be having different levels of implementation, with different levels of security for the application for which the route is being created. Whenever a route is demanded from a source to a destination, the source will judge the security requirement of the application, and then, with the ranking information at the neighboring nodes, it will thus route the packet to its destination. It most cases, it will be considered that a higher secured route will have a trade-off with respect to performance for a longer path.

A. Ranked Neighbor Discovery (RND)

A straightforward idea for wormhole detection is to estimate the real physical distance between the nodes that are believed to be neighbors. With general knowledge, we can say that, if a node that is physically further away than what the routing table information reflects, then there is presence of a wormhole in the network [1, 3, 5].

The distance estimation techniques of temporal leashing will the main point of focus for the modified form of the algorithm for neighbor discovery. The main idea of the mechanism is to add some information to the packets that restricts their maximum allowed transmission distance. With this information, we are actually limiting the maximum distance that the packet will travel. A temporal leash is based on timing information, and it ensures that the packet has an upper bound on its lifetime.

In this process, it a requirement that the nodes have tightly synchronized clocks, such that the maximum difference Δt between any two nodes' clocks is in the order of a few hundred nanoseconds. Thus, we see that the difference of time is a crucial factor in this case. As a result, we need to have a very efficient technique, which will efficiently synchronize the sender and the receiver. In this regard, we will use the TESLA with Instant Key-disclosure (TIK), to authenticate temporal leashes in packets. TESLA combines the advantages of digital signatures and Message Authentication Codes (MAC). The method of TESLA is discussed in the referred paper, and instead, we here discuss the concept of the TIK protocol for synchronization.

It has been observed that the authentication delay of TESLA can be removed in an environment where the nodes' clocks are tightly synchronized, and the TIK protocol is based on this fact. TESLA requires that the MAC value of the packet is received earlier by the receiver than the time at which the TESLA key used for computing the MAC is disclosed by the sender. This can be achieved by sending the MAC value at the beginning of the transmission and disclosing the TESLA key at the end of the same transmission, as shown in figure-1. When the entire MAC is received, the receiver's clock shows:

$$t_r + T_{MAC}$$

When the key disclosure is started, the sender's clock shows:

$$t_s + T_{MAC} + T_{PKT}$$

At the same moment, time at the receiver can be a minimum value of:

$$t_s - \Delta t + T_{MAC} + T_{PKT}$$

Hence, the receiver thus checks for the TESLA condition to be satisfied (i.e., the full MAC is received before any bit of the key with which it was computed is released), and the receiver can start the verification of the MAC essentially without any delay:

$$t_r + T_{MAC} < t_s - \Delta t + T_{MAC} + T_{PKT}$$

where t_s is known to the receiver from the temporal leash in the packet. Clearly, in order for this to work, very precise timings are needed and, in particular, Δt must be very small (or otherwise packets need to be extremely long).

After the phase of synchronization, the process of distance estimation begins. When sending a packet, the sender includes in the packet the time t_s of sending the first bit of the packet. When receiving a packet, t_s is compared to the time t_r of receiving the first bit of the packet at the receiver. More precisely, the receiver computes an upper bound on its distance d' to the sender as:

$$d' = V_{light}(t_r - t_s - \Delta t) \quad [V_{light} = \text{Speed of light}] \qquad (1)$$

After that, the sender node will then calculate the trust values, or the rank of its neighbors, based on its predefined maximum range, or the maximum radius of its transmission, T.

We consider in this case, that the defined range of ranks, R, for the neighbors is on a scale of 0 to 4. Thus, the value of d', previously obtained from equation (1), is now compared to T, as shown in the table below.

Table 1: Checking of d′ and T, and corresponding R

Distance Estimation	Rank Assigned, *R*
$d' \leq T/4$	4
$T/4 < d' \leq T/2$	3
$T/2 < d' \leq 3T/4$	2
$3RT/4 < d' \leq T$	1
$T > d'$	0

In figure 2, we see a node going though the phase of the ranked neighbor discovery. In this specific scenario, we consider T = 150m. Thus, the messages revealed three neighbors, A, B, and C, placed at 100m, 125m, and 50m respectively. Therefore, the assigned ranks to the respective nodes are shown table 2.

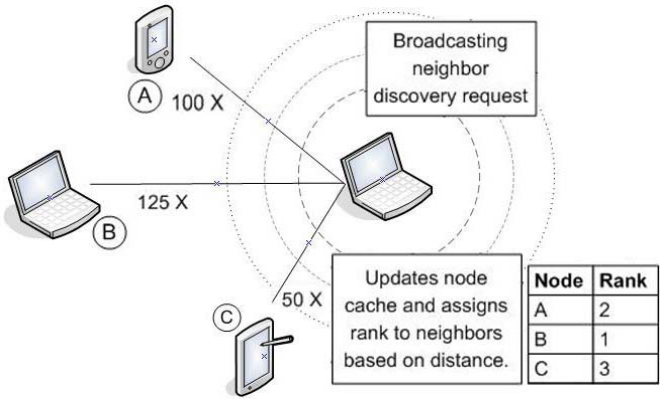


Figure 2: A node going through the phase of neighbor discovery, and assigning ranks to them.

Table 2: Distance estimation results from figure 2 and the assigned ranks to the neighbors

Max Transmission Distance, T = 150 m		
Node	**Distance of neighbor, d′ (m)**	**Rank Assigned, *R***
A	100 *(T/2<d′≤3T/4)*	2
B	125 *(3RT/4<d′≤T)*	1
C	50 *(T/4<d′≤T/2)*	3

With the trust ranks assigned to the discovered neighbors, the source node will update its routing table information in its cache. When the requirement arises for the routing of a packet, it will then go on to proceed with a security adaptive routing, discussed in the next sub-section.

B. Security Adaptive Ad-Hoc On-demand Vector Routing (SA-AODV)

The ad-hoc on-demand distance vector (AODV) routing is a routing protocol for Mobile Ad-hoc Networks (MANETs) and other wireless ad-hoc networks. AODV is capable of both unicast and multicast routing. The fact that AODV is a reactive protocol implies that it sets up a route only when it is required. AODV is, as we can see from the name, a distance-vector routing protocol. AODV avoids the *counting-to-infinity* problem of other distance-vector protocols by using sequence numbers on route updates.

Considering the fact that AODV is a reactive protocol, and it avoids the counting to infinity problem, by the use of sequencing numbers, we will consider this to the basic backbone protocol for a security adaptive protocol, the *SA-AODV*.

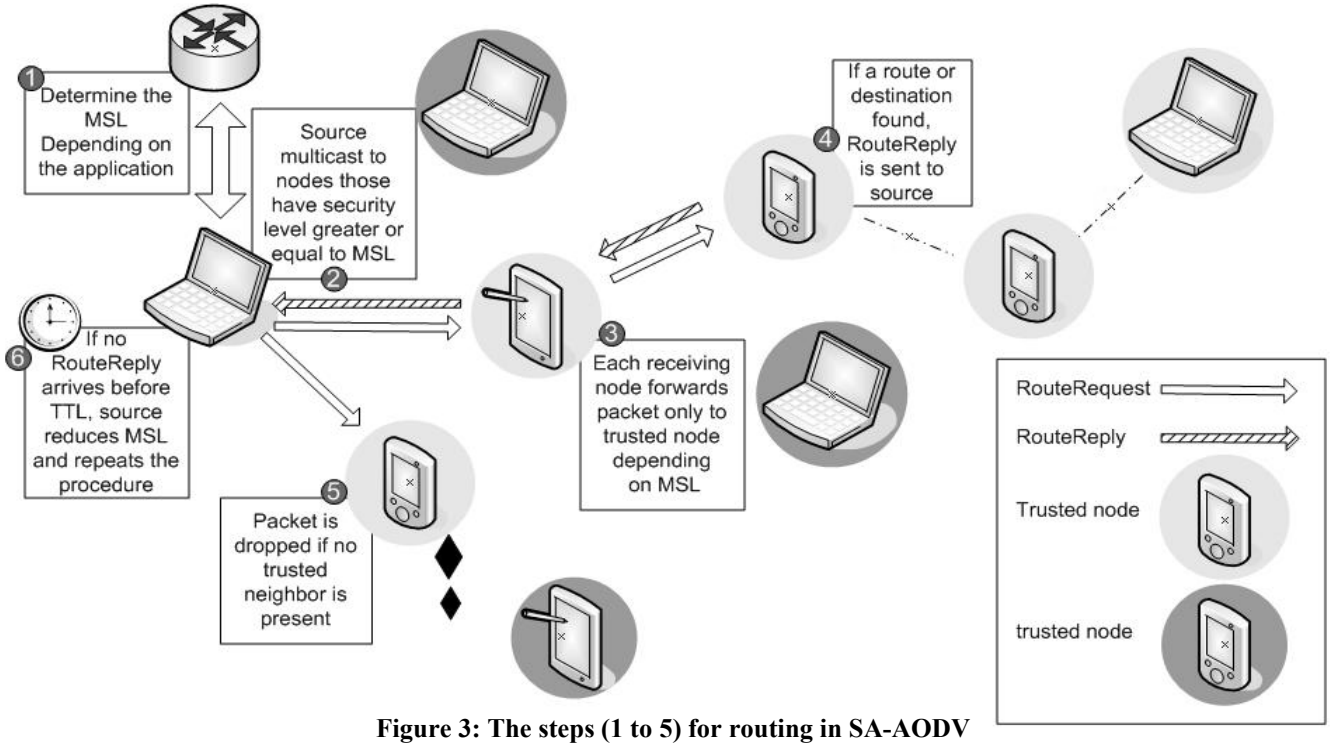


Figure 3: The steps (1 to 5) for routing in SA-AODV

By the use of a RND algorithm, we were able to distinctively distinguish the wormholes present in the network. After that, we assigned the different levels of trust, denoted by the rank value, and did never discard any information about any node, may that be an adversary, or a valid node.

When a route is requested from a source node, A, to a destination node, B, in AODV, a route request is broadcasted. In this case, the basic principal will be same, but the only difference lies in the fact that, before a route request is broadcasted, the security level requirement has to be defined, being called here as the *Minimum Security Level* (MSL), from a MSL-Database (MSL-DB). This will be done on a predefined scaling basis, related to the rank values for the nodes. Thus, with a security requirement level, the node now directively broadcasts the route request only to the nodes with a minimum level of trust. This specific operation is being possible to be implemented, as because AODV is able to unicast, as well as multicast. The neighbors, upon receipt of the request, will react the same way, as it is done in the traditional AODV, except for the slight change in the way it rebroadcast, as mentioned above.

The process of directive broadcasting, with reference to the MSL, will help the *Route Requests* reach the destination, but through a route, which has a defined MSL, with respect to its physical distance estimation. Thus, upon receipt of the *Route Request* by the destination node, it sends back the *Route Reply*. Therefore, the route discovery mechanism will return a path from the source to the destination, with a defined security requirement, defined here as a *Security Adapted Route* (SAR).

In case a route is not being able to be established with the initial MSL, with which the *Route Request* was broadcasted, an *Error* packet will be generated. The source node will be notified of the fact that the route request that has been sent is not returning a path with the defined MSL. In this scenario, the source will now define a new minimum MSL, by decrementing the MSL value, and will rebroadcast the *Route Request* to a new set of nodes. This pro-activity ensures that no matter what is the current state of the network, we will always get the best route with the required security level for a specific demanded link for a specific application, from a source to a destination.

In figure-3, a scenario for routing in SA-AODV is demonstrated. As to the figure, it can be seen a source node at first determines the MSL from the database (step 1). The *Route Request* is then directively broadcasted to only those nodes, with rank equal to or greater to MSL (step 2). Each and every subsequent node, that receives the *Route Request*, forwards the packet, maintaining the MSL requirement (step 3). As the *Route Requests* are being forwarded in this manner, a node receives the request, and finds an entry on its routing table being maintained in its cache. The node then sends the *Route Reply* back to the source node, with the route information for the SAR (step 4). As the

routing is being done from the source to another node, there might be a lot of packets being circulated in loops, and also nodes which does not have any neighbors with the defined MSL. For these cases, as soon as the nodes find a packet with an old sequence number, or the node does not find any neighbors to forward it to, the packet is dropped (step 5), and this ensures that the network is not jammed with unwanted and useless *Route Request* packets. If a route is not found, after a certain *time-to-live* (TTL), a new request is broadcasted, with a decremented value of MSL (step 6).

C. The Security Adaptive Protocol as a Solution

The proposed model is only a theoretical model, designed with the main objective of focusing on the security issues of a wireless environment. The main egress of wireless communication is the challenge of physical distances to be doled out by the routing protocol. This loophole is the vulnerability which adversary nodes try to explore, and thus, when dealing with a secured application over the wireless network, it is a valid trade off with the overall performance of the network if security is concerned.

If the advantages of the model are mooted upon, the following features will be exposed:

- Sender and receiver are synchronized before the transmission begins.
- Exchange of keys ensures authenticity of the transmission.
- An upper bound for the real physical distances is being estimated.
- Wormholes and tunnels are being discovered.
- Ranks for the trust values are being assigned, which ensures the validity of the positions of the nodes.
- The MSL for different applications ensures the communicating nodes of the security requirement.
- The adjustment of the MSL ensures that at any present time, the most secured path available is always returned.
- The feature of directive broadcasting of route requests ensures minimum number of packets in the network.
- The use of sequence numbers let the nodes identify the old useless request packets, keeps the network free from unnecessary traffic.

The model is definitely not the best proposal for any ideal wireless environment. The exchange of keys, the synchronization, and the other security features will definitely reduce the level of performance. The additional field of MSL will also increase the size of the route request packets, and thus again, slowing down the process. There will also be the requirement to introduce a database to maintain the MSL for the different applications. The mobility of the nodes is also a concern. This is because, after the phase of discovery, between the periods of the beaconing, if the nodes are displaced from their reference positions there will again be the factor of lost packets.

V. Future works and conclusion

This proposed model, being on theoretical grounds, does not ensure the optimum performance in implementation levels. Thus, close to real world simulations will be the next focus for the continuum on this topic. Apart from that, cryptographic viewpoints are also being considered, and forthcoming works are intended to include them too.

There are always a lot of scopes for improvement for advanced technologies. For wireless networks, being the most researched arena, continual improvements, and better and more efficient techniques are being developed every day. Even after all these, we can never be able to make the ultimate optimum protocol, ideal for all sorts of environments, platforms, and applications. All we can try to do is to generate the best option for a specific wireless environment.

References

[1] Panos Papadimitratos, Marcin Poturalski, Patrik Schaller, Pascal Lafourcade, David Basin, Srdjan Capkun, Jean-Pierre Hubaux - *Secure Neighbour Discovery: A Fundamental Element for Mobile Ad Hoc Networking*, IEEE February 2008.

[2] Levente Buttyán and Jean-Pierre Hubaux - *Security and cooperation in Wireless Networks*, July 27, 2007.

[3] YihChun Hu, Adrian Perrig, David B. Johnson - *Rushing Attacks and Defense in Wireless Ad Hoc Network Routing Protocols*, September 2003.

[4] Menezes, P. van, Oorschot, and S. Vanstone - *Handbook of Applied Cryptography.*

[5] Yih-Chun Hu, Adrian Perrig, David B. Johnson - *Packet Leashes: A Defense against Wormhole Attacks in Wireless Networks.*

[6] Amit Kumar Saha, Khoa Anh To, Santashil PalChaudhuri, Shu Du, David B. Johnson - *Physical Implementation and Evaluation of Ad Hoc Network Routing Protocols using Unmodified Simulation Models*, April 2005.

[7] David B. Johnson and David A. Maltz - *Dynamic Source Routing in Ad Hoc Wireless Networks in Mobile Computing*, edited by Tomasz Imielinski and Hank Korth, Kluwer Academic Publishers, 1996.

[8] Charles E. Perkins and Elizabeth M. Royer - *Ad-Hoc On-Demand Distance Vector Routing*, February 1999.

[9] A. Perrig, R. Canetti, J. D. Tygar, and D. Song - *Efficient authentication and signing of multicast streams over lossy channels.* In Proceedings of the IEEE Symposium on Research in Security and Privacy, May 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Policy Based Admission Control and Handoff Decision Algorithm for Next Generation All-IP Wireless Network

Sulata Mitra
Dept. of Computer Sc. & Tech., Bengal Engineering Science University, India
e-mail:mitra_sulata@hotmail.com

***Abstract*—The present work is a policy based admission control algorithm and handoff decision algorithm for next generation all-IP wireless network. It uses policy based admission control algorithm to admit a new call and policy based handoff decision algorithm to admit a handoff call. The proposed scheme also uses the network initiated vertical handoff algorithm to perform load balancing, to maximize battery power life time of a mobile node, to minimize power consumption, system delay and loss of packet, and finally to maximize the throughput of the network. Results based on a detailed performance evaluation study are also presented to demonstrate the efficacy of the proposed scheme.**

I. Introduction

In the next generation wireless network, a user may roam over a series of networks during his global travel. Internet browsing "on-the-move", video conferencing and file transfer are some of the new expected services in near future. The present work considers the all-IP integration architecture of the dynamic mobility management scheme [1].

The mobile node (MN) sends MN route request message to select a new route in case the MN has data packets for transmission or reception. The proposed scheme considers such request message as new call request and uses policy based admission control (PBAC) algorithm to establish the new route. The MN sends MN handoff request message to update an existing route in case the remaining battery power life time becomes equal to a threshold. In the proposed scheme the updation of an existing route indicates the selection of a new route either in the same domain or in the other domain so that transmission or reception of the remaining data packets of MN is over within the remaining battery power life time. The proposed scheme considers such request message as handoff call request and uses policy based handoff decision (PBHD) algorithm to update the existing route. The handoff call request has higher priority over new call request. The PBAC and PBHD algorithm uses one of three functions which are known as intra subnet (intra_subnet), intra domain (intra_domain) and inter domain (inter_domain) function. The policy enforcement point (PEP) is maintained at each local agent (LA), subnet agent (SA), mobility agent (MA) and global mobility agent (GMA). The policy decision point (PDP) is maintained at each SA, MA and GMA to trigger intra subnet function, intra domain function and inter domain function respectively. GMA triggers network initiated vertical handoff algorithm to transfer the existing route from one domain to other domain in case the power consumption of the domain having the existing route crosses a threshold.

The PBAC algorithm selects the most appropriate route for a MN using which transmission or reception of data packets is over within the battery power life time of the MN. In case of any additional delay in the existing route, PBHD algorithm updates it to reduce the packet loss and the wastage of battery power life time of the MN who is using the existing route. The network initiated vertical handoff algorithm triggers vertical handoff dynamically depending upon the network condition. It initiates vertical handoff for MNs having maximum excess cost and minimum allowable delay from cellular domain to WLAN domain due to low cost, high speed and high bandwidth of WLAN domain whereas it initiates vertical handoff for MNs having minimum excess cost and maximum allowable delay from WLAN domain to cellular domain due to high cost and low bandwidth of cellular domain. This algorithm helps to balance the effective load of the two domain by maintaining the power consumption of both the domain below a threshold. As a result the average value of the delay, cost, network satisfaction and throughput of the network remain almost constant.

II. Previous work

Nair and Jhu introduced [2] network latency, congestion, battery power, service type as important performance criteria to evaluate seamless vertical mobility. An end-to-end mobility management system is proposed in [3] to reduce unnecessary handoff and ping-pong effect by using measurements on the condition of different networks. Nasser et al. proposed a vertical handoff decision method [4] to calculate the service quality for available networks and selects the network with the highest quality. The vertical handoff algorithms in [2,4] are not adequate to coordinate the QoS of many individual mobile users or adapt to newly emerging performance requirements for handoff and changing network status. The vertical handoff decision function for heterogeneous wireless network [5] is a measurement of network quality. But the authors did not provide any performance analysis. An active application oriented handoff decision algorithm [6] was proposed for multi interface mobile terminals to reduce the power consumption caused by unnecessary handoff and other unnecessary interface activation.

III. Present work

In this section the proposed scheme is considered for discussion. When a MN say M sends MN route request message for the selection of a new route or MN handoff request message for updating an existing route to LA having LA_id=LA1=110 under the coverage area of SA having SA_id=SA1=11, MA having MA_id=MA1=1 and GMA having GMA_id=N1=0 (Fig.1), a new route is established using PBAC algorithm or an existing route is updated using PBHD algorithm. The route is identified as M->LA1->SA1->MA1->N1->CN i.e. M->110->11->1->0->CN (Fig.1) in case of transmission or

CN->N1->MA1->SA1->LA1->M i.e. CN->0->1->11->110->M (Fig.1) in case of reception. N1, MA1, SA1 and LA1 are the nodes associated with the route. CN is the correspondent node.

3.1 MN route request message: A MN sends this message to its current LA for the selection of a new route. The MN route request message from MN_i (i^{th} MN) contains 5 tuple as route identification (r_id=$route_i$), packet size (p_sz=P_i), number of data packet (n_pac=T_i), data rate (d_r=d_i) and time out value (t_r=t_i). In case of 100000 MNs, the number of bits require to represent a MN in binary is 17. The $route_i$ is considered as identical to MN_i who is using that route. So the number of bits to represent $route_i$ in binary is 17. The requested packet size of each MN is assumed to be in the range 8 KB to 12 KB. So the maximum number of bits require to represent packet size in binary is 4. The number of bits require to represent the number of data packets in binary is $log_2 T_i$. The requested data rate of each MN can be one of the values from the set (64 kbps, 128 kbps, 256 kbps). So the maximum possible data rate is 256 kbps and the number of bits require to represent data rate in binary is 8. The requested time out value of each MN is assumed as 60 sec. So the number of bits require to represent time out value in binary is 6. So for MN_i the length of MN route request message (rrm_bit_i) in binary is 35+$log_2(T_i)$ bits.

3.2 MN handoff request message: The MN_i sends this message to its current LA in case the remaining battery power life time becomes equal to a threshold. It contains remaining battery power life time information as user dissatisfaction. In the proposed scheme the initial battery power life time (life_time) is considered as 0.6x10^6 sec. So the maximum number of bits to represent battery power life time in binary is 20 and hence for MN_i the length of MN handoff request message (mim_bit_i) in binary is 20 bits.

3.3 Policy database at GMA, MA and SA: Each GMA, MA and SA maintains a policy database to keep the record of all the routes passing through them. The record in the database corresponding to i^{th} route contains $route_i$, route, T_i, P_i, d_i, t_i information. The $route_i$ is considered as the search key to read the record corresponding to i^{th} route from the database. All the nodes associated with i^{th} route insert record in the policy database after establishing i^{th} route, update the record in the policy database after modifying i^{th} route and delete the record from the policy database when route duration time corresponding to i^{th} route is over. Fig.1 shows the existing routes in the cellular and WLAN domain. Fig.2 and Fig.3 show the policy database maintain by all the nodes in Fig.1. GMA of a network always monitor the status of all the existing route through it in both the domain. If a route remains idle for a long time, GMA informs LA through MA and SA associated with the idle route to delete the record of the idle route from the policy database.

3.4 Operation performs by PEP at each LA: If more than one MN send MN route request or MN handoff request message to the same LA simultaneously, the PEP at LA schedules the MNs in a queue (Q_l). The MNs having MN handoff request are scheduled in the front end of the queue in the ascending order of their remaining battery power life time and the MNs having MN route request are scheduled in the rear end of the queue in the ascending order of their time out value. The queuing delay of MN_i (D_q_i) is the sum of the time requires to serve the request for (i-1) number of MNs in front of i^{th} MN in the queue.

3.5 Computation performs by PDP at each GMA, MA and SA: The PDP at each GMA, MA and SA in Fig.1 performs the following computations for each route passing through them. The

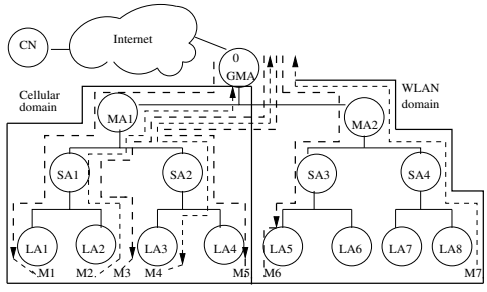


Fig. 1. Existing routes in each domain

(a)

r_id	route	n_pac	p_sz	d_r	t_r
M1	0–>1–>11–>110–>M1	T1	P1	d1	t1
M2	M2–>111–>11–>1–>0	T2	P2	d2	t2
M3	0–>1–>11–>111–>M3	T3	P3	d3	t3
M4	M4–>120–>12–>1–>0	T4	P4	d4	t4
M5	0–>1–>12–>121–>M5	T5	P5	d5	t5
M6	0–>2–>21–>210–>M6	T6	P6	d6	t6
M7	M7–>221–>22–>2–>0	T7	P7	d7	t7

(b)

r_id	route	n_pac	p_sz	d_r	t_r
M1	0–>1–>11–>110–>M1	T1	P1	d1	t1
M2	M2–>111–>11–>1–>0	T2	P2	d2	t2
M3	0–>1–>11–>111–>M3	T3	P3	d3	t3
M4	M4–>120–>12–>1–>0	T4	P4	d4	t4
M5	0–>1–>12–>121–>M5	T5	P5	d5	t5

(c)

r_id	route	n_pac	p_sz	d_r	t_r
M6	0–>2–>21–>210–>M6	T6	P6	d6	t6
M7	M7–>221–>22–>2–>0	T7	P7	d7	t7

Fig. 2. (a)Policy database at node 0, (b)Policy database at node 1, (c)Policy database at node 2

PDP at GMA, MA and SA associated with i^{th} route searches the policy database using the search key $route_i$ and reads the values of T_i, P_i, d_i, t_i for the computation of various parameters corresponding to i^{th} route. The computation performed by the PDP at all the nodes associated with i^{th} route is discussed in this section.

1. The PDP at each GMA maintains two counter for each MN. The counter H_i for MN_i indicates the number of times the existing i^{th} route is updated whereas the counter V_i for MN_i indicates the number of times the existing i^{th} route is transferred from one domain to another domain due to vertical handoff.
2. The PDP at GMA, MA and SA associated with i^{th} route computes the route duration time of i^{th} route using by MN_i ($r\_d\_time_i$) as the ratio of ($T_i * P_i$) and d_i in sec.
3. The PDP at GMA, MA and SA associated with i^{th} route computes the desired bandwidth to maintain i^{th} route ($desire\_BW_i$) for transmission or reception of the data packets of MN_i. It is computed as half of the data rate (d_i) by the concept of sampling theorem.
4. The PDP at GMA, MA and SA associated with i^{th} route computes the delay introduced by i^{th} route ($Delay_i$). It maintains two counter N_i and $count_i$ after selecting i^{th} route and after starting transmission or reception using i^{th} route. At any instant of time N_i indicates how many packets has already been transmitted or received and $count_i$ indicates how much time is elapsed to transmit or receive N_i number of data packets using the existing i^{th} route. Now time requires to transmit or receive N_i number of data packets is $(N_i * P_i)/d_i$ sec. So $Delay_i$ is computed as the difference of $count_i$ and $(N_i * P_i)/d_i$ in sec. Let i^{th} route is used also by MN_j and MN_s. The average delay introduced by i^{th} route (d_{avg_i}) is $(Delay\_i + Delay\_j + Delay\_s)/3$ sec.
5. The PDP at GMA, MA and SA associated with i^{th} route computes the number of packet loss on i^{th} route (PL_i) as the ratio of ($d_i * Delay_i$) and P_i.
6. The PDP at each GMA, MA and SA computes the time requires to select i^{th} route ($r\_s\_time_i$) for transmission or reception of T_i number of data packets of MN_i. It also computes the time requires to update i^{th} route ($r\_u\_time_i$) for

(a)

r_id	route	n_pac	p_sz	d_r	t_r
M1	0–>1–>11–>110–>M1	T1	P1	d1	t1
M2	M2–>111–>11–>1–>0	T2	P2	d2	t2
M3	0–>1–>11–>111–>M3	T3	P3	d3	t3

(b)

r_id	route	n_pac	p_sz	d_r	t_r
M4	M4–>120–>12–>1–>0	T4	P4	d4	t4
M5	0–>1–>12–>121–>M5	T5	P5	d5	t5

(c)

r_id	route	n_pac	p_sz	d_r	t_r
M6	0–>2–>21–>210–>M6	T6	P6	d6	t6

(d)

r_id	route	n_pac	p_sz	d_r	t_r
M7	M7–>221–>22–>2–>0	T7	P7	d7	t7

Fig. 3. (a)Policy database at node 11, (b)Policy database at node 12, (c)Policy database at node 21, (d)Policy database at node 22

transmission or reception of $T_i - N_i$ number of data packets of MN_i assuming that the transmission or reception of N_i number of data packets of MN_i is over using the existing route. $r\_s\_time_i$ depends upon the number of bits ($r\_s\_bit_i$) in wireless message (MN route request, D_q_i) requires to exchange among various nodes for selecting i^{th} route. It is the product of $r\_s\_bit_i$ and time/bit. $r\_u\_time_i$ depends upon the number of bits ($r\_u\_bit_i$) in wireless message (MN handoff request message, D_q_i) requires to exchange among various nodes for updating i^{th} route. It is the product of $r\_u\_bit_i$ and time/bit. Let MN_i sends MN route request message or MN handoff request message for route selection or updation to LA_id=k under the coverage area of SA_id=p, MA_id=j, and GMA_id=n. If i^{th} route is selected or updated using intra subnet function and is associated with LA_id=t, SA_id=p, MA_id=j, GMA_id=n, the computation of $r\_s\_bit_i$ and $r\_u\_bit_i$ are as follows:

The number of MN route request message exchange among the nodes (MN,k,p,j,n,t,CN) is 6 (as MN->k->p->t, p->j->n->CN) and the number of D_q_i exchange among the nodes (k,p,j,n) is 3 (as k->p->j->n). The nodes p,j,n insert the record corresponding to i^{th} route of MN_i in policy database. $r\_s\_bit_i = 6 * rrm\_bit_i + 3 * log_2(D\_q_i)$.

The number of MN handoff request message exchange among the nodes (MN,k,p,j,n,t) is 5 (as MN->k->p->t, p->j->n) and the number of D_q_i exchange among the nodes (k,p,j,n) is 3 (as k->p->j->n). The nodes p,j,n update the record corresponding to i^{th} route of MN_i in policy database. $r\_u\_bit_i = 5 * mim\_bit_i + 3 * log_2(D\_q_i)$.

If i^{th} route is selected or updated using intra domain function and is associated with LA_id=m, SA_id=s, MA_id=j, GMA_id=n, the computation of $r\_s\_bit_i$ and $r\_u\_bit_i$ are as follows:

The number of MN route request message exchange among the nodes (MN,k,p,j,s,m,n,CN) is 7 (as MN->k->p->j->s->m, j->n->CN) and the number of D_q_i exchange among the nodes (k,p,j,s,n) is 4 (as k->p->j->s, j->n). The nodes (s,j,n) insert the record corresponding to i^{th} route of MN_i in policy database. $r\_s\_bit_i = 7 * rrm\_bit_i + 4 * log_2(D\_q_i)$.

The number of MN handoff request message exchange among the nodes (MN,k,p,j,s,m,n) is 6 (as MN->k->p->j->s->m, j->n) and the number of D_q_i exchange among the nodes (k,p,j,s,n) is 4 (as k->p->j->s, j->n). The nodes (j,n) update and the node s inserts the record corresponding to i^{th} route of MN_i in policy database. $r\_u\_bit_i = 6 * mim\_bit_i + 4 * log_2(D\_q_i)$.

If the route is selected or updated using inter domain function and is associated with LA_id=a, SA_id=b, MA_id=r, GMA_id=n, the computation of $r\_s\_bit_i$ and $r\_u\_bit_i$ are as follows:

The number of MN route request message exchange among the nodes (MN,k,p,j,n,r,b,a,CN) is 8 (as MN->k->p->j->n->r->b->a, n->CN) and the number of D_q_i exchange among the nodes (k,p,j,n,r,b) is 5 (as k->p->j->n->r->b). The nodes b,r,n insert the record corresponding to i^{th} route of MN_i in policy database. $r\_s\_bit_i = 8 * rrm\_bit_i + 5 * log_2(D\_q_i)$.

The number of MN handoff request message exchange among the nodes (MN,k,p,j,n,r,b,a) is 7 (as MN->k->p->j->n->r->b->a) and the number of D_q_i exchange among the nodes (k,p,j,n,r,b) is 5 (as k->p->j->n->r->b). The node n updates and the nodes (r,b) insert the record corresponding to i^{th} route of MN_i in policy database. $r\_u\_bit_i = 7 * mim\_bit_i + 5 * log_2(D\_q_i)$. Finally $r\_s\_time_i = r\_s\_bit_i * time/bit$ and $r\_u\_time_i = r\_u\_bit_i * time/bit$.

7. The PDP at GMA associated with i^{th} route computes the time requires to transfer the existing i^{th} route from one domain to other domain ($r\_nv\_time_i$) due to network initiated vertical handoff. It depends upon the number of bits ($r\_nv\_bit_i$) in wireless message (MN route request message as the route is selected in a new domain, D_q_i) requires to exchange among various nodes for transferring i^{th} route from one domain to other domain. It is the product of $r\_nv\_bit_i$ and time/bit. Let i^{th} route is transferred from the domain having MA_id=j to the domain having MA_id=r and it is associated with LA_id=c, SA_id=d, MA_id=r, GMA_id=n. The number of MN route request message exchange among the nodes (c,d,r,n) is 3 (as n->r->d->c) and the number of D_q_i exchange among the nodes (d,r,n) is 2 (as n->r->d). So $r\_nv\_bit_i = 3 * rrm\_bit_i + 2 * log_2(D\_q_i)$ and $r\_nv\_time_i = r\_nv\_bit_i * time/bit$. The node n updates and the nodes (r,d) insert the record corresponding to i^{th} route of MN_i in policy database.

8. The PDP at GMA of each network computes congestion per MN, excess cost per MN and maximum allowable delay per MN during transmission or reception of the data packets dynamically. The congestion for MN_i (C_i) is $Delay_i + D\_q_i + (H_i * r\_u\_bit_i + V_i * r\_nv\_bit_i) * time/bit$. In case of congestion in the network MN uses the network facility for longer period of time which incurs huge cost. So the excess cost for MN_i (δC_i) is assumed as proportional to C_i. A queue Q_c is maintained at GMA to schedule the MNs continuing transmission or reception in the descending order of their excess cost. The maximum allowable delay for MN_i is assumed as half of the sum of time out value and remaining battery power life time. A queue Q_d is maintained at GMA to schedule the MNs continuing transmission or reception in the ascending order of their maximum allowable delay.

9.The PDP at each GMA computes average Cost ($Cost_n$) and average Delay ($Delay_n$) of the network. $Cost_n$ ($Delay_n$) is the ratio of the sum of cost (delay) to maintain all the routes R_n in the network and R_n. The cost (delay) to maintain i^{th} route in the network is the product of the number of bits require to exchange and cost/bit (time/bit). Now the number of bits require to exchange (bit_i) is the sum of the number of bits require to select i^{th} route, number of bits require to transmit or receive using i^{th} route, number of bits require to update i^{th} route and number of bits require to transfer i^{th} route.

$bit_i = r\_s\_bit_i + T_i * P_i + H_i * r\_u\_bit_i + V_i * r\_nv\_bit_i$

$Cost_n = (\sum_{i=1}^{R_n} bit_i * (cost/bit))/R_n$ unit, where cost/bit is assumed as 1 unit for WLAN domain and 2 unit for cellular domain.

The delay of MN_i (D_i) is $(bit_i - T_i * P_i) * time/bit + D\_q_i + Delay_i$ in msec. So $Delay_n = (\sum_{i=1}^{R_n} D_i)/R_n$ in msec.

Considering wireless network data rate as 30 kbps, [Reference websites are as follows: http://www.computerworld.com/mobiletopics/mobile /technology/story/0,10801,87486,00.html

http://www.ee.latrobe.edu.au/mf/ELE52PMC /Lectures/PMC-06.ppt] the bit/time is assumed as 30 kbps in the present work.
10. The PDP at GMA associated with i^{th} route computes the remaining battery power life time of MN_i (r_i) dynamically. The computation of r_i after transmission or reception of N_i number of packets is as follows: Time requires for transmitting or receiving N_i number of packets of MN_i ($time_i$) using i^{th} route is $D_i + (N_i * P_i)/d_i$ $r_i = life\_time - time_i$
10. The PDP at each GMA computes the power consumption of a domain periodically as the sum of power consumption to maintain upload and download using the routes in that domain. Now the power consumption of i^{th} route using by MN_i (PC_i) is the product of ($r\_d\_time_i + D_i$) and consumption rate (C). Again C is equal to the ratio of initial battery power of a MN and life time. C is assumed as 5 mJ/sec. So the average power consumption of the network (P_n) is $(\sum_{i=1}^{R_n} PC_i)/R_n$, where R_n is the number of routes in the network.
11.The PDP at each GMA computes average throughput (T_n) of the network. The time requires to complete the transmission or reception of T_i number of packets of MN_i (τ_i) is $bit_i * time/bit + Delay_i + D\_q_i$. The throughput of i^{th} route in the network (TH_i) is the ratio of T_i and τ_i in packets/min. $T_n = (\sum_{i=1}^{R_n} TH_i)/R_n$

3.6 Computation performs by MN_i: Like all other nodes associated with i^{th} route, MN_i computes $Delay_i$, $r\_s\_time_i$ in case of route selection and $r\_u\_time_i$ in case of route updation. Knowing D_q_i, $Delay_i$ and $r\_s\_time_i$ or $r\_u\_time_i$, MN_i computes $time_i$ and r_i. r_{i_th} is the threshold value of r_i and it is assumed as $(T_i - N_i)/d_i + Delay_i + D\_q_i$. If r_i is equal to r_{i_th}, MN_i sends MN handoff request message mentioning r_i to its current LA for transmission or reception of $T_i - N_i$ number of data packets.

3.7 Computation of quality function: The PDP at each SA computes the quality function of all the LAs under it, the PDP at each MA computes the quality function of all the SAs under it and the PDP at each GMA computes the quality function of all the MAs under it. $R_{k\_p\_j\_n}$ is assumed as the total number of routes passing through LA_id=k, SA_id=p, MA_id=j, GMA_id=n and i^{th} route using by MN_i for transmission or reception of data packets is assumed as one of the routes out of $R_{k\_p\_j\_n}$.

Computation of quality function by PDP at each SA: The PDP at each SA corresponding to each subnet computes three parameters as average local delay, average local packet loss and local unused bandwidth of all the LAs under it. At any instant of time SA of each subnet computes the above three parameters for LA(S) associated with the record which is inserted or updated in the policy database at that SA. The computation of the above three parameters for LA_id=k under the coverage area of SA_id=p, MA_id=j and GMA_id=n are discussed below:
The average local delay at LA_id=k ($av\_del_{k\_p\_j\_n}$) is the ratio of the sum of delay of all the routes $R_{k\_p\_j\_n}$ and $R_{k\_p\_j\_n}$. So $av\_del_{k\_p\_j\_n} = (\sum_{i=1}^{R_{k\_p\_j\_n}} Delay_i)/R_{k\_p\_j\_n}$.
The average local packet loss at LA_id=k ($av\_pac_{k\_p\_j\_n}$) is the ratio of the sum of packet loss of all the routes $R_{k\_p\_j\_n}$ and $R_{k\_p\_j\_n}$. So $av\_pac_{k\_p\_j\_n} = (\sum_{i=1}^{R_{k\_p\_j\_n}} PL_i)/R_{k\_p\_j\_n}$.
The local unused bandwidth at LA_id=k ($un\_BW_{k\_p\_j\_n}$) is the difference of the available bandwidth ($av\_BW_{k\_p\_j\_n}$) and used bandwidth ($u\_BW_{k\_p\_j\_n}$) at the same LA. Now $u\_BW_{k\_p\_j\_n}$ is the sum of the desired bandwidth of all the routes $R_{k\_p\_j\_n}$. So $u\_BW_{k\_p\_j\_n} = \sum_{i=1}^{R_{k\_p\_j\_n}} desire\_BW_i$.
Finally the PDP of each SA corresponding to each subnet computes the quality function of LA(s). The quality function of LA_id=k ($Q_{k\_p\_j\_n}$) is the ratio of $un\_BW_{k\_p\_j\_n}$ and $(av\_del_{k\_p\_j\_n} * av\_pac_{k\_p\_j\_n})$.

Computation of quality function by PDP at each MA: The PDP at each MA corresponding to each domain computes three parameters as average subnet delay, average subnet packet loss and subnet unused bandwidth of all the SAs under it. At any instant of time MA of each domain computes the above three parameters for SA(S) associated with the record which is inserted or updated in the policy database at that MA. The computation of the above three parameters for SA_id=p under the coverage area of MA_id=j and GMA_id=n are discussed below:
The average subnet delay at SA_id=p ($avg\_del_{p\_j\_n}$) is the ratio of the sum of delay of all the LAs ($no\_of\_LA_{p\_j\_n}$) under the coverage area of SA_id=p to $no\_of\_LA_{p\_j\_n}$. The delay at LA_id=k (D_k) under the coverage area of SA_id=p is computed as $\sum_{i=1}^{R_{k\_p\_j\_n}} Delay_i$. So $avg\_del_{p\_j\_n} = (\sum_{k=1}^{no\_of\_LA_{p\_j\_n}} D_k)/no\_of\_LA_{p\_j\_n}$.
The average subnet packet loss at SA_id=p ($avg\_pac_{p\_j\_n}$) is the ratio of the sum of packet loss of all the LAs $no\_of\_LA_{p\_j\_n}$ under the coverage area of SA_id=p and $no\_of\_LA_{p\_j\_n}$. The packet loss at LA_id=k (P_k) under the coverage area of SA_id=p is computed as $\sum_{i=1}^{R_{k\_p\_j\_n}} PL_i$. So $avg\_pac_{p\_j\_n} = (\sum_{k=1}^{no\_of\_LA_{p\_j\_n}} P_k)/no\_of\_LA_{p\_j\_n}$.
The subnet unused bandwidth at SA_id=p ($un\_BW_{p\_j\_n}$) is the sum of unused bandwidth of all the LAs $no\_of\_LA_{p\_j\_n}$ under the coverage area of SA_id=p. The unused bandwidth at LA_id=k (B_k) under the coverage area of SA_id=p is computed as $av\_BW_{k\_p\_j\_n} - u\_BW_{k\_p\_j\_n}$. So $un\_BW_{p\_j\_n} = (\sum_{k=1}^{no\_of\_LA_{p\_j\_n}} B_k)$.
Finally the PDP of each MA corresponding to each domain computes the quality function of SA(s). The quality function of SA_id=p ($Q_{p\_j\_n}$) is the ratio of $un\_BW_{p\_j\_n}$ and $(av\_del_{p\_j\_n} * av\_pac_{p\_j\_n})$.

Computation of quality function by PDP at each GMA: The PDP at each GMA corresponding to each network computes three parameters as average domain delay, average domain packet loss and domain unused bandwidth of all the MAs under it. At any instant of time GMA of each network computes the above three parameters for MA(S) associated with the record which is inserted or updated in the policy database at that GMA. The computation of the above three parameters for MA_id=j under the coverage area of GMA_id=n are discussed below:
The average domain delay at MA_id=j ($avg\_del_{j\_n}$) is the ratio of the sum of delay of all the SAs ($no\_of\_SA_{j\_n}$) under the coverage area of MA_id=j and $no\_of\_SA_{j\_n}$. The delay at SA_id=p (D_p) under the coverage area of MA_id=j is computed as $\sum_{k=1}^{no\_of\_LA_{p\_j\_n}} D_k$. So

$avg\_del_{j\_n} = (\sum_{p=1}^{no\_of\_SA_{j\_n}} D_p)/no\_of\_SA_{j\_n}$.

The average domain packet loss at MA_id=j ($avg\_pac_{j\_n}$) is the ratio of the sum of packet loss of all the SAs $no\_of\_SA_{j\_n}$ under the coverage area of MA_id=j and $no\_of\_SA_{j\_n}$. The packet loss at SA_id=p (P_p) under the coverage area of MA_id=j is computed as $\sum_{k=1}^{no\_of\_LA_{p\_j\_n}} P_k$.

So $avg\_pac_{j\_n} = (\sum_{p=1}^{no\_of\_SA_{j\_n}} P_p)/no\_of\_SA_{j\_n}$.

The domain unused bandwidth at MA_id=j ($un\_BW_{j\_n}$) is the sum of unused bandwidth of all the SAs $no\_of\_SA_{j\_n}$ under the coverage area of MA_id=j. The unused bandwidth at SA_id=p (B_p) under the coverage area of MA_id=j is computed as $\sum_{k=1}^{no\_of\_LA_{p\_j\_n}} av\_BW_{k\_p\_j\_n} - u\_BW_{k\_p\_j\_n}$. So $un\_BW_{j\_n} = (\sum_{p=1}^{no\_of\_SA_{j\_n}} B_p)$.

Finally the PDP of each GMA corresponding to each network computes the quality function of MA(s). The quality function of MA_id=j (Q_{j_n}) is the ratio of $un\_BW_{j\_n}$ and ($av\_del_{j\_n} * av\_pac_{j\_n}$).

IV. PBAC (PBHD) algorithm

The PDP at each SA uses intra_subnet function to select LA_id=t under the coverage area of the corresponding SA wherein the Policy I is satisfied. The PDP at each MA uses intra_domain function to select SA_id=s under the coverage area of the corresponding MA wherein the Policy II is satisfied. The PDP at each GMA uses inter_domain function to select MA_id=r under the coverage area of the corresponding GMA wherein the Policy III is satisfied. Policy I, Policy II and Policy III are as follows:

Policy I for LA_id=t, SA_id=p, MA_id=j, GMA_id=n: (a)$Q_{t\_p\_j\_n}$ is maximum, (b)$desire\_BW_i < un\_BW_{t\_p\_j\_n}$, (c)$d_{avg\_i} + y + r\_d\_time_i + D\_q_i \leq x$, where x is t_i, y is $r\_s\_time_i$ and $r\_d\_time_i$ is $T_i * P_i/d_i$ in case of PBAC algorithm. x is r_i, y is $r\_u\_time_i$ and $r\_d\_time_i$ is $(T_i - N_i)P_i/d_i$ in case of PBHD algorithm.

Policy II for SA_id=s, MA_id=j, GMA_id=n: (a)$Q_{s\_j\_n}$ is maximum, (b)$desire\_BW_i < un\_BW_{s\_j\_n}$

Policy III for MA_id=r, GMA_id=n: (a)Q_{r_n} is maximum, (b)$desire\_BW_i < un\_BW_{r\_n}$

Let MN_i sends MN route request (MN handoff request) message to select (update) i^{th} route to its current LA having LA_id=k under the coverage area of SA having SA_id=p, MA having MA_id=j and GMA having GMA_id=n. The function performed by the PEP at LA_id=k for MN_i in front of Q_l is as follows:

It sends MN route request (MN handoff request) message along with D_q_i to the PDP at SA_id to trigger intra_subnet(SA_id,MA_id,GMA_id) function where SA_id=p, MA_id=j, GMA_id=n.

intra_subnet(SA_id,MA_id,GMA_id): The PDP at SA_id searches for LA_id=t under its coverage area wherein Policy I is satisfied. If found, new LA_id=t. The PDP at SA_id selects (updates) i^{th} route passing through LA_id, SA_id, MA_id, GMA_id. It sends MN route request (MN handoff request) message along with D_q_i to GMA_id through MA_id and MN route request (MN handoff request) message to LA_id. GMA_id, MA_id, SA_id inserts (updates) record corresponding to i^{th} route in policy database in case of PBAC (PBHD) algorithm. In case of PBHD algorithm T_i is replaced by $T_i - N_i$ in policy database at GMA_id, MA_id and SA_id in case transmission or reception of N_i number of data packets is already over using the existing i^{th} route. The PDP at GMA_id increases C_{new} ($C_{handoff}$) as new call (handoff call) request counter by 1. Otherwise the PEP at SA_id sends MN route request (MN handoff request) message along with D_q_i to PDP at MA_id to trigger intra_domain(MA_id,GMA_id) function where MA_id=j and GMA_id=n.

intra_domain(MA_id,GMA_id): The PDP at MA_id searches for SA_id=s under its coverage area wherein Policy II is satisfied. If found, new SA_id=s. The PEP at MA_id sends MN route request (MN handoff request) message along with D_q_i to PDP at SA_id to trigger intra_subnet(SA_id,MA_id,GMA_id) function where SA_id=s, MA_id=j and GMA_id=n. Otherwise the PEP at MA_id sends MN route request (MN handoff request) message along with D_q_i to PDP at GMA_id to trigger inter_domain(GMA_id) function where GMA_id=n.

inter_domain(GMA_id): The PDP at GMA_id searches for MA_id=r under its coverage area wherein Policy III is satisfied. If found, new MA_id=r. The PEP at GMA_id sends MN route request (MN handoff request) message along with D_q_i to PDP at MA_id to trigger intra_domain(MA_id,GMA_id) function where MA_id=r and GMA_id=n. Otherwise in case of PBAC algorithm the PDP at GMA_id increases call block counter (b_c) by 1. It also computes call blocking probability (CBP) as the ratio of b_c and $C_{new} + b_c$. In case of PBHD algorithm the PDP at GMA_id increases call drop counter (d_c) by 1. It computes call dropping probability (CDP) as the ratio of (d_c) and $C_{handoff} + d_c$. The PDP at GMA_id also computes satisfaction of MN_i as N_i/T_i and satisfaction of the network (N_s) as $(\sum_{i=1}^{U} N_i)/(\sum_{i=1}^{U} T_i)$.

V. Network initiated vertical handoff algorithm

If the power consumption of a domain crosses a threshold (P_{th}), the PDP at GMA triggers vertical handoff algorithm. P_{th} is assumed as 0.6 watt in case of download and 1 watt in case of upload.

If the power consumption of the cellular domain corresponding to upload (download) crosses its threshold, the PDP at GMA searches both Q_c, Q_d for MNs having maximum excess cost or minimum allowable delay during transmission (reception) of data packets and repeats func(GMA_id,MA_id) for such MNs to trigger vertical handoff from cellular domain to WLAN domain till the power consumption of cellular domain corresponding to upload (download) goes below its threshold.

If the power consumption of the WLAN domain corresponding to upload (download) crosses its threshold, the PDP at GMA searches both Q_c, Q_d for MNs having minimum excess cost or maximum allowable delay during transmission (reception) of data packets and repeats func(GMA_id,MA_id) for such MNs to trigger vertical handoff from WLAN domain to cellular domain till the power consumption of WLAN domain corresponding to upload (download) goes below its threshold. The PDP at GMA also identifies MNs having long idle period and initiates vertical handoff from WLAN domain to cellular domain for mobility management [1] of such MNs as in idle mode the power consumption in WLAN domain is almost 9 times higher than that in cellular domain.

The func(GMA_id,MA_id) for MN_i where GMA_id=n and MA_id=WLAN domain in case of vertical handoff from cellular domain to WLAN domain whereas with GMA_id=n and MA_id=cellular domain in case of vertical handoff from WLAN

domain to cellular domain is discussed below.

func(GMA_id,MA_id):

Step 1: The PEP at GMA_id sends MN route request message along with D_q_i to MA_id.

Step 2: The PDP at MA_id searches for SA_id=u under its coverage area wherein Policy II is satisfied.

Step 3: The PEP at MA_id sends MN route request message along with D_q_i to SA_id=u.

Step 4: The PDP at SA_id searches for LA_id=v under its coverage area wherein Policy I is satisfied.

Step 5: The PEP at SA_id sends MN route request message to LA_id=v.

Step 6: i^{th} route passing through SA_id=u and LA_id=v under the coverage area of GMA_id and MA_id is selected for transmission or reception of the remaining data packets of MN_i.

VI. Simulation result

In this section the simulation results are considered for discussion. The curves are plotted as a function of traffic load in the network. The traffic load is computed as the ratio of arrival rate and departure rate of service request. In the present work both the arrival rate and departure rate of service request is considered as poisson distribution. Fig.4 shows the plot of CBP and CDP vs. traffic load. From the plot it can be observed that both CBP and CDP increase with traffic load but CDP is lesser than CBP due to high priority of handoff call over new call. Fig.5 shows the plot of N_s vs. traffic load. From the plot it can be observed that N_s reduces with traffic load. Fig.6 and Fig.7 show the plot of $Delay_n$ and $Cost_n$ vs. traffic load. From the plot it can be observed that both $Delay_n$ and $Cost_n$ increase with traffic load. Fig.8 shows the plot of the average throughput of the network (T_n) vs. traffic load. From the plot it can be observed that after traffic load 10^2 T_n becomes almost constant.

The network congestion increases with traffic load. As a result CBP, CDP, $Delay_n$, $Cost_n$ increases with traffic load and N_s, T_n reduces with traffic load. But when traffic load is greater than equal to 10^2, the PDP at GMA starts to trigger vertical handoff algorithm due to huge power consumption. In case the power consumption of the cellular domain crosses its threshold GMA triggers vertical handoff from cellular domain to WLAN domain. After such vertical handoff CBP, CDP, $Delay_n$, $Cost_n$ reduces and N_s, T_n increases. In case the power consumption of the WLAN domain crosses its threshold GMA triggers vertical handoff from WLAN domain to cellular domain. After such vertical handoff CBP, CDP, $Delay_n$, $Cost_n$ increases and N_s, T_n reduces. Moreover the vertical handoff algorithm tries to balance the power consumption of the two domain which in turn balances the effective load of the two domain. As a result the rate of change of CBP, CDP, N_s, $Delay_n$, $Cost_n$ and T_n with traffic load after traffic load equal to 10^2 lies within a limit.

VII. Conclusion

The present work is a dynamic resource management scheme for next generation all-IP wireless network. It maintains the average value of CBP, CDP, delay, cost, network satisfaction and throughput almost constant by balancing the effective load of the two domain in the network. The performance of the scheme is evaluated considering only the data class of traffic. It can be extended by evaluating its performance in presence of other traffic classes such as voice and video in the network.

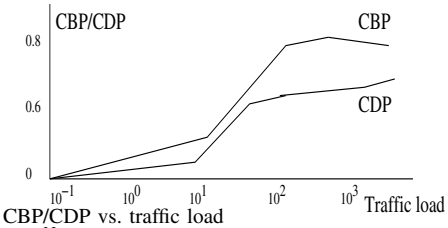


Fig. 4. CBP/CDP vs. traffic load

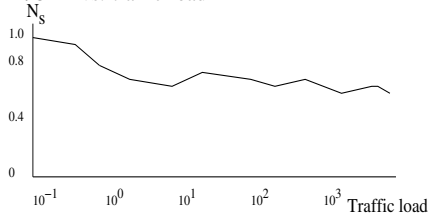


Fig. 5. Network satisfaction vs. traffic load

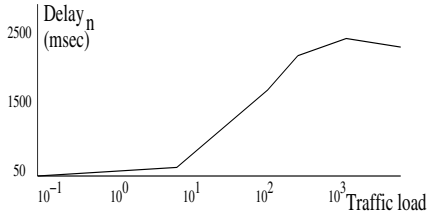


Fig. 6. $Delay_n$ vs. traffic load

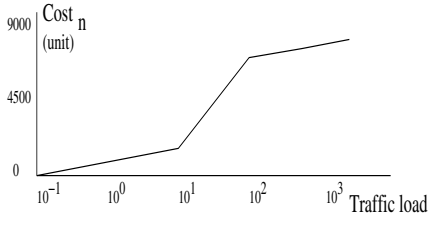


Fig. 7. $Cost_n$ vs. traffic load

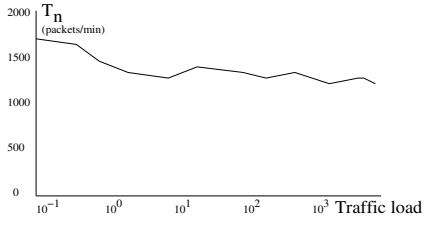


Fig. 8. T_n vs. traffic load

References

[1] S.Mitra, 'Dynamic Mobility Management for Next Generation All-IP Wireless Network' AsiaCSN 2008.

[2] 3GPP TR 23.234 v7.1.0, '3GPP System to WLAN Interworking; System Description (Release 7) March 2006', http://www.3gpp.org/specs/specs.htm.

[3] J.McNair and F.Zhu, 'Vertical Handoffs in Fourth-Generation Multinetwork Environments', IEEE Wireless Communications Magazine, June 2004.

[4] Chuanxiong Guo, Zihua Guo, Qian Zhang and Wenwu Zhu, 'A seamless and proactive End-to-End Mobility Solution for Roaming across Heterogeneous Wireless Networks', IEEE Journal on Selected Areas in Communications, vol.22, no.5, pp.834-848, June 2004.

[5] N.Nasser, A.Hasswa and H.Hassanein, 'Handoffs in Fourth Generation Heterogeneous Networks', IEEE Communications Magazine, vol.44, no.10, pp. 96-103, October 2006.

[6] T.B.Zahariadis, 'Guest Editorial:Migration toward 4G Wireless Communications', IEEE Wireless Communications Magazine, June 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Routing in Mobile Ad hoc Networks: Cases of Long-hop and Short-hop

Tanveer Ahmed Bhuiyan and Mohammed Tarique
Department of Electrical and Electronic Engineering
American International University-Bangladesh
Banani, Dhaka
E-mail:tariquemohammed@aiub.edu

Rumana Islam
East West University
Mohakhali, Dhaka
E-mail:runa712000@yahoo.com

***Abstract*—A fundamental problem of Mobile Ad hoc Network (MANET) is to determine whether it is advantageous to route packet over many short-hops or few long-hops. The main arguments for short-hop routing are minimum energy consumption and high signal-to-interference ratio. Minimum energy consumption maximizes network life. High signal-to-interference ratio maximizes network capacity. On the other hand, main arguments for using long-hop routing are low end-to-end packet delay and high reliability against node mobility. Low end-to-end delay increases network capacity. High reliability reduces packet loss in the network. In this paper, we shed more lights on these issues by listing the network parameters that need to be considered while choosing short-hop or long-hop routing. The simulation results show that short-hop routing is a good candidate to maximize network life and to minimize packet loss. On the other than, long-hop routing is preferable to ensure network connectivity and to minimize end-to-end packet delay.**

I. Introduction

In the recent years, Mobile Ad hoc Network (MANET) is considered as a suitable means of providing instant networking to a group of mobile nodes. MANET is self-organizing and self-configuring. No centralized administration is required to operate and maintain the network. No infrastructure is needed to set-up a MANET. For this reason, MANET is considered a suitable choice of networking where infrastructure has been destroyed by natural calamity. MANET is also considered a suitable choice of networking where infrastructure based network is hard to build. In MANET, mobile nodes communicate with each other in a multi-hop fashion. That means a source mobile node sends packet to a destination mobile node via other mobile nodes located in between them. Routing protocol is the most important element of MANET. One of the fundamental problem of multi-hop communication is to decide whether mobile node should use many short-hops or few long-hops. The length of hop is determined by the transmission power level of mobile node. Higher transmission power means longer hop. Lower transmission power means shorter hop. The long-hop and short-hop issues have been addressed by researchers since the inception of packet radio network [1]. A crude analysis in [2] shows that transmission range should be small, but not too small so that there will be dis-connectivity in the network. Another work [4] has shown that there is a trade-off between the transmission range and the network bandwidth. Higher transmission range means loss of bandwidth. The analysis presented in [4] also shows that a mobile node should adjust the transmission power so that it will have six neighbors.This work has been extended by [3]. The authors concluded in [3] that mobile nodes should adjust transmission ranges so that they will have number of neighbors is on the order of eight [3]. A similar work [5] has proved that the critical transmission range of a mobile node in a MANET should be minimum to a level that ensures network connectivity. Another related work [7] shows that the mobile node should adjust transmission power maximize its battery life. According to the scheme presented in [7] mobile nodes formed group (i.e. cluster). Mobile nodes adjust their transmission ranges so that they can reach to the furthest node in a given cluster. One common limitations of all these power transmission strategies is that the mobile nodes are assumed to be static. But mobility has been considered in [8]. In this paper, the author claimed that the transmission ranges of the nodes should be adjusted so that maximum number of packets can be delivered to the destination under mobility conditions. Simulation results presented in this paper show that the number of neighbors should be increased when the node mobility is increased in the network. But there is no optimum number of neighbors that can ensure the maximum number of packet delivery to the destination. A very recent work presented in [9] shows that long-hop routing is better. The authors discussed 18 cases in this paper where long-hop routing protocol shows better performance compared to short hop routing. The major limitation of this work is that the authors claims are based on laboratory experiments of a small network consisting on 10 sensor nodes. For this kind of small network, it is hard to judge whether short-hop or long-hop routing is better. Because the performance of a routing protocol varies with the network size.

In this paper, we investigate the effects of long-hop and short-hop routing on the performances of large MANET. The objective is to discover the cases when we should use short-hop routing and when we should use long-hop routing. Specifically, we focus on network performances in terms of overhead control messages, energy consumption, packet loss, end-to-end delay and node mobility. Overhead control

messages generated in the network occupy bandwidth and hence affect the performance of network. Overhead control messages are originated from the route discovery mechanism used by the routing protocol. Reactive routing protocols like Dynamic Source Routing (DSR) protocol uses a 'flooding' technique during the route discovery process [6]. Although overhead control messages generated in the network is not very significant when network size is small, but it is shown in [10] a huge number of overhead packets is generated in the network when network size large. In this paper, we investigate the impact of transmission ranges on the overhead control messages generated in the network. Energy consumption is another important issue in MANET. Energy consumption determines network life. We investigate dependency of energy consumption on transmission range. Packet loss in ad hoc network is another major issue. There are two main reasons of packet loss in the network (1) packet collisions due to simultaneous packet transmissions by different mobile nodes, and (2) packet drop due to limited buffer size of mobile node. In this paper we will relate packet loss with the transmission ranges of mobile node. End-to-end delay of packet is another important design parameter of MANET. The end-to-end delay per packet depends on the traffic conditions of the network as well as the number of hops a packet travels from source to destination. Hence end-to-end delay of packet also depends on the transmission ranges of the mobile nodes. We will investigate the impact of transmission ranges on the end-to-end delay of packets also. Node mobility affects routing decision and also packet loss in MANET. High mobility increases route 'breakage' rate and it also increases packet loss. The route breakage also depends on the transmission range. To investigate all these issues of network performances, we choose DSR protocol as the routing protocol of the network. While it is likely that network performances will vary with the routing protocol used , the results obtained with DSR protocol can be generalized to most on-demand ad hoc routing protocols.

II. The DSR protocol

The DSR protocol consists of two main mechanisms :(1) route discovery, and (2) route maintenance. Route discovery is the mechanism by which a source node discovers a route to a destination. During route discovery process, a source node initiates the route discovery by broadcasting a request message to its neighbors. When the neighboring nodes receive the request packet, they add their addresses in the request packet and re-broadcast that request message. This process goes on until the request packet is received by the destinaton node. A route discovery mechanism is shown in Figure 1. In that figure, node A is attempting to discover a route to node E. To initiate the route discovery, node A transmits a 'route request' packet as a single local broadcast packet, which is received by all nodes currently within wireless transmission range of A, including node B in this example. Each Route Request identifies the initiator and target of the route discovery, and also contains a unique request identification (i.e., 2 in this example), determined by the initiator of the request. Each

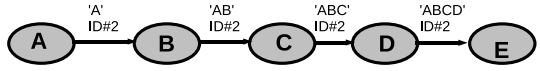


Fig. 1. Route discovery in DSR protocol

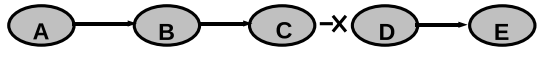


Fig. 2. Route maintenance in DSR protocol

route request also contains a listing of the address of each intermediate node through which this particular copy of the route request has been forwarded. When another node receives this route request (i.e., node B in this example), if it is the target of the route discovery, it returns a 'route reply' to the initiator of the route discovery, giving a copy of the accumulated route record from the route request. When the initiator receives this route reply, it caches this route in its route cache for use in sending subsequent packets to this destination. Otherwise, if this node receiving the route request has recently seen another route request message from this initiator bearing this same request identification and target address or if this nodes own address is already listed in the route record of the route request, this node discards the request. Otherwise, this node appends its own address to the route record in the route request and propagates it by transmitting it as a local broadcast packet (with the same request identification). In this example, node B re-broadcast the route request, which is received by node C; nodes C and D each also, in turn, re-broadcast the request, resulting in the request packet being received by node E.

Route maintenance is a mechanism by which a node is able to detect changes in the network topology. While originating or forwarding a packet using a source route, each node transmitting the packet is responsible for confirming that data can flow over the link from that node to the next hop. For example, in the network scenario as shown in Figure 2 Node A has originated a packet for node E using a source route through intermediate nodes B, C, and D. In this case, each node is responsible to monitor the link between itself to the next hop. For example, node A is responsible for the link from A to B. An acknowledgement can provide confirmation that a link is capable of carrying data, and in wireless networks,

acknowledgements are often provided by an existing standard part of the MAC protocol such as IEEE 802.11 [20]. In this example, when node C detects that the link between itself to node D is broken, node C creates a route error message and sends that packet to node A. After receiving the route error message, node A marks the route as 'invalid' in the route cache and tries to find an alternative route to the destination node E. If no such route is found in the route cache, node A initiates a new route discovery process.

III. Reasons for Routing over long hops and short hops

The performance of routing protocol like DSR depends on the transmission ranges of the mobile nodes. In order to investigate how transmission range of mobile node affect the network performances, we focus on the following performance metrics: routing overhead packets generated in the network, energy consumption, packet loss, mobility and end-to-end delay.

A. Routing overhead

Routing overhead packets generated in the network occupies a significant portion of network bandwidth specially when the network size is large. Hence a considerable portion of network bandwidth is occupied by the overhead packets. The objective will be to reduce the number of overhead packets so that network bandwidth is used more efficiently by useful data packet. The overhead control messages are genereated from the route discovery process of reactive routing protocol like DSR. During this route discovery process a mobile node is obliged to re-broadcast a request message when it receives it from other nodes. This kind of re-broadcasting is called 'flooding'. Flooding can adversely affect the performance of a network when network size is large. The main problems of flooding are (1) contention, (2) collision, and (3) redundancy. A detailed analysis of these problems can be found in [11]. Many routing protocols have been proposed in [11], [12], [13], [14] and [16] to reduce the 'flooding' problem. But the 'flooding' problem can be reduced by increasing the transmission range of mobile node as well. If the transmission range of mobile node is increased, the route request packet will travel for less number of hop from the source to the destination. Since route request packet travels less number of hops, there will be less number of re-broadcasting of a request packet. To investigate how the overhead control packet generated in the network varies with respect to different transmission ranges, we simulated a network consisting of 100 mobile nodes in Network Simulator (NS-2) [15]. Those nodes were placed randomly over an area of 1000m $\times$ 1000m. Ten connections were set up randomly in the network. While setting up each connection, each source initiates a route discovery process. Once conenction is set-up, Constant Bit Rate (CBR) agent was used to generate packets and the packet genaration rate was 2 packets per second. Each simulation was tested for 250 seconds simulation time. IEEE MAC layer was used as the MAC layer. We increased the network size by keeping the node density constant so that the network connectivity is not affected. We increased the network area to 1500m $\times$ 1000m when we increased the number of nodes to 150 nodes. When the network area was 2000m $\times$ 1000m, we placed 200 mobile nodes in the network. The total number of overhead control packet generated in the network and the total number of data packet delivered to the destination were monitored during each simulation. We measured the overhead per packet as the ratio of the total overhead packets generated in the network and the total number of data packets delivered to the destination. We created ten different topologies by using random variables. Then we repeated the test by using three different transmission ranges of 200m, 250m and 300m to examine the effect of transmission ranges on overhead packet. The results are shown in Figure 3.

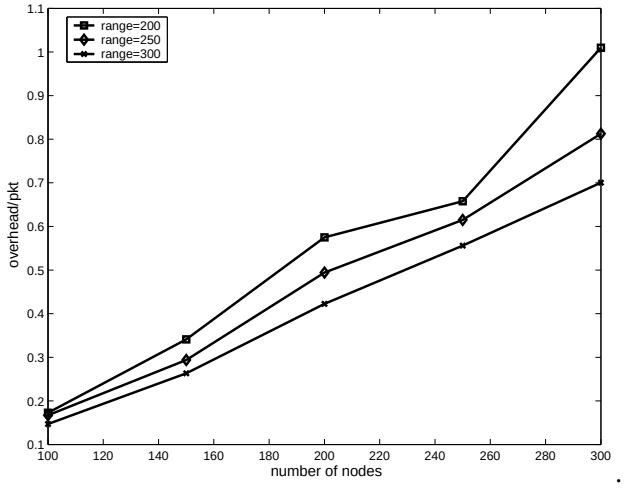


Fig. 3. Overhead packets with different transmission ranges

It is depicted from this figure that the overhead packet generated in the network decreases with the increase in transmission range. But for small network the overhead per data packet is almost same irrespective of different transmission ranges. For example, overhead per data packet is almost 0.2 for transmission ranges of 200m, 250m and 300m. When the network size is large, for exmaple 200, the overhead per packet are 0.45 and 0.55 when transmission ranges are 250m and 200m respectively. So there is a 10 percent reduction of overhead when we increased the transmission range by 50 meter. But the difference in the overhead packet becomes more visible when we further increased the network size. For example, when the network is the largest (i.e. 300 nodes), there is a significant amount of decrease in the number of overhead packets. The figure shows that the overhead packets is reduced by 20% when we increased the transmission range from 200m to 250m. The overhead is further reduced by 30% when the transmission range was increased further by another 50m. The reason for this kind of overhead reduction is due to the fact that when the transmission range is high, the request packet is travelling few long-hops and hence for a given network size request packets are broadcasted fewer times during the route discovery process. On the other hand, when the transmission

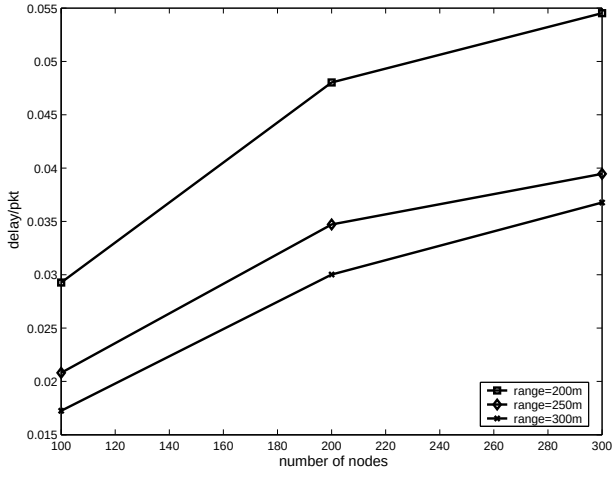


Fig. 4. End-to-end delay comparison

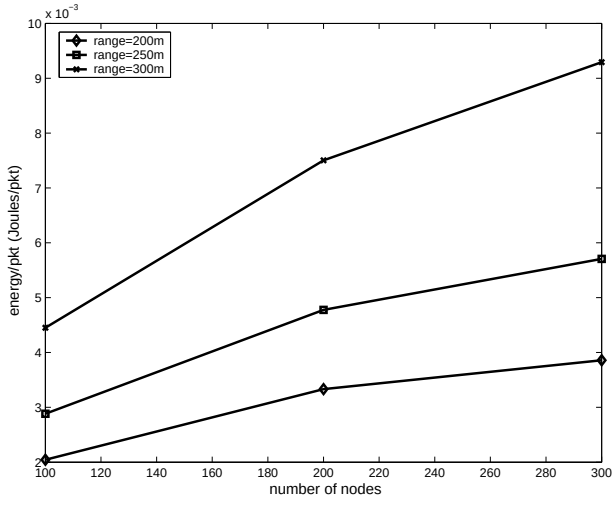


Fig. 5. Energy consumption for different transmission ranges

range is small, the request packets travel for many short-hops. Hence there are more broadcasts of a request packet.

B. End-to-end delay

The end-to-end delay per packet is the time taken by a data packet to travel from its point of origin to its destination. After receiving a packet, a mobile node needs to access the medium by a medium access control algorithm like IEEE 802.11 [20]. A node can only transmit a packet when it finds the surrounding medium is free. Hence a packet needs to wait at each hop while travelling from the source to the destination. If many short-hops is used, a packet needs to wait at many nodes and hence it takes longer time to travel. On the other hand, when long-hop is used, a packet needs to wait at few number of mobile nodes. The obvious consequence is that loner-hop will lower the delay per packet. On the other hand, short-hop will increase delay. The delay performances of the network for different transmission ranges are shown in Figure 4. It is depicted from this figure that end-to-end delay per packet is low when the transmission range is the maximum. The end-to-end delay increases when we decrease the transmission range. But there not very significant improvement is delay irrespective of transmission ranges when the network size is small. Because for a small network size, packet travels only a few hops (one or two) from source to destination. But delay improvement is more significant when the network size is large. For example, when the network size is of 200 nodes and when the transmission range is increased from 200m to 250m, the delay per packet is reduced by 20%. That delay is further reduced by 30% when transmission range is further increased from 250m to 300m. For a network size of 300 nodes, there is a reduction of delay by 40% when the transmission range is increased from 200m to 250m. The figure shows that delay per packet is further reduced by 30% when the transmission range is increased from 250m to 300m.

C. Energy consumption

The energy consumption and energy model in ad hoc network has been investigated in [18] and [19]. According to these models, the energy spent at wireless node's network card while transmitting a packet is described by the following equation

$$E(D, P_t) = K_1 P_t D + K_2 \quad (1)$$

where the values of K_1 and K_2 are 4 μ-sec/byte and 42μ Joules respectively. Equation 1 will be used in the rest of this paper as the energy consumption model. In order to investigate how a mobile node spends energy while transmitting packets, we focus on three types of packet namely data packet, Medium Access Control (MAC) packets and routing packets. The MAC layer packets are namely Clear-to-Send (CTS), Request-to-send (RTS) and Acknowledgement(ACK) packets. The routing packets are namely route request packet, route reply packet and route error packets. The MAC packets and the routing packets altogether are called overhead control packets. The number of overhead packets generated in the network depends on the network size. When the network size is small, the number of overhead packets is very small. But when the network size is large, the number of overhead packets is huge. Hence a small amount of energy is consumed by overhead packet when the network size is small. But a considerable portion of node energy is spent for transmitting overhead packets when network size is large. In our study we focus on the total energy consumption by different types of packet. We meausred energy consumption by energy per data packet delivered to the destination. The energy per data packet is the ratio of total energy consumed and total data packet delivered to the destination. The energy consumption per packet under different network size is shown in Figure 5. It is depicted from this figure that the energy consumption per packet increases with the network size. But the energy consumption per packet is low when the transmission range is the minimum (i.e. 200m). The energy consumption per packet increases with the transmission range.

For example, when the network size is 100 nodes, the energy consumptions per packet are 2.0 mJ, 3.0 mJ and 4.0 mJ for

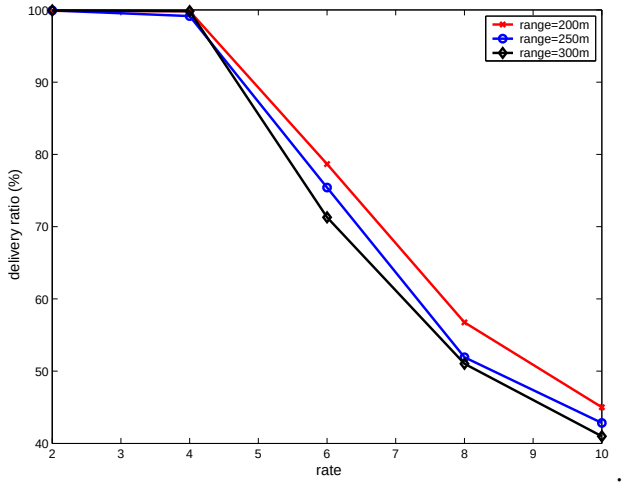


Fig. 6. Delivery ratio performances

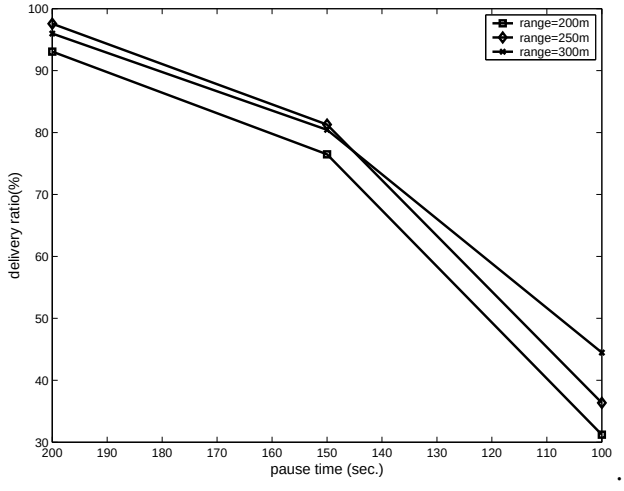


Fig. 7. Network performance under mobility cases

the transmission ranges of 200m, 250m and 300m respectively. The energy consumption per packet is increased by 30% and 33% when the transmission range is increased from 200m to 250m and 250m to 300m respectively. When network size is large (i.e., 300 nodes), the energy consumptions per packet are 3.5 mJ, 5.5 mJ and 8.5 mJ for transmission ranges of 200m, 250m and 300m respectively. Hence energy consumption per packet increases by almost 50% when the transmission range is increased from 200m to 250m and energy consumption increases by amlost 60% when the transmission range is increased from 250m to 300m. We can conlude from the energy consumption patterns depicted in Figure 5 that to save battery of the mobile node, transmission power should be as minimum as possible to save energy and hence to maximize network life.

D. Delivery ratio

Short-hop routing has gained a lot of support to reduce packet loss in the network. The proponents argue that the probability of packet collision will be reduced if short-hop routing is used. Hence there will be less packet loss in the network. The packet collision probability for IEEE 802.11 MAC layer has been formulated in [17], which is given by

$$\gamma = 1 - e^{-(n-1)\alpha} \tag{2}$$

where n is the number of neighbors in a given region, α is the parameter of the exponential back-off duration and this $\frac{1}{\alpha}$ has units of time. When the transmission range is low, the number of neighbor in a given region will decrease. Hence the packet loss will be reduced. However, the parameter α depends on how much traffic is there in a given region of the network. In order to test the packet loss probability, we created a network consists of 200 nodes placed over an area of 2000m $\times$ 1000m. We varied the packet generation rate in the network. We measured the packet loss as the delivery ratio of the network. The delivery ratio is the ratio between the data packet delivered to the destination and data packet sent to the destination. The delivery ratio performance is shown in Figure 6. It is depicted from that figure that there is very negligible packet loss in the network when there is not too much traffic in the network. The delivery ratio is almost 100% upto packet generation rate of 4.0 packet/sec. After that point, the delivery ratio decreases rapidly. The delivery ratio is always better when the transmission range is minimum (i.e., 200m). Thus the result supports the popular idea that the transmission range of the mobile node should be as minimum as possible to reduce packet loss.

E. Mobility

Mobility affects the performance of a network. When a mobile node moves out of the transmission range of one another, there will be a route breakage. More mobility of mobile nodes means more route breakages. Route breakage is coped with route maintenance mechanism as mentioned in section II. When a route breakage occurs in the network, some packets are lost because those packets do not have route to travel. Hence the packet loss in the network will increase with the mobility of the nodes. To test the effects of mobility on packet loss, we used a mobility model called random-waypoint model which is available with the Network Simulator [15]. According to this model, mobile node targets a random point in the network as the destination position and moves there at a pre-specified speed. After reaching that point, they stays there for some time called 'pause time'. To test the performances of a network under different mobility conditions, we created a network consisting of 200 nodes placed over an area of 2000m $\times$ 1000m. The mobile speed was set at 20 m/second. We varied the transmission ranges and tested network performances for different topologies. We set pause time to 200 second. Then we decreased the pause time. High pause time means less mobility in the network. Low pause time means high mobility in the network. The results are shown in Figure 7. It is depicted from that figure that the delivery ratio is higher when we increased the transmission range. Under higher transmission range, the mobile should take longer time to move out of the reach of its neighbor. On the other hand, mobile node will quickly moves

out of the reach of its neighbor if transmission range is small. The figure shows that the delivery ratio is almost 100% when the network has the least amount of mobility (i.e., pause time is 200 second). But for the high mobility (i.e., pause time is 100 second), the delivery ratio is 30% when the transmission range is 200m. The delivery ratio improves when the transmission range was increased. For exmaple, the delivery ratios were 35% and 45% when the transmission ranges were 250m and 300m respectively. Hence the simulation results show that long-hop is suitable if there is high mobility conditions in the network.

IV. Conclusions

In this paper, we investigated the performances of MANET under short-hop and long-hop routing conditions. The simulation results show that it is hard to find an optimum transmission range that will maximize the network performances in terms of delay, number of overhead, energy consumption and delivery ratio. But it is shown clearly that the decision whether short-hop routing or long-routing is a better choice depends upon the netwok conditions. How many nodes are there, how much traffic network is carrying, what is the mobility condition of the network etc. When the main objective of network is to reduce end-to-end delay per packet, long-hop routing shows better performance. But long-hop consumes more energy. Hnece it will reduce network life. So if the objective of a network is that it should be operative as long as possible, short-hop routing is a good choice. When there is more mobility in the network, long-hop routing helps to reduce packet loss. If a network has problem with high overhead packets, long-hop routing is preferable because it produces less overhead in the network. Hence we can conclude that there is no definite answer when one should-use shot-hop and when one should use long-hop routing. Network designers should choose either long-hop routing or short-hop routing decision depending upon the network conditions and performance objectives of network.

References

[1] J. Jubin and J. Tornow, *The DARPA Packet Radio Network Protocols*, In the Proceedings of the IEEE, 75(1), 1987, pp. 21-32

[2] D. Bertekas and R. Gallager, *Data Networks*, Prentice Hall, Upper Saddle River, NJ 07458, Second Edition, 1992, pp. 349-350

[3] H. Takagi and L. Kleinrock, *Throughput Delay Characteristics of Some Slotted-aloha Packet Radio Networks*, IEEE Transaction on Communication, Vol. 33, 1985, pp. 1200-1207

[4] L. Kleinrock and J. Silvester, *Optimum transmission Radii for packet Radio Networks or why six is a Magic Number*, In the Proceedings of the IEEE National Telecommunication Conference, Birmingham, Alabama, December 1978, pp. 4.3.1-4.3.5

[5] M. Shanchez, P. Manzoni and Z.J. Haas, *Determination of Critical Transmission Range in Ad hoc Networks*, In the Proceedings of the International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, May 2-4, 2007, pp. 1-12

[6] J. Broch, D. B. Johnson, and D. A. Maltz, *The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks*, IETF Internet-Draft, draft-ietf-manet-dsr-00.txt, March 1998

[7] T.A. Elbatt, S.V. Krishnamurthy, D. Connors, and S. Dao, *Power Management for Throughput Enhancement in Wireles Ad hoc Networks*, In the Proceedings of the IEEE International Conference on Communications (ICC), New Orleans, LA, June 2000, pp. 1503-1513

[8] E. M. Royer, P. Micheal, M. Smith and L. E. Moser, *An analysis of the Optimum Node Density for Ad hoc Mobile Networks*, In the proceedings of the Internatioanl Conference on Communication (ICC), Helsinki, Finalnd, June 11-14, 2001, Vol.3, pp. 857-861

[9] M.Haenggi and D. Puccinelli, *Routing in Ad Hoc Networks: A case of long hops*, IEEE Communication Magazine, October 2005, pp. 93-101

[10] M. Tarique and K. E. Tepe, *A New Hierarchical Design for Wireless Ad hoc Network with Cross Layer Design*, International Journal of Ad hoc and Ubiquitous Computing, Vol. 2, No. 1/2, 2007, pp. 12-21

[11] S.Y. Ni, Y.C. Tseng, Y. S. Chen, and J. P. Sheu, *The Broadcast Storm Problem in a Mobile Ad Hoc Networks*, MobiCom99, Seattle, Washington, August 15-20 1999, pp. 151-162

[12] B. Krishnamachari , S.B. Wicker, and R. Bejar, *Phase transition phenomenon in wireless ad-hoc networks*, In the proceedings of GLOBOCOM, San Antonio,Texas, November,2001, pp. 2921-2915

[13] Y. Sasson, D. Cavin and A. Schiper, *Probabilistic Broadcast for flooding in Wireless Mobile Ad hoc Networks*, Swiss Federal Institute of Technology, Switzerland, Technical Report IC/2002/54

[14] J. Z. Haas, Y. J. Halpern, and L. Li, *Gossip Based ad hoc routing*,In the proceedings of IEEE INFOCOM 2002, New York, Volume 3, June 2002, pp. 1707-1716

[15] K. Fall and K. Varadhan, *NS Notes and Documentation Technical Report*,University of California Berkeley, LBL, USC/ISI, and Xeron PARC

[16] H. Lim and C. Kim, *Multicast tree construction and flooding in wireless ad hoc networks*, In the proceedings of ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems,Boston,2000

[17] k., Anurag, d. Manjunath and j. Kuri, *Communication Networking: An Analytical Approach*, Morgun Kauffman Publishers, San Francisco, CA, USA, pp. 715-716

[18] S. Doshi, S. Bhandare and T.X. Brown, *An On-demand Minimum Energy Routing Protocol for Wireless Ad hoc Network*, The ACM SIGMOBILE Mobile Computing and Communication Review, Vol. 6, No. 2. pp. 50-66

[19] L.M. Feeny and M. Nilsson, *Investigating energy consumption of wireless network interface in ad hoc networking environment*, Vol. 3, April 2001, pp. 1548-1557

[20] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, *MACAW: A media access protocol for wireless LANs*, In the proceedings of ACM SIGCOMM, London, UK, August 1994, pp. 212-225

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Upper Bound on Blocking Probability for Vertically Stacked Optical Banyan Networks with Link Failures and Given Crosstalk Constraint

Basra Sultana[1] and M. R. Khandker[1]

[1]Department of Applied Physics and Electronic Engineering,
University of Rajshahi, Rajshahi – 6205, Bangladesh.
E-mail: basra_apee@walla.com, khandker@ru.ac.bd

Abstract - Vertical stacking of multiple copies of an optical banyan network is a novel scheme for building nonblocking optical switching networks. The resulting network, namely vertically stacked optical banyan (VSOB) network, preserves all the properties of the banyan network, but increases the hardware cost significantly under first order crosstalk-free constraint. Therefore, blocking behavior analysis could be an effective approach to studying network performance, and finding a graceful compromise between hardware costs and blocking probability with different degree of crosstalk constraint and link failure probability. However, upper bound on blocking probability for such networks with link failures only has been presented in the literature. In this paper, we present the simulation results for upper bound on blocking probability considering both link-failures and given degree of crosstalk constraint. We find how crosstalk adds a new dimension to the performance analysis of practical VSOB networks where link failures present. The simulation results presented in this paper can guide network designer in finding a tradeoff among the blocking probability, the degree of crosstalk and link failures of VSOB networks.

I. Introduction

Optical mesh networks are considered more capacity-efficient and survivable for serving as the backbones for next generation internet. A key network element equipped with a switching node of optical mesh networks is the optical switch, which has the capability of switching huge data at an ultra-high speed. The main factors those have to be considered while designing any optical switching networks are hardware cost, blocking probability, crosstalk, switching speed etc. The basic 2×2 switching element (SE) in a large optical switching network is usually a directional coupler (DC) [1,8]. DC's can switch multiple wavelengths at the same time, which is important for the future optical cross-connects (OXCs).

Crosstalk is an intrinsic shortcoming of the DC. It is the effect of the undesirable coupling between the signals carried in the two waveguides of the coupler [1,2]. When two optical signals meet at a DC, a small portion of the signal power will be directed to the unintended output channel. Crosstalk suppression becomes particularly important in networks, where a signal propagates through many nodes and accumulates crosstalk from different elements at each node from the system view. In order to obtain an approximate idea of the crosstalk requirements, suppose that a signal accumulates crosstalk from N sources, each with crosstalk level ε. This neglects the fact that some interfering channels may have higher powers than the desired channel. Networks are very likely to contain amplifiers and to be limited by signal-spontaneous beat noise. For example, if we have 10 interfering equal-power crosstalk elements, each producing intrachannel crosstalk, then we must have a crosstalk suppression of below 35dB in each element, in order to have an overall penalty of less than 1dB [8]. Thus, Crosstalk reduction is an important issue in designing the systems that are based on DC's. The crosstalk issue can be tackled at either the device level or the system level. The two methods complement each other. The focus of this paper is on the system-level approach. As will be seen, crosstalk adds a new dimension to the theory of building a nonblocking and negligible blocking switching network.

Banyan networks [9,4,3,5] are a class of attractive switching structures for constructing DC-based optical switches, because they have a smaller and exactly the same number of SEs along any path between an input-output pair such that an absolutely loss uniformity and smaller attenuation of optical signals are guaranteed in this class of switching networks. However, with the banyan topology only a unique path can be found from each network input to each network output, in which the network is degraded as a blocking one. The general scheme for building banyan-based nonblocking optical switching networks is to vertically stack the multiple copies of regular optical banyan network [7,11] as illustrated in Fig.1. We use VSOB to denote the optical switching networks built on vertical stacking scheme of banyan network.

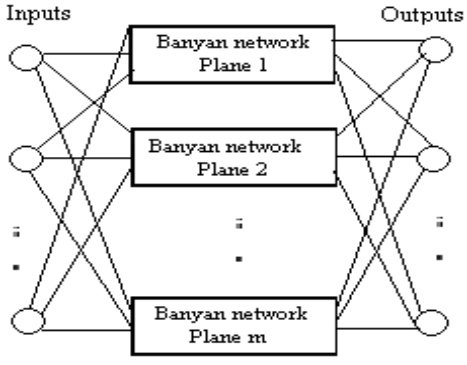


Fig. 1. Vertically stacked optical banyan (VSOB) network.

Numerous result are available for VSOB networks, such as [7], and their main focus has been on determining the maximum and minimum number of stacked copies (planes) required for a nonblocking VSOB networks without link-failures if conservative routing algorithm is used for routing a request to a plane. These results indicate that VSOB structure, although is attractive, usually requires either a high hardware cost or a relatively larger depth for building a nonblocking network. In paper [12], the maximum number of planes required for nonblocking a VSOB networks with link-failures and the blocking probability of VSOB networks for given failure probability & load have been determined. In paper [13], the lower bound on number of planes for VSOB networks with link-failures have been determined.

However, no results are available for evaluating the probabilities of VSOB networks with link failures under various degree of crosstalk constraint. As the first important step toward the blocking behavior analysis of general VSOB networks under any crosstalk constraint, we presented in this paper the blocking probabilities of VSOB networks having link-failures under crosstalk constraint with c=1,2,3..... where c is denoted to the degree of crosstalk and their upper bound with respect to the number of planes required to make the network nonblocking. The simulation results can guide network designers to initiate a compromise among the hardware cost, and the blocking probability of a VSOB networks under different degree of crosstalk.

The rest of the paper is organized as follows: Section 2 provides preliminaries that facilitate our further discussions. Section 3 presents our contribution. Section 4 concludes the paper.

II. Preliminaries

In VSOB network, blocking happens when two connections intend to use the same link, which is referred to as link-blocking. However, there is another type of connection-related blocking, which occurs when some paths (including the new one) violate the crosstalk-free constraint after adding the new connection. In such a situation, the connection is not allowed to be allocated even if the path is available. We refer to this second type of blocking as crosstalk-blocking throughout the paper. It is the combination of the two types of blocking that makes the design principle different.

A connection request may be blocked by link failures in a faulty VSOB network, which is referred to as the failure-blocking. We assume that the links in VSOB networks may fail independently and these failures are permanent. Thus, both crosstalk-blocking and failure-blocking should be fully considered in the blocking analysis of a faulty VSOB network as illustrated in Fig.2 for a 8×8 network.

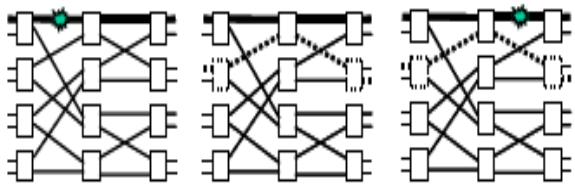


Fig. 2. Blocking in a VSOB network. (a) Failure-blocking. (b) Crosstalk-blocking. (c) Combination of failure blocking and crosstalk-blocking.

For A typical $N\times N$ banyan network consists of log_2N stages, each containing $N/2$ 2×2 SEs. Regular banyan network has a unique path between an input-output pair, and a basic technique for creating multiple paths between an input-output pair is the vertical stacking of multiple copies of the banyan network. In this paper, we focus on the banyan network that has m multiple copies as shown in fig.3.

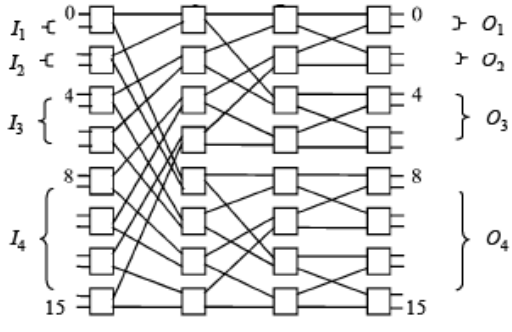


Fig. 3. 16×16 banyan network.

Due to the topological symmetry, all paths in banyan networks have the same property in terms of blocking. We define the blocking probability to be the probability that a feasible connection request is blocked, where a feasible connection request is a connection request between an idle input port and an idle output port of the network. Without loss of generality, we focus on the path between the first input and the first output (which is termed as the tagged path hereafter). All the SEs and links on the tagged path are called tagged SEs and tagged links respectively. The stages of SEs are numbered from left (stage 1) to right (stage log N) and the stages of links are also numbered from left (stage 1) to right (stage log N +1). For the tagged path, an input intersecting set $I_i=\{2^{i-1}, 2^{i-1}+1,\ldots, 2^{i}-1\}$ at stage i is defined as the set of all inputs that intersect a tagged SE at stage i. Likewise, an output intersecting set $O_i=\{2^{i-1}, 2^{i-1}+1,\ldots, 2^{i}-1\}$ associated with

stage i contains all the outputs that intersect a tagged SE at stage $\log N - i + 1$.

When two light signals go through an SE simultaneously, crosstalk is generated at the SE. Such SE is referred to as a crosstalk SE (CSE). The degree of crosstalk of the switching system is defined as the number of CSE's allowed along a path.

A restricted SE (RSE) is a 2×2 SE which carries only one light signal at a time. Although crosstalk at an RSE is very small, it may not be entirely zero. For example, when a light signal passes through an RSE, a small portion of the signal will leave at the other unintended output channel. This stray signal can arrive at the input of the next stage SE and generate some crosstalk. Since crosstalk generated by the stray signal is much smaller than the regular crosstalk, we will ignore it in our analysis.

Following the typical assumption as in [7,10] on probabilistic analysis of multistage interconnection networks, we neglect the correlation among signals arriving at input and outputs ports, and consider that the statuses (busy or idle) of individual input and output ports in the network are independent. This assumption is justified by the fact that the correlation among signals at inputs and outputs, though exists for fixed communication patterns, and becomes negligible for arbitrary communication patterns in large size networks, which is the trend of future optical switching networks that can switch huge data at high speeds.

A. Upper Bound on Number of Planes with link failures

Fig.4 shows the number of planes required to make VSOB network nonblocking with link-failures. This figure does not consider the effect of crosstalk on the blocking probability.

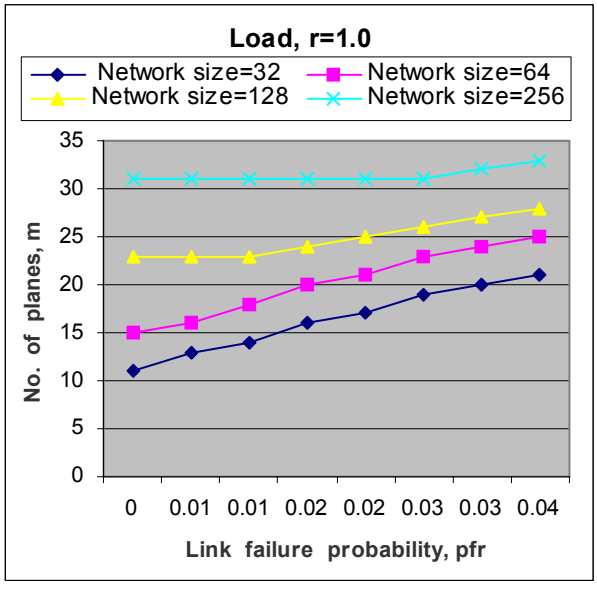


Fig. 4. Nonblocking condition for VSOB networks with link-failures.

III. VSOB Network with Link Failures and allow Crosstalk

In this section, we analyzed the blocking behavior of VSOB networks having link-failures and allow crosstalk. We determined the maximum number of planes required to make VSOB networks nonblocking.

A. Network Simulator

The network simulator we developed consists of six major modules as shown in fig. 5.

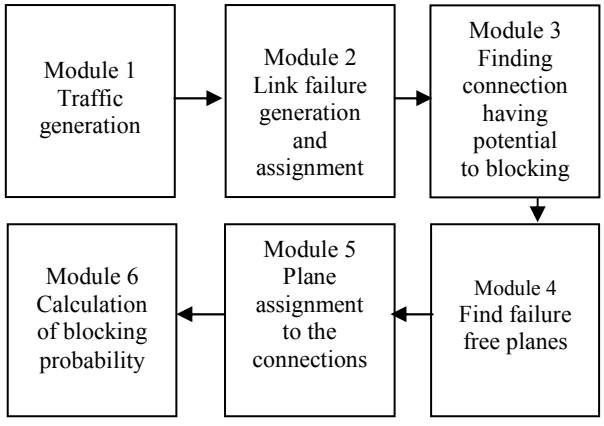


Fig. 5. Block diagram of network simulator.

We consider here the permutation request as the traffic. Due to the symmetric architecture of VSOB *(N, T)* network, every connection request has the same probability to be blocked. In our simulation, we fix the connection request of input–output pair 0–0 and investigate the blocking probability of this connection request only that may result by other contentious connections.

Module 1: This module randomly generates a permutation request for the VSOB *(N, T)* network based on the workload r (here workload r is defined as the occupancy probability of a port).

Module 2: This module generates link failures based on the given *pfr* (here *pfr* is defined as the probability that a link is failed or broken) and then assign those failures randomly to different links.

In this section we present some definitions, which are used in the discussion of our algorithm:

Definition 1 **Blocking connections, C_{bc}**: The connections those are potential to blocking the tagged path.

Definition 2 **Crosstalk-blocking connections, C_{cbc}**: The connections those are potential to crosstalk-blocking the tagged path.

Definition 3 **Failure free planes, P_i**: The planes those are free from link-failures on the path of each C_{bc}.

Definition 4 **Maximum allowed crosstalk, C_m:** C_m is defined the maximum allowed crosstalk along the tagged path.

Definition 4 **Allowed crosstalk, c:** c is defined the allowed crosstalk along the tagged path at any time.

Module 3: This module finds C_{bc} which is determined by the following relation:

$$I_i + O_j < \log_2 N + 2 \qquad (1)$$

Here, input i has originated from input intersecting group I_i and destined to output j that belongs to output intersecting group O_j (see Fig.2).

We find C_{cbc} that must satisfy the following relation

$$I_i + O_j = \log_2 N + 1 \qquad (2)$$

Module 4: First we check all the planes if there is a failure on the tagged path and make a list of planes, say P_{tagged}, in which no links on the tagged path are failed or broken. Then finds P_i from P_{tagged}, and sort it in the ascending order where i –th entry represents the number of free planes for input i.

Module 5: This module assigns connection requests to different planes in the following way. The tagged path is assigned first to a plane randomly chosen from the list of free planes of tagged path. Then a connection from the list of C_{cbc} is assigned to the plane in which tagged path assign and allow one c. If c less than C_m then the next connection from the list of C_{cbc} is picked and assigned to the plane as mentioned above. When C_m reached then the other connections from the list of C_{bc} are assigned to separate planes. This plane assignment algorithm ensures the use of maximum number of planes for routing a permutation request. At last if there is a connection for which no such plane is available then the connection request pattern is recorded as a blocked connection pattern.

Module 6: In this module the blocking probability is estimated by the ratio of number of connection requests in which the 0–0 request is blocked to the total number of connection requests generated.

A.1. Upper Bound on Number of Planes Given Crosstalk Constraint

The simulation results for upper bound on blocking probability of VSOB networks with link failures under crosstalk constraint is given below:

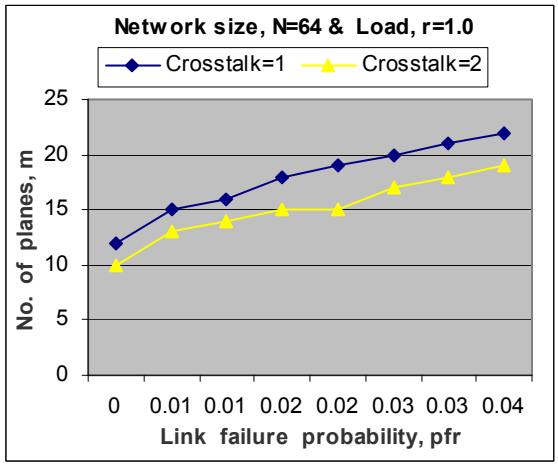


Fig 6. Maximum number of planes to make VSOB network nonblocking with given crosstalk .

Fig.6 shows that if we allow small amount of crosstalk then the number of planes required to making the VSOB networks nonblocking decreases.

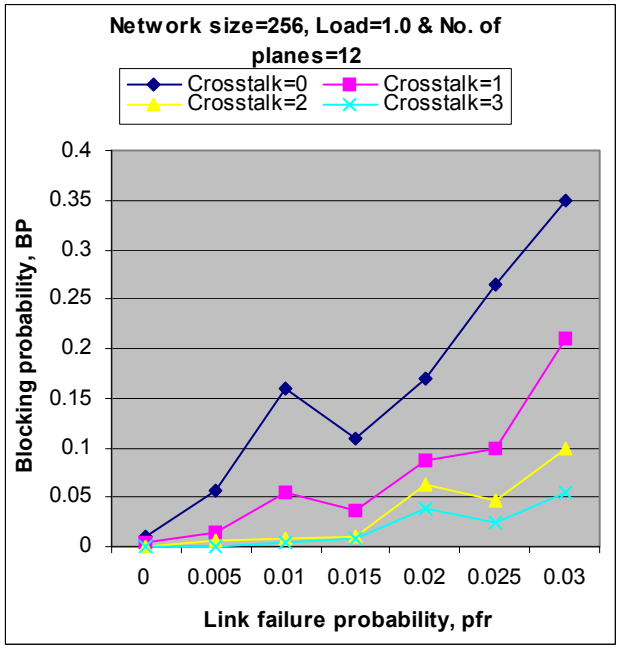


Fig. 7. Blocking probability decreases for certain range of link-failures and then increases again.

It is also interesting to note from Fig.7 that the blocking probability not always increases with the increase of link failures; blocking probability also decreases for certain range of link failures and then increases again. For N=256, r=1, m=12 & c=0; if we introduce 0.015 link failure to the networks then BP increases 91.3%. Now if we allow 2 CSEs to the network then BP decreases 90.9%. That means the amount of performance degradation caused by link failures can be compensated by introducing CSEs. This may help switch designer choose a switch networks for different degree of performance with comparably low hardware cost.

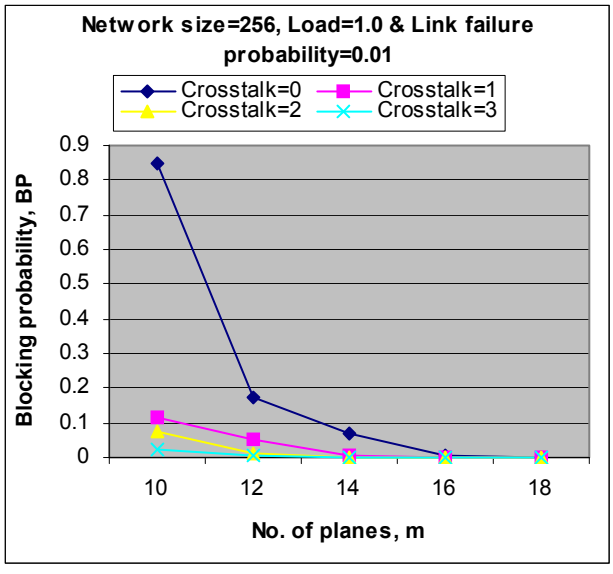


Fig. 8. Blocking probability decreases exponentially with number of planes.

It is shown from Fig.8 that for N=256, pfr=0.01, r=1.0, c=0, m=10; if we allow 2 CSEs then BP decreases by 90.9% than c=0 and if we increase the number of planes from 10 to 12 then BP decreases by 79.88%. This picture derives the fact that if we allow small amount of crosstalk then the blocking probability decreases rate is high compare the number of planes.

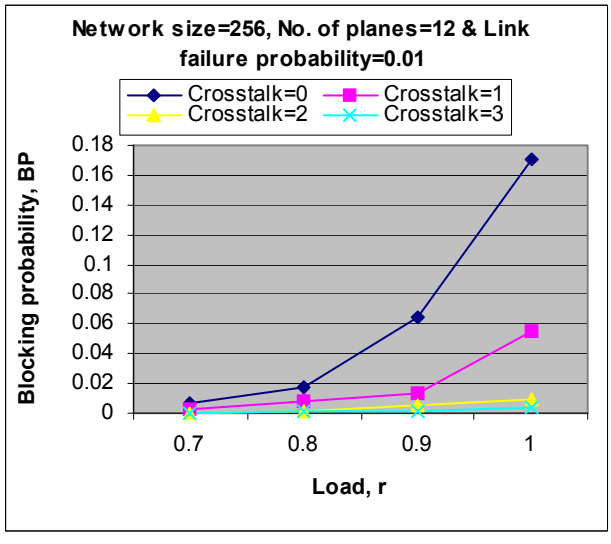


Fig. 9. load vs blocking probability for different Crosstalk.

The effect on blocking probability with load for different crosstalk is shown in Fig.9. If we reduce load 1 to 0.8 then BP decreases 85.8% for crosstalk=1, BP decreases 89.9% for crosstalk =2.

A.2. Worst Case Scenario

In the above simulation, the traffic generation module randomly generates a permutation request. In that type permutation, the probability of worst-case permutation generation is very small. From paper [14], the probability of worst-case scenario is given by

$$P_{worst} = 2.57\times\text{exp-}10 \quad \text{for } N=64,\ r=0.9.$$
$$P_{worst} = 2.45\times\text{exp-}20 \quad \text{for } N=128,\ r=0.9.$$

The above results indicate that the probability of worst-case scenario is very small for both even and odd number of stages. So we only generate a subset of all possible worst-case permutation and the simulation results for that case is given below:

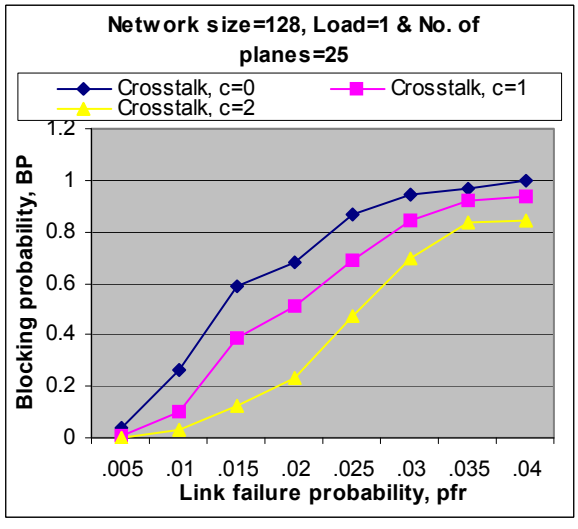


Fig. 10. Link failure probability vs blocking probability for different crosstalk.

Fig.10 shows that blocking probability gradually increases with link failure probability for different crosstalk. For N=128, pfr=0.01, r=1.0, m=25; if we allow 2 CSEs to the network then BP decreases by 88% than c=0.

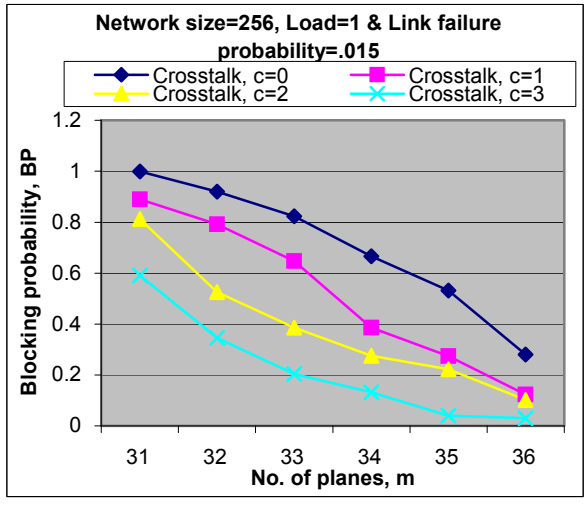


Fig. 11. Number of planes vs blocking probability for different crosstalk.

Fig.11 shows that blocking probability decreases gradually with planes. For *m*=31-33, BP decreases 17% for c=0, BP decreases 52% for c=2.

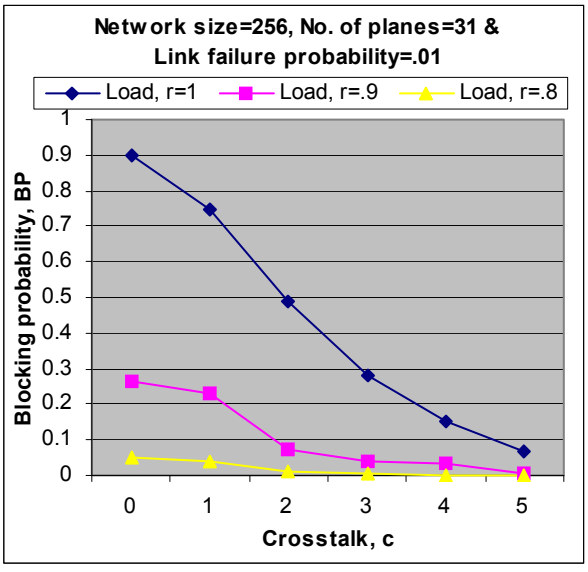


Fig. 12. Crosstalk vs blocking probability for different load under worst-case permutation.

The effect on blocking probability with various crosstalks for a fixed network size, link failure probability and number of planes is shown in Fig.12. For N=256, m=31, pfr=0.01; if we allow 2 crosstalk then BP decreases 45% for r=1, BP decreases 72% for r=0.8.

IV. Conclusion

We have presented results of upper bound on number of planes required to make the VSOB networks nonblocking having link-failures and given certain degree of crosstalk constraint. The simulated results in section 3 can provide network developers with a guidance of quantitatively determining the effects of various degrees of crosstalk and reduction in number of planes on the overall blocking behaviors of VSOB networks having link-failures. The results can also show how the crosstalk adds a new dimension to the VSOB networks and effects of crosstalk on hardware cost and blocking probability.

References

1. *H. S. Hinton, An introduction to photonic switching fabrics,* New York: Plenum, 1993.
2. R. Chinni etal., Crosstalk in a lossy directional coupler switch, IEEE/OSA Journal of Lightwave Technology, vol. 13, no. 7, pp. 1530-1535, July 1995.
3. G. R. Goke and G. J. Lipovski, Banyan networks for partitioning multiprocessor systems, Proc.1st Annu. Symp. Comp. Arch, pp.21-28, 1973.
4. F. T. Leighton, Introduction to parallel algorithms and architectures: Arrays, Trees, Hypercubes, Morgan Kaufmann, 1992.
5. J. H. Patel, Performance of processor-memory interconnections for multiprocessors, IEEE Trans. Comput., vol. C-30, pp. 771-780, Oct. 1981.
6. M. M. Vaez and C. -T. Lea, Strictly nonblocking directional-coupler-based switching networks under crosstalk constraint, IEEE Trans. Commun., vol. 48, no.2, pp. 316-323, Feb. 2000.
7. X. Jiang, H. Shen, Md. Mamun-ur-R. Khandker, S. Horiguchi, Blocking Behavior of Crosstalk free Optical Banyan Networks on Vertical Stacking, IEEE/ACM Transactions on Networking, Vol. 11, No. 6, Dece. 2003.
8. Rajib Ramaswami and Kumer N. Sivarajan, Optical networks, Morgan Kaufmann Publishers, ISBN:1-55860-655-6.
9. C.Yu, X. Jiang and S. Horiguchi, Performance Modeling for Vertically Stacked Optical Banyan Networks with Extra Stage, International Journal of Computational Science and Engineering (IJCSE), Vol.2, No.1-2, pp.81-87, May 2006.
10. Xiaohong Jiang, H. Shan and S. Horiguchi, Bloaking Probability of Vertically Stacked Optical Banyan Networks Under Random Routing, Proc. Of GLOBECOM 2003, Dec.1-5, San Francisco, USA.
11. Xiaohong Jiang, H. Shan, S. Horiguchi and P.H.Ho, Performance Modeling for All-Optical Photonic Switches Based on the Vertical Stacking of Banyan Network Structures, IEEE Journal of Selected Areas on Communications, vol.23/8, pp.1620-1631, Aug.2005.
12. Basra Sultana, M.R.Khandker, X.Jiang and S.Horiguchi: "Blocking Probability of Vertically Stacked Optical Banyan Networks with Link Failures", Proc. The International Workshop on High Performance and Highly Survivable Routers and Networks, Tohoku University, Sendai, Japan, pp. 157-169, March 14, 2007.
13. Basra Sultana, M.R.Khandker: "Lower Bound on Number of Planes for Vertically Stacked Optical Banyan Networks with Link Failures", Proc. 10th International Conference on Computer and Information Technology, pp. 178-183, 27-29 December, 2007, Dhaka, Bangladesh.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Blocking Behavior Analysis of Extended Pruned Vertically Stacked Optical Banyan Networks with Link Failures

Basra Sultana[1] and M. R. Khandker[1]

[1]Department of Applied Physics and Electronic Engineering,
University of Rajshahi, Rajshahi – 6205, Bangladesh.
E-mail: basra_apee@walla.com, khandker@ru.ac.bd

Abstract - To guarantee a high switching speed, routing in vertically stacked optical banyan (VSOB) networks needs special attention so that connections are established as fast as possible. Previously proposed Pruned Vertically Stacked Optical banyan (P-VSOB) networks used Plane Fixed Routing (PFR) algorithm, and has O(log_2N) time complexity. Blocking probability has also been analyzed for these kinds of networks having link failures. This paper deals with the blocking behavior of EP-VSOB (Extended Pruned VSOB) networks having link failures. In EP-VSOB networks a few regular banyan planes are added with the P-VSOB networks. Necessary routing algorithms, namely, PFR_LS and PFR_RS show that this switching network can reduce the blocking probability to very low value even with zero-crosstalk constraint while keeping the hardware cost almost the same as that of P-VSOB networks. Both these algorithm also have the time complexity O(log_2N). The blocking behavior of EP-VSOB networks is much better than P-VSOB networks. Simulation results show that the blocking probability of the VSOB networks does not always increase with the increase of link-failures; blocking probability decreases for certain range of link-failures, and then increases again.

I. Introduction

Optical switching network plays an important role in optical networks, which has the capability of switching huge data at an ultra-high speed. The main factors those have to be considered while designing any optical switching networks are hardware cost, blocking probability, crosstalk, switching speed etc. Researchers have proposed many switching architectures having different merits in the literature. A few important works have been listed in the references.

Directional-coupler (DC) [1,2] can handle the optical signals of some terabits per second and with multiple wavelengths, and this make it ideal for serving as the basic 2×2 switching element (SE) in high-speed optical switches. However, one of the major shortcomings of a DC is crosstalk. By ensuring that only one signal passes through a DC at a time, crosstalk can be eliminated.

Banyan networks [3,4,5,6] are attractive for constructing DC- based optical switches for their small depth, absolute loss uniformity, lower hardware cost and a simple switch setting ability (self-routing). With banyan structure, only a unique path exists between input-output pair in which the network is simply degraded as a blocking one. The blocking probability of banyan network is high. The whole network can be made nonblocking by vertically stacking multiple copies of an optical banyan network [7], known as vertically stacked optical banyan (VSOB) network (Fig.1). The VSOB network increases the hardware cost but decreases the blocking probability, and has O($Nlog_2N$) time complexity. Due to high hardware cost and routing complexity of VSOB network, researchers proposed Pruned VSOB (P-VSOB) network [8]. In P-VSOB network, PFR routing algorithm is used in which a plane of the VSOB network is fixed for a group of inputs and achieved $O(\log_2 N)$ time complexity. Although P-VSOB networks guarantee considerably low blocking probability, but for some high performance applications, this blocking probability may be considered too high. To deal with this situation, Khandker *et. al.* have proposed Extended-Pruned-VSOB (EP-VSOB) network [9] that has the potential to reducing the blocking probability while keeping the crosstalk constraint to zero. Necessary routing algorithm, namely PFR_LS and PFR_RS modifying the PFR algorithms have also been proposed. The EP-VSOB network has almost the same routing complexity as that of the P-VSOB network with significantly lower blocking probability. This switch network achieves the speed close to P-VSOB networks and blocking probability close to VSOB networks.

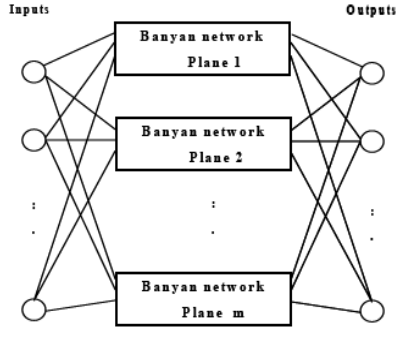


Fig. 1. Vertically stacked optical banyan network.

Due to the increasing importance and requirement for fault-tolerance in optical switches for large mesh WDM networks, performance analysis of VSOB networks at the presence of probability of link failures becomes critical for the practical adoption of the VSOB networks in the current internet applications. Paper[10] deals the blocking behaviour analysis of VSOB networks with link-failures and the author showed that VSOB network is a fault tolerant network. In paper [11], the blocking probabilities of P-VSOB networks having link failures have been determined. P-VSOB networks can accommodate some link failures safely without degrading the performance. In this paper, we analyze the blocking behaviour of EP-VSOB networks when some links are failed or broken in the network.

The rest of the paper is organized as follows. We discuss the published work briefly in section 2; especially EP-VSOB networks. Section 3 presents our contribution. We conclude this paper in section 4.

II. Preliminaries

A connection request may be blocked by link failures in a faulty VSOB network, which is referred to as the *failure-blocking*. We assume that the links in VSOB networks may fail independently and these failures are permanent. Thus, both crosstalk-blocking and failure-blocking should be fully considered in the blocking analysis of a faulty VSOB networks as illustrated in Fig.2 for a 8×8 network.

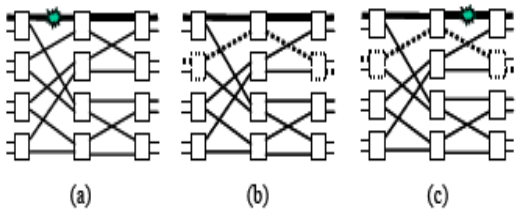


Fig. 2. Blocking in a VSOB network. (a) Failure-blocking. (b) Crosstalk- blocking.(c) Combination of failure-blocking and crosstalk-blocking.

Due to their topological symmetry, all paths in a banyan network have the same property in terms of blocking. Through out this paper, we will select the path between the first input and the first output and try to set up a connection between them. We call the path between this input-output pair the tagged path. In this paper, we consider the banyan networks that support one-to-one communication. We define the blocking probability to be the probability that a feasible connection request is blocked, where a feasible connection request is a connection request between an idle input port and an idle output port of the network

A. Pruned-VSOB networks

P-VSOB(N,T) denotes an N×N Pruned-VSOB network that has T fixed planes. In P-VSOB network, plane fixed routing (PFR) algorithm is used for routing input signals to requested outputs. In PFR algorithm, we select one input from each input group Ii and tie them with a plane of VSOB(N, T) network permanently. Input groups Ii is defined as the set of inputs {i,i+1,i+2, …,i+T-1} as shown in Fig.3. We first define the following subsets of input taking on from each input sets:

$$G_i = \left\{ i, i+T, i+2T, \cdots, i+\left(\frac{N}{T}-1\right)T \right\}; o \le i \le T-1 \quad (1)$$

$$T = 2^{[(\log_2 N + 1)/2]} \quad (2)$$

In PFR algorithm all N inputs are always evenly distributed among the planes and each plane has only N/T connections. Since these N/T connections are fixed at their corresponding inputs in the plane, only 1 input of a group will be active; all other inputs will remain unused. Therefore, input switches connected to unused inputs and switches in the successive stages corresponding to these unused inputs are redundant. All these redundant switching elements are eliminated. Fig.4 explains the idea.

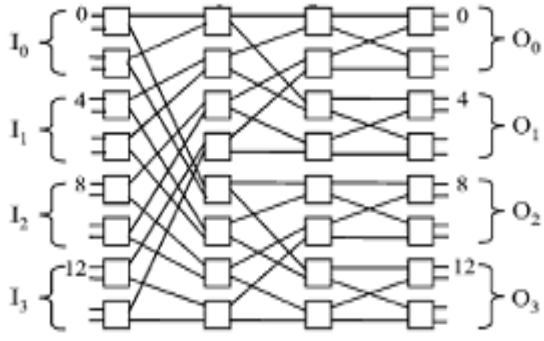


Fig. 3. Illustration of inputs and outputs grouping.

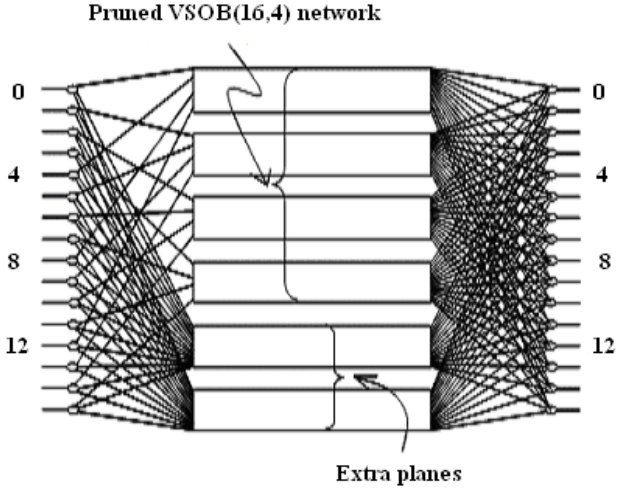


Fig. 4. The structure of a EP-VSOB (16,4+2) network.

B. Extended Pruned VSOB(*N,T+K*) network

In this section, we describe previously proposed Extended-Pruned-VSOB network. EP-VSOB(N,T+K) denotes an *N*×*N* EP-VSOB network that has *T* fixed planes (pruned banyan) and *K* extra planes (regular banyan). The main

idea of the EP-VSOB is to add a small number of regular banyan planes to the P-VSOB(N,T) network such that the connections blocked in the fixed T planes of the P-VSOB(N,T) network have more chances to be established through these extra planes.

Depending on how a plane is chosen from the extra planes, we propose two routing algorithm for fast connection establishment in an EP-VSOB($N,T+K$) network, namely PFR with linear search algorithm and PFR with random routing algorithm.

B.1. PFR with Linear Search (PFR_LS)

In the PFR_LS algorithm for an EP-VSOB($N,T+K$) network, each connection now has (1+K) chances to be established through the network. Whenever a request arrives at an input, first it is sent to the fixed plane of that input. If the request cannot be established in the fixed plane, the input searches for a free plane among the extra K planes. The searching starts with the first plane and continues orderly up to the last one so that each plane is being searched at most one time by an input. If the connection request still fails to find a free plane among all K extra planes, it is considered as a blocked request. The time complexity of the PFR_LS algorithm O(log_2N) when K is constant and much smaller compared to T.

B.2. PFR with Random Selection (PFR_RS)

In the PFR_RS algorithm for an EP-VSOB($N,T+K$) network, each connection has two chances to be established through the network. The first chance is to establish the connection through its fixed plane. If the connection request is blocked in its fixed plane, it still has a second chance to be established through another plane selected randomly from K extra planes. If the connection request still could not be established through the randomly selected plane, it is considered as a blocked request. It is easy to see that the time complexity of the PFR_RS algorithm remains the optimum O(log_2N).

The simulation results of EP-VSOB networks using without link failures are given below

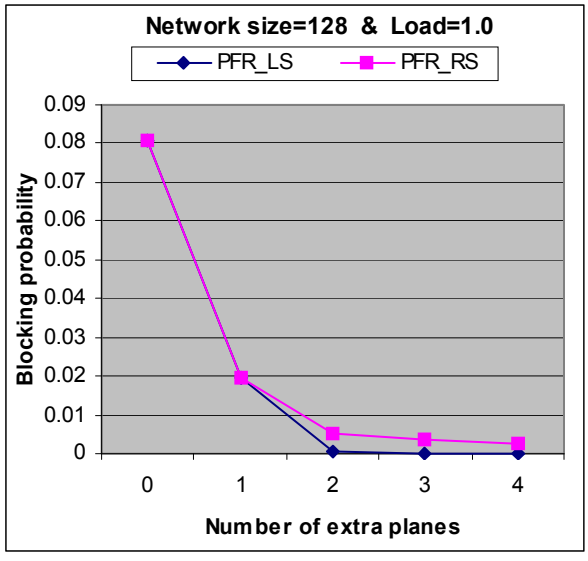


Fig. 5. Blocking probability EP-VSOB networks with PFR_RS algorithm.

III. EP-VSOB Networks with Link Failures

In this section, we analyzed the blocking behaviour of EP-VSOB networks having link-failures using PFR_LS and PFR_RS algorithm.

A. Network simulator

The network simulator we developed consists of six major modules as shown in figure 6.

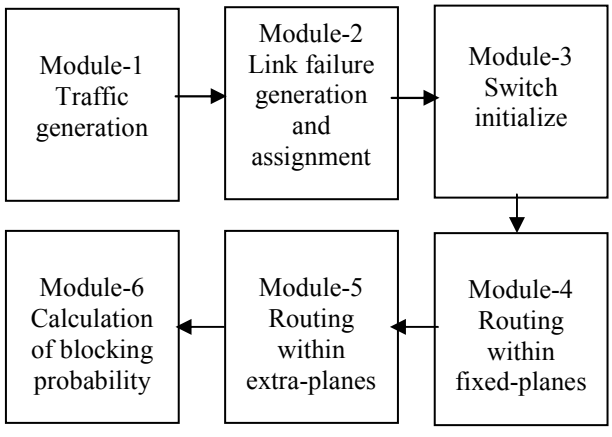


Fig. 6. Block diagram of the network simulator.

We consider here the permutation request as the traffic since a permutation does not have output contention, and therefore, gives real blocking probability of the switch network only. Due to the symmetric architecture of P-VSOB *(N, T) network*, every connection request has the same probability to be blocked. In our simulation, we fix the connection request of input–output pair 0–0 and investigate the blocking probability of this connection request only.

Module 1: This module randomly generates a permutation request for the VSOB*(N, T)* network based on the workload *r* (here workload *r* is defined as the occupancy probability of a port).

Module 2: This module generates link failures based on the given *pfr* (here *pfr* is defined as the probability that a link is failed or broken) and then assign those failures randomly to different links.

Module 3: This module initializes the switches with the permutation request.

Module 4: This module attempts to assign connection requests to different planes using PFR algorithm. We only consider the plane that contains the tagged path. We try to establish the tagged path in the selected plane. If the tagged path is not established, then we try to assign other connections in their selected planes.

Module 5: If the tagged path is not established in the selected plane then we try to establish the tagged path (along with other connections that are not established in their selected plane) in the extra planes. In case of PFR_LS algorithm, the searching starts with the first plane and continues orderly up to the last one. If the tagged path still fails to find a free plane among all extra planes, it is considered as a blocked request. In case of PFR_RS algorithm, we randomly selected a plane from extra planes. If the tagged path still could not be established through the randomly selected plane, it is considered as a blocked request.

Module 6: In this module the blocking probability is estimated by the ratio of number of connection requests in which the 0–0 request is blocked to the total number of connection requests generated.

A.1. PFR_LS algorithm

The simulation results of EP-VSOB networks using PFR_LS algorithm with link failures are given below:

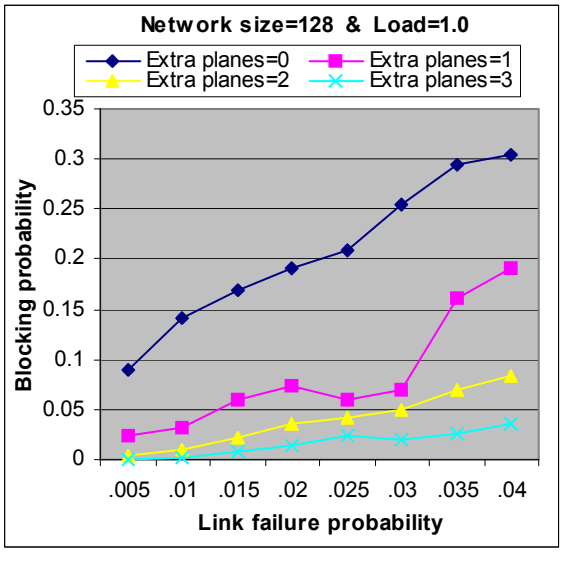


Fig. 7. Blocking probability decreases for certain range of link-failures and then increases again.

It is interesting to note from Fig.7 that the blocking probability not always increases with the increase of link failure; blocking probability decreases for certain range of link failures and then increases again. The reason for decreasing the blocking probability may be as follows. When there are some links failed on the path of potential blocking connections, they can not interfere with the tagged path. This phenomenon increases the tagged path's chance of being successful.

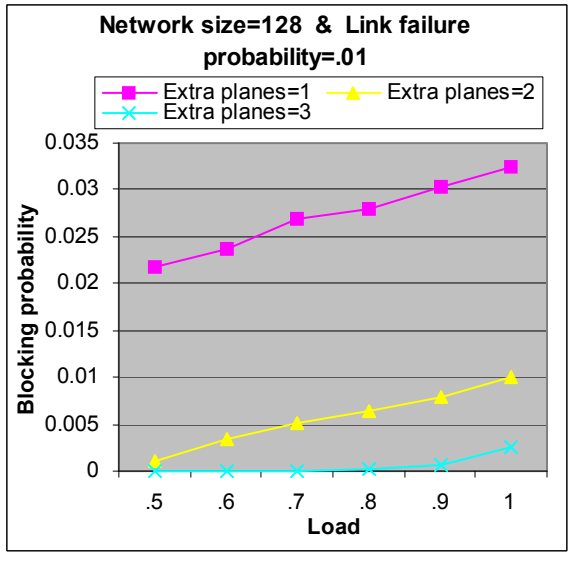


Fig. 8. Load vs blocking probability for different planes.

The effect on blocking probability with various load for a fixed network size and link failure probability is shown in Fig.8. The effect on blocking probability depends on the number of extra planes; if we reduce load 1 to 0.8 then BP decreases 14% for extra plane =1, BP decreases 35% for extra planes =2.

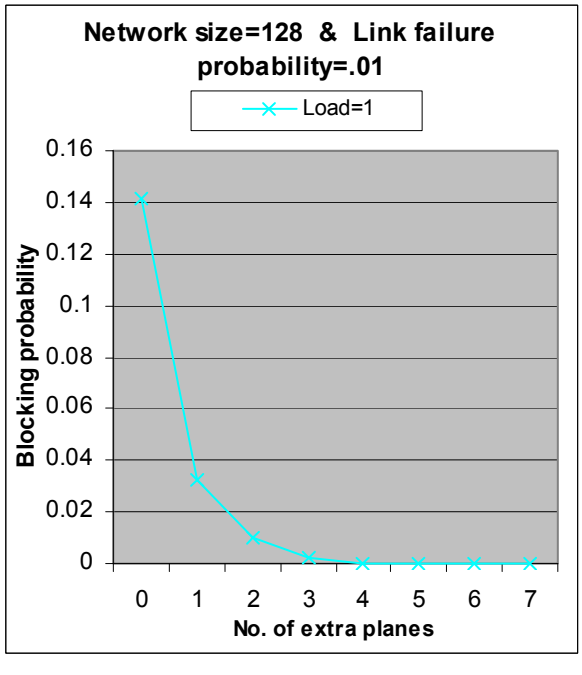


Fig. 9. Blocking probability decreases dramatically with planes.

Fig.9 shows that blocking probability decreases exponentially with the number of extra planes. For N=128, pfr=0.01 & r=1; if we add 2 extra planes to the networks then the blocking probability decreases 77% and 7 extra planes required to make the networks nonblocking.

A.2. PFR_RS algorithm

The simulation results of EP-VSOB networks using PFR_RS algorithm with link failures are given below

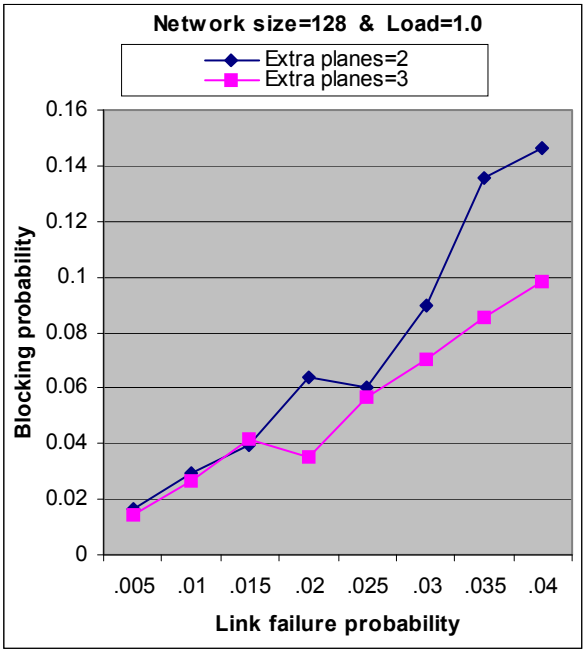


Fig. 10. Link failure vs blocking probability for different planes.

Fig.10 also shows that the blocking probability not always increases with the increase of link failure; blocking probability decreases for certain range of link failures and then increases again. In fig.10, we only present the results of extra planes 2 & 3 because for extra plane 1 the blocking behaviour of both PFR_LS & PFR_RS algorithm is the same.

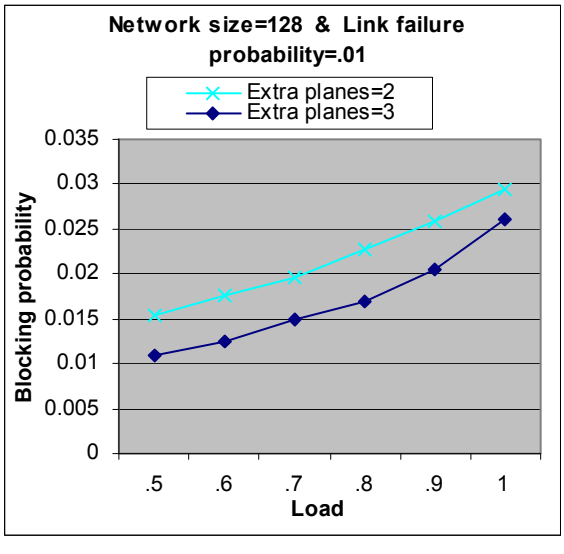


Fig. 11. Blocking probability increases linearly with load.

Blocking probability increases linearly with load as shown in fig.11. If we reduce load 1 to 0.8 then BP decreases 22.8% for extra plane =2, BP decreases 35% for extra planes =3.

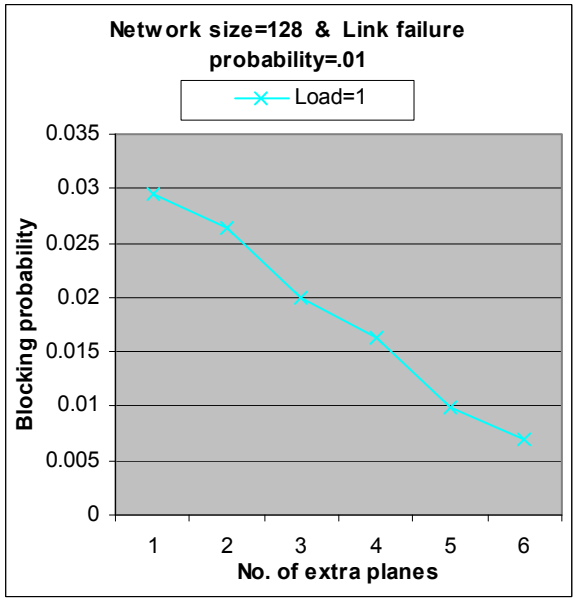


Fig. 12. Blocking probability decreases linearly with planes.

Fig.12 shows that blocking probability decreases linearly with the number of extra planes. For N=128, pfr=0.01 & r=1; if we add 2 extra planes to the networks then the blocking probability decreases 81% and 85% for 3 extra planes. It is observe from fig.9 that the decreases rate of blocking probability with the number of extra planes is low.

From fig.5 & 7 we can say that the blocking probability of P-VSOB(128,16) networks with 1% link-failures increases by 57% than P-VSOB networks without link failures. From fig.5 & 7, the blocking probability of EP-VSOB(128,16+2) networks with the same link-failures increases by 8% than EP-VSOB networks without link failures using PFR_LS algorithm and from fig.5 & 10, the blocking probability of EP-VSOB(128,16+2) networks with the same link-failures increases by 17% than EP-VSOB networks without link failures using PFR_RS algorithm.

From fig.9 and 12 we can say that for N=128, r=1 & pfr=0.01; if we add 3 extra planes to the networks then blocking probability decreases 98% for PFR_LS algorithm and 85.8% for PFR_RS algorithm.

IV. Conclusion

In this paper, we have presented simulation results of the blocking probability of EP-VSOB networks with link-failures. To guarantee the fast connection setup in the EP-VSOB networks, we use two PFR based routing algorithm, namely PFR_LS & PFR_RS algorithm. Extensive simulation results indicate that EP-VSOB networks achieve better fault-tolerant property than P-VSOB networks. Also the performance of PFR_LS algorithm is better than PFR_RS algorithm. We believe the

findings of this paper will make the adoption of banyan-based optical switches very attractive in most practical applications. A mathematical model for blocking probability of this network is under research.

References

[1] V.R.Chinni etal., "Crosstalk in a lossy directional coupler switch," *J.Lightwave Technol.,* vol.13, no.7, pp.1530-1535,July 1995.

[2] H.S.Hinton, *An introduction to Photonic Switching Fabrics*, New York: Plenum, 1993.

[3] G.R.Goke and G.J.Lipovski, "Banyan networks for partitioning multiprocessor systems," *Proc.1st Annu. Symp. Comp. Arch,*, pp.21-28,1973.

[4] C.P.Kruskal and M.Snir, "The performance of multistage interconnection networks for multiprocessors," *IEEE Trans. Commun*., vol.COM-32, pp.1091-1098, Dec.1983.

[5] F. Thomson Leighton, *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes*, Morgan Kaufmann, 1992.

[6] J.H.Patel, "Performance of processor-memory interconnections for multiprocessors," *IEEE Trans. Comput*., vol.C-30, pp.771-780, Oct.1981.

[7] C. -T. Lea, "Muti-$\log_2$N networks and their applications in high speed electronic and photonic switching systems," *IEEE Trans. Commun.,* vol. 38, pp. 1740-1749, Oct. 1990.

[8] Md. M. Rashid Khandker, X. Jiang, Pin-Han Ho, S. Horiguchi, H. T. Mouftah, "Performance of Fast Routing Algorithm in Large Optical Switches Built on the Vertical Stacking of Banyan Structures," Cluster Computing: The Journal of Networks, Software Tools and Applications, Vol 7, No. 3, pp. 219-224, 2004.

[9] Md. Mamun-ur-Rashid Khandker, Xiaohong Jiang, Masuru Fukushi, Susumu Horihuchi, "Pruned optical banyan networks on vertical stacking scheme for faster connection establishment", Graduate School of Information Sciences, Tohoku University, Sendai, Japan, Received 9 March 2005, accepted 8 September 2005.

[10]X. Jiang, P.H. Ho, H. Shen and S. Horiguchi: "Fault Tolerance Analysis of Optical Switching Systems Built on the Vertical Stacking of Banyan Network", Proc. the 2004 IEEE Workshop on High Performance Switching and Routing.

[11]Basra Sultana, M.R.Khandker, X.Jiang and S.Horiguchi: "Blocking Probability of Vertically Stacked Optical Banyan Networks with Link Failures", Proc. The International Workshop on High Performance and Highly Survivable Routers and Networks, Tohoku University, Sendai, Japan, pp. 157-169, March 14, 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Analysis of Real-Time Multimedia Traffic in the Context of Self-Similarity

Rajibul Alam Joarder[1], S. Parveen[2] MIEEE, H. Sarwar[3], S. K. Sanyal[4] SMIEEE, S. Rafique[5]

[1]Bangladesh National University, [2,5]University of Dhaka, [3]United International University, [4]Jadavpur University

Abstract- **Multimedia networks are now providing versatile services to meet different types of subscriber needs including voice, video, data, etc. Voice and video conferencing applications over these IP networks have already gained wide acceptance in today's end-user communities. The traffic characteristics and intensity as a function of time, geographic source can be realized by observing traffic distribution. Multimedia traffic has characteristics very different from Poisson characteristics, generating high rate data at one time and low rate data at another. Variable bit rate (VBR) real-time applications such as compressed video and audio and also the data sources tend to generate bursty traffic patterns and exhibit certain degree of correlation between arrivals and show long-range dependence in time i.e. self-similarity [1, 2, 3, 4, 5]. Therefore, it is important to identify and characterize network traffic flows to analyze network performance. In the present work the analysis of real time audio and video streaming traffic, colleted from the videoconferencing session of an enterprise high-speed hybrid multimedia network, has been done in the context of self-similarity.**

I. INTRODUCTION

In the new millennium, the ever changing and dynamic field of communication engineering is experiencing an explosive growth particularly in multimedia traffic (audio, video, text) and demanding the tremendous need for high-speed multimedia data transfer at a high bit rate (Tbit/s). Voice and low rate data services are insufficient for users in a world where high-speed WWW access is taken for granted. High-speed real time multimedia applications such as video on demand, voice over IP, teleconferencing, video conferencing, high-definition television (HDTV), digital libraries, remote tutoring systems require a sound understanding of the voice, video and data traffic characteristics because they possess different characteristics at optimized performance. Therefore, traffic characterization is one of the important network components to be designed carefully to ensure that a multimedia network is able to support numerous connections with QoS guarantees. A good number of recent research works convincingly demonstrate that network traffic is self-similar or long-range dependent (LRD) [1, 2, 4]. The remarkable discovery of self-similarity of aggregated Ethernet LAN traffic by Leland et al [2] has opened up a new research path for network traffic analysis and modeling. Later Paxon et al [1] showed self-similar burstiness manifesting itself in pre-world wide web WAN IP traffic. Crovella et al [4] examined reasons why TCP, FTP and TELNET traffic is self-similar. They found that because of the distribution of file sizes the distribution of transmission times may be heavy-tailed. They also found that silent times that occur due to user thinking were also heavy-tailed. Later Willinger et al [6] demonstrated that if the individual ON/OFF sources are described by heavy-tailed infinite variance distributions such as Pareto distribution, the aggregate flow shows self-similarity or long-range dependence. In a multimedia network the complexity of traffic is a natural consequence of integrating a diverse range of traffic sources over a single communication channel. The since, traditional Poisson or Markovian traffic models have been proven, by several studies [1, 2] inappropriate for predicting the performance of networks with long-range dependent traffic. To the best of our knowledge, very little work was done in the analysis of throughput performance of the videoconferencing traffic of real enterprise network. The present work gives a comprehensive treatment of one such approach. As simulation studies fail to provide a realistic environment in which voice and video over IP traffic characteristics can be understood at a fundamental level [7], hence for collecting real world videoconferencing traces from a enterprise network, a standard packet sniffer tool Ethereal was used to capture the traces and then analysis has been done on it in the context of self-similarity.

This paper is organise as follows. In section II, an brife overview of traffic self similarity is presented. Section III gives the network architecture from which the real time traffic traces were captured. Section IV deals with the statistical analysis of the captured video conferencing traffic.

In section V the discussion on the obtained observation is given. We concluding in section VI.

II. SELF-SIMILARITY

Self-similarity is a phenomenon that behaves the same when viewed at different degree of magnification or different scales on dimensions. Self-similarity indicates similar statistical behavior over a wide range of time scales and manifests itself in a number of different ways [8]. Real world traffic is highly variable over a wide range of time scales. The variability over wide time scales implies that bursts do not average out over long enough time scales. No matter at what scale we look, we will always find a similar amount of packet bursts. Self-similarity of network traffic represents Long Range Dependence (LRD) which means a correlation property over time scales on the order of a few hundreds of milliseconds and larger. LRD captures the "memory" of the behavior. Past values affect the present. It is quantified by a single scalar number: The Hurst Exponent ($0.5<H<1$) [9]. With the same arrival rate, increasing H causes increasing an average queue length. Recent research efforts have identified various reasons of self-similarity on many levels of the transmission scheme. In Fig. 1 the self-similar pattern of network traffic is shown.

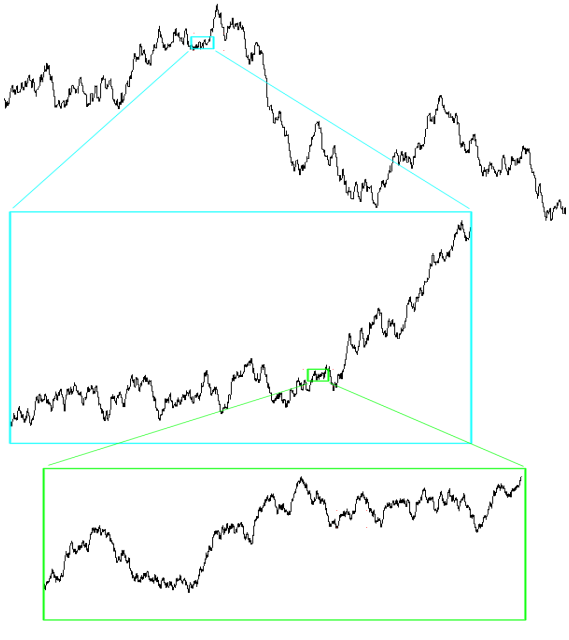


Fig.1. Self-similar pattern of Network traffic

III. ARCHITECTURE OF THE TESTBED

In the second step of the work, the behavior of TCP for different types of traffic has been investigated in detail from a real network environment. To do so, the multimedia real traffic was captured by the packet sniffer tool Ethereal from a Enterprise network with Virtual Private Network (VPN) connectivity. The network has both the wired (10/100 Mbps) and wireless (802.11g) connectivity having 120 nodes, eight different servers for specific applications and Internet connectivity speed of 4MB. Among the servers two are dedicated for audio/video conferencing. The network is well protected by firewalls. The overall view of the network is shown in Fig. 2. The traces of audio, video & text data were taken from the firewall when audio video conferencing and web browsing traffic were transmitting between the firewall and the remote server. Three routers were configured to collect Text, Audio & Video traffic separately. The captured file was of different sizes and taken at different times of a day under various loads.

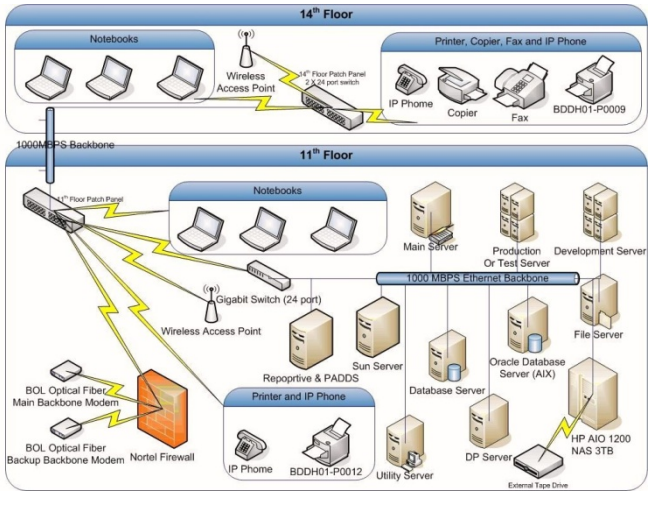


Fig. 2. Architecture of the testbed

Traffic Captured by Ethereal

As we want to focus on the characteristics of network traffic and its effects on network performance so we viewed the traffic at the network layer. Fig. 3 represents the high-level view of the captured traffic traces by Ethereal. The snap shot of Fig. 3 illustrate the frame sequences and the corresponding time stamps of the captured traffic. The next two columns represent the source and destination IP addresses among which the communications were established. The next column represents the protocols used for that particular transmission. Before analyzing the behavior of the TCP connection in detail, the packets displayed in the Ethereal window were first filtered to get the series of TCP & HTTP messages between the client computer and the server.

TCP performance parameters like throughput and Round Trip Time (RTT) have been investigated between the client computer and the server. Throughput graphs for Video traffic is given in Fig. 4**.** where simultaneous packet accumulations the same time period indicates the bursty traffic nature of video data.

	Time	Source	Destination	Protocol
18	2.046463	162.44.191.11	162.44.191.77	DCERPC
19	2.046786	162.44.191.77	162.44.191.11	DCERPC
20	2.046982	162.44.191.11	162.44.191.77	DCERPC
21	2.047079	162.44.191.77	162.44.191.11	DCERPC
22	2.047264	162.44.191.11	162.44.191.77	DCERPC
23	2.047372	162.44.191.77	162.44.191.11	DCERPC
24	2.047564	162.44.191.11	162.44.191.77	DCERPC
25	2.089764	162.44.191.51	162.44.191.127	NBNS
26	2.200864	162.44.191.77	162.44.191.11	TCP
27	2.265776	162.44.191.77	162.44.191.11	SMB
28	2.266091	162.44.191.11	162.44.191.77	SMB
29	2.267338	162.44.191.77	162.44.191.11	SMB
30	2.267530	162.44.191.11	162.44.191.77	SMB
31	2.397324	87.120.81.8	162.44.191.77	TCP
32	2.401504	162.44.191.77	162.44.191.11	TCP
33	2.415662	87.120.81.8	162.44.191.77	SSL
34	2.602160	162.44.191.77	87.120.81.8	TCP
35	2.677957	162.44.191.11	162.44.191.77	Jabber
36	2.802799	162.44.191.77	162.44.191.11	TCP

Fig. 3 part of the high-level view of the captured traffic trace by Ethereal

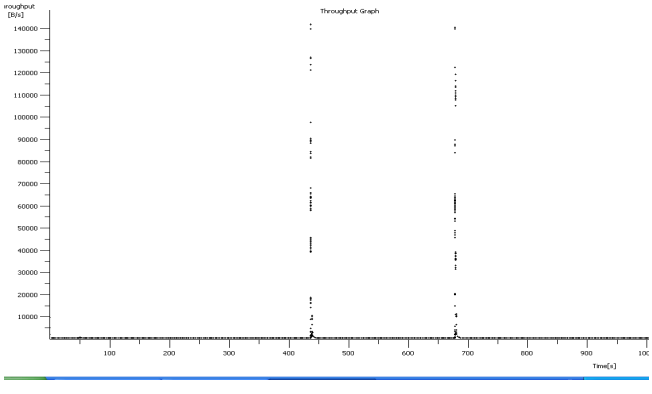


Fig. 4 Throughput graph of Video traffic

IV. STATISTICAL ANALYSIS OF AUDIO & VIDEO TRAFFIC BY MATLAB

In this section an overview of our statistical analysis of the frame size traces of audio and video have been given. The traces were processed using MATLAB to generate the required graphs. For the statistical evaluation of the traces the following notation has been introduced,

Let N denote the number of considered frames of a given traffic sequence. In case of audio sequence this would be N = 38661 and for video sequence N = 213745.The individual frame sizes are denoted by $X_1 \ldots\ldots X_N$. The mean frame size $\overline{X}$ is estimated as

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (1)$$

The aggregated frame size for the aggregation level of frames is denoted by $\overline{X}_a(j)$ and is estimated as

$$\overline{X}_a(j) = \frac{1}{a}\sum_{i=(j-1)a+1}^{ja} X_j \qquad (2)$$

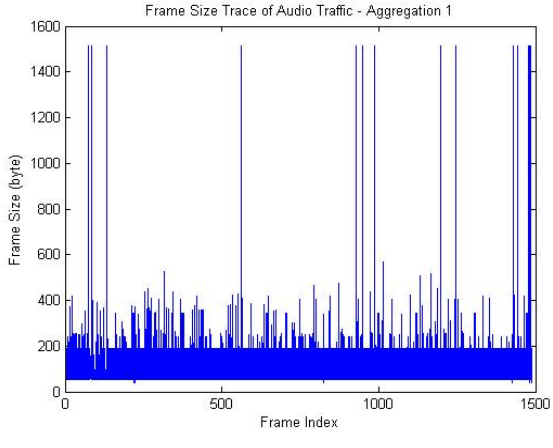


Fig. 5 Frame size distribution of audio traffic

In Fig. 5 and 7 the frame sizes versus time are given for audio and video data respectively. The aggregated frame sizes versus time is given in Fig. 6 and 8 for the aggregation level a = 4. The characteristics of the sequences are much better illustrated in the aggregation plot for both the cases. The traffic data shown in Fig. 5 and 7 is representative of the bursty nature of network traffic. The data is clearly not exponentially distributed. From the plots of Fig. 6 and 8 we see that the video sequences are clearly more variable bit rate (VBR) encoded than the audio sequence.

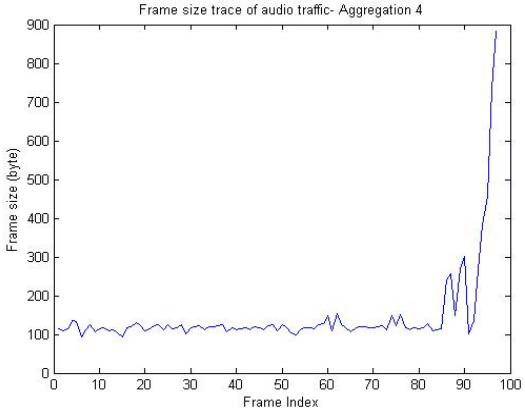


Fig. 6 Aggregated packet level of audio traffic

The Variance S^2_X of the frame size is estimated as

$$S^2{}_X = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 \qquad (3)$$

The coefficient of variation CoV of the frame size is estimated as

$$CoV = \frac{S_X}{\bar{X}} \qquad (4)$$

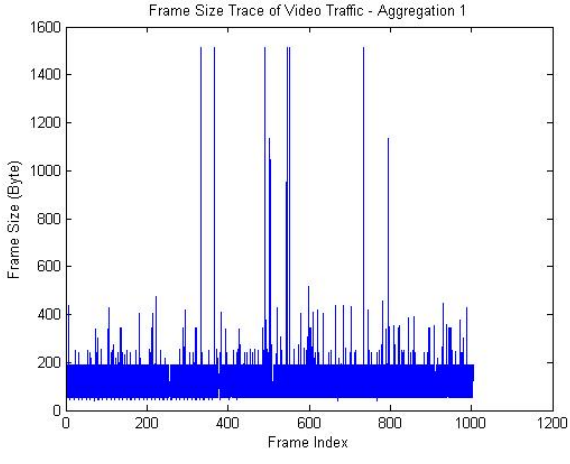


Fig. 7 Frame size distribution of video traffic

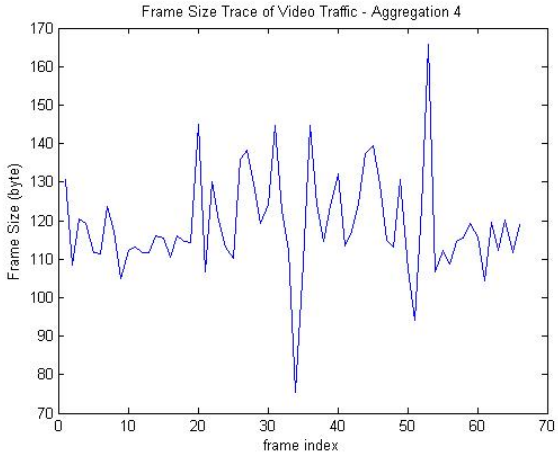


Fig. 8 Aggregated packet level of video traffic

Besides the mean and variance of the frame, the frame size distribution is very important for the network design. In addition to that the distribution of the frame sizes is needed in order to make any statistical modeling of the traffic possible. Frame size histograms or probability distributions allow us to make observations concerning the variability of the encoded data and the necessary requirements for the purpose of real-time transport of the data over a combination of wired and wireless networks. In Fig. 9 and 10 the illustration of the cumulative frame size distributions as a function of frame size for audio and video sequences respectively.

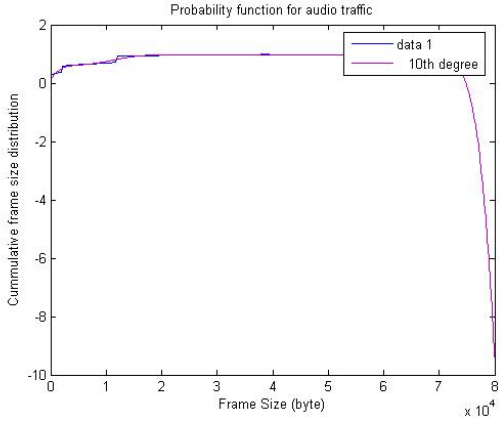


Fig.9 Cumulative distribution function of audio traffic

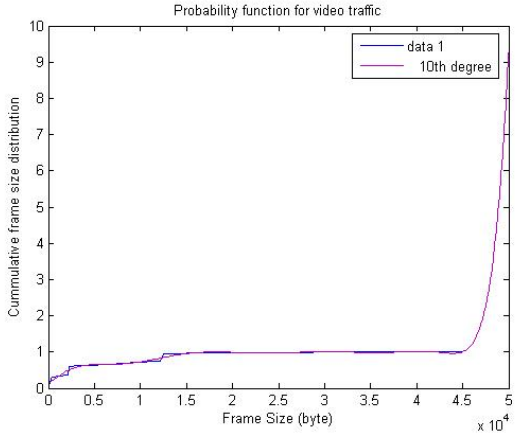


Fig.10 Cumulative distribution function of video traffic

The Hurst parameter, or self-similarity parameter, H, is a key measure of self-similarity [1, 2, 10]. H is a measure of the persistence of a statistical phenomenon and is a measure of the length of the long range dependence of a stochastic process. A Hurst parameter of 0.5< H<1 indicates the presence of self-similarity. The H parameter can be estimated from a graphical interpolation of the so-called R/S plot. The R/S plot gives the graphical interpretation of the rescaled adjusted range statistic by utilizing the following method [9]. The length of the complete series N is subdivided into blocks with a length k, for which the partial sums Y (k) are calculated as in Equation (5). Then the variance of all these aggregations is calculated. The resulting R/S value is evaluated as shown in Equation (7) for a single block.

$$Y(k) = \sum_{i=1}^{k} X_i \qquad (5)$$

$$S^2{}_X(k) = \frac{1}{k}.\sum_{i=1}^{k}\left[X_i^2 - \left(\frac{1}{k}\right)^2 .Y(k)^2 \right] \qquad (6)$$

$$\frac{R}{S}(N) = \frac{1}{S_X(k)}\left[\max_{0\le t\le k}\left(Y(t) - \frac{t}{k}.Y(k) \right) - \min_{0\le t\le k}\left(Y(t) - \frac{t}{k}.Y(k) \right) \right] \qquad (7)$$

If plotted on a log/log scale for R/S versus differently sized blocks, the result will be several different points. This plot is also called pox plot for the R/S statistics and is illustrated in Fig. 11 and Fig. 12 for the audio and video traces. The Hurst parameter H is then estimated by fitting a line to the point of the plot. The slope of the line then gives the value of H. In our present work the value of H observed for the audio sequences is 0.95 and for video sequences it is 0.55. These two graphical view suggest that both the audio & video traffic sequence is self-similar in nature as in both the cases the values are larger than 0.5.

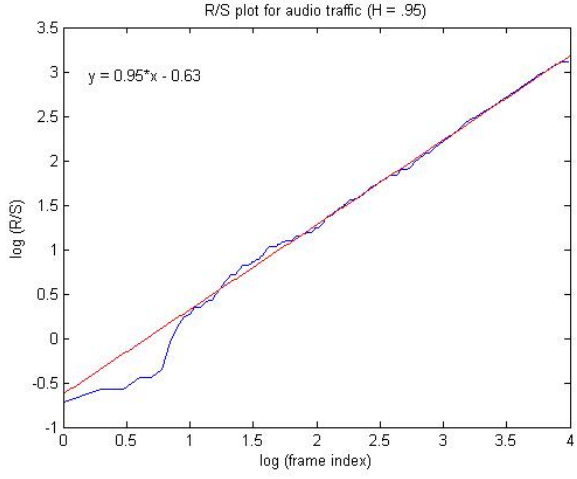


Fig. 11 Measure of H parameter for audio traffic

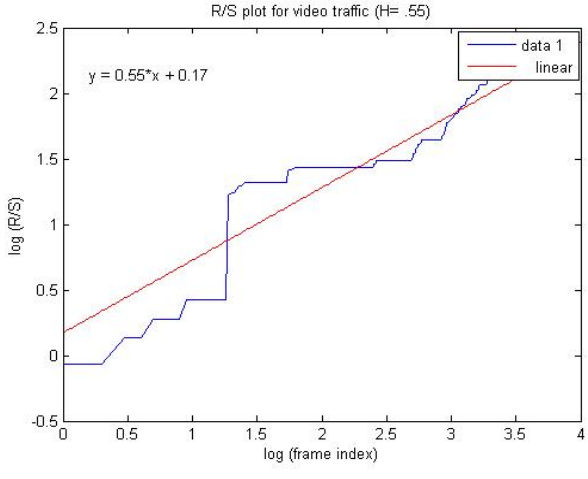


Fig. 12 Measure of H parameter for video traffic

V. DISCUSSION

The present Internet continues to experience more and more non-TCP traffic such as IP phone and video stream that may

not be TCP friendly since they do not use the TCP congestion control algorithm when confronted with router queue overflow [11]. The result of the present study will lead to better understanding of the characteristics of high-speed real-time multimedia streams in the Internet which could then result in better network design trade-offs and design of new network protocols that supplement Internet multimedia systems. Poisson model considers network arrival as a random process. Self-similarity uses autocorrelation and does not consider the network traffic to be random. Poisson Model does not scale the Bursty Traffic properly. So there is always the need to develop new models for traffic generation. In future new variants of the existing models are to be developed.

VI. CONCLUSION

In this work, the nature of real traffic collected from the videoconferencing session of an enterprise network has been observed. It is seen that real traffic usually does not follow any regular pattern. Burstyness of audio and video traffic have been clearly observed from the throughput graph of Ethereal and statistically analyzed aggregated frame size distribution graph. From the statistical analysis it has also been found that the audio and video traffic captured from videoconferencing traces shows self-similarity that indicates the need for new model to scale bursty traffic accurately.

VI. ACKNOWLEDGEMENT

We acknowledge the support of M&H Informatics Ltd for providing us network traffic data of video conference sessons.

REFERENCES

[1] V. Paxon and S. Floyed, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, pp. 226- 244, June 1995.

[2] W.E. Leland, M.S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," (extended version), *IEEE/ACM Trans. Networking*, vol. 2, pp. 1-15, Feb. 1994.

[3] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit –rate video traffic," *IEEE Trans. Commun.*, vol.43, pp.1566- 1579, July 1995.

[4] Mark E. Crovella, A. Bestvros, "Self-similarity in World Wide Web Traffic: Evidence and Possibility," *IEEE/ACM Transactions on Networking,* vol. 5, no. 6, December 1997.

[5] Huei-Wen Ferng, Jin-Fu Chang, "Connection-wise end-to-end performance analysis of queuing networks with MMPP inputs," *International Journal of Performance Evaluation, Elsevier 2001,* vol. 43, pp 39-62.

[6] W. Willinger, M.S. Taqqu, R. Sherman and D.V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM trans. Networking,* vol. 5, pp. 71- 86, Feb. 1997.

[7] Sharmin Parveen, S.K. Sanyal and Shahida Rafique, "Analysis of Self-Similarity of Multimedia Traffic in a Broadband Network," in Proceedings of 3rd International Conference on Computers and Devices for Communication (CODEC-06), Dec 2006, pp. 210-213.

[8] William Stallings, High-Speed Networks and Internets- Performance and Quality of Service, 2nd Edn., Pearson Education, 2002.

[9] Robertazzi, T.G., *Computer Networks and Systems Queueing Theory and Performance Evalution*. 3rdedn Springer-Verlag, Heidelberg, Germany, 2002 Reprint.

[10] F. Anjum and L. Tassiulas, "Balanced-RED: An algorithm to achieve fairness in the Internet," Proceedings of IEEE INFOCOM'99, March 1999.

[11] Behrouz A. Forouzan, "Data Communications and Networking, 3rd Edn, Tata McGraw-Hill Edition.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

RC4A Stream Cipher for WLAN Security: A Hardware Approach

Abdullah Al Noman[1,2], Roslina b. Mohd. Sidek [2], Abdul Rahman b. Ramli [3], Liakot Ali [4]

[1]Quantum Information Department, MIMOS Berhad, Technology Park Malaysia, 5700, Kuala Lumpur, Malaysia.
[2]Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, Serdang, Selangor D.E., Malaysia.
[3]Department of Computer and Communication System Engineering, Faculty of Engineering,Universiti Putra Malaysia, 43400, Serdang, Selangor D.E., Malaysia.
[4] Institute for Information & Communication Technology, Bokshi Bazar, BUET, Dhaka Bangladesh
E-mail: simbul74@gmail.com, roslina@eng.upm.edu.my, arr@eng.upm.edu.my, liakot@iict.buet.ac.bd

Abstract - Wireless networks are on the cutting edge of modern technology and rapidly gaining popularity in today's world due to their excellent usability. For secure wireless data transmission, Wired Equivalent Privacy (WEP), IEEE 802.11 standard defined security protocol, is employed. WEP has a potential limitation that stems from its adaptation of RC4 stream cipher algorithm. As a result, there is a pressing need for new WLAN security measure. Therefore, this paper presents hardware implementation of RC4A stream cipher and proposes to replace RC4 in WLAN security scheme, due to weakness of RC4.The design of the cipher was implemented by Verilog HDL. For hardware implementation of the design, an Altera Field Programmable Gate Array (FPGA) device, EP20K200EFC484-2X from APEX family, APEX 20KE, was used.

I. Introduction

Wireless Local Area Network (WLAN) offers organizations and users a both convenient and flexible means of communication. The popularity of wireless LANs is a testament primarily to their convenience, cost efficiency, and ease of integration with other networks and network components.It also provides mobility, enhances productivity, and lowers installation costs. The most prominent feature about WLAN is the absence of wires and its mobility. It is cable-free, no-strings-attached networking. As compares to the traditional network, WLAN requires no complicate configuration on its physical topology. Wireless LANs use electromagnetic airwaves to communicate information from one point to another without relying on any physical connection However, as data travels through the air, it can easily be tapped by any one including unauthenticated personnel using sniffer. There are number of issues that have to be considered when setting up a WLAN. The most vital is the security [5] particularly for applications hosting important information. For instance, networks transmitting credit card numbers for verification or storing sensitive information are certainly candidates for emphasizing security. In these cases and others, proactively safeguard the network against security attacks is a very important problem. The designers have tried to overcome the security concern by devising a user authentication and data encryption system known as wired equivalent privacy (WEP) [1 and 16]. In the paper we mainly focus on the WEP relevant secure problems.

WEP has a prospective weakness since its adaptation of RC4 stream cipher algorithm [17]. As a result, these networks are very susceptible to security violations and need significant encryption algorithm. In this circumstance, RC4A stream cipher can be alternative to replace RC4. As far as security concern, RC4A has an enhanced security over the RC4 against most of the known plaintext attacks [2].RC4A pseudorandom bit generator passed all the statistical tests listed in [3].

For secure high speed networks the hardware always appears to be the ultimate choice because hardware implementation of cryptographic algorithms is intrinsically more physically secure and run faster than software. Reconfigurable logic, based on Field-Programmable Gate Arrays (FPGA) devices, provides a hardware solution to algorithm flexibility [15]. FPGA is an ideal platform to provide hardware arithmetic acceleration for use in many cryptographic applications. Implementation of cryptographic algorithms in FPGA devices usually achieves superior performance when compared with software-based ones.

In this paper, for WLAN security RC4A stream cipher is implemented in FPGA. Proposed implementation supports variable key lengths from 8 bit to 512 bit. For initialization proposed implementation require 7.6 µs or 256 clock cycle, for Key Scheduling Algorithm (KSA) it require 61.2 µs or 2042 clock cycle and for Pseudo Random number Generation Algorithm (PRGA) it require 22.98 µs or 766 clock cycle. The proposed hardware implementation achieves a data throughput up to 22.28 MB/sec or 177.98 Mbits/s at frequency of 33.33 MHz. Keystream of proposed hardware implementation has passed the all the statistical tests listed in [3] which prove the randomness of the keystream. The cipher was designed using Verilog hardware description language and implemented into a single Altera APEX^{TM} 20K200E Field Programmable Gate Array (FPGA).

The paper is organized as follow. First weakness of WEP and RC4 algorithm is presented.RC4A algorithm is discussed then. This is followed by architecture of hardware, design methodology, discussion on the FPGA implementation and performance of the hardware. Finally the conclusion is presented.

978-1-4244-2014-8/08/$25.00 (c)2008 IEEE

II. WEP, RC4 Stream Cipher and Their Weakness

The concept of WEP is to prevent eavesdroppers by encrypting data transmitted over the WLAN from one point to another. Wired Equivalent Privacy (WEP) is a security protocol that is part of IEEE 802.11 standard for wireless networks. WEP is still widely employed around the world due to the fact that old network interface cards cannot match the requirements of newer security protocols. WEP uses a pre-shared key for encryption and user authentication. WEP was developed to protect link-level data during wireless transmission [14].

WEP adopts RC4 algorithm, a stream cipher, developed by RSA security. RC4 is a symmetric algorithm relies on a single shared key that is used at one end to encrypt plaintext into cipher text, and decrypt it at the other end [2 and 14]. It is a variable key-size stream cipher with byte-oriented operations. The algorithm is based on the use of a random permutation. It works in Output Feedback (OFB) mode of operation. [8]. Confidentiality in WEP is achieved using RC4 stream cipher. While remarkable in its simplicity, RC4 falls short of the high standards of security set by cryptographers, and some ways of using RC4 can lead to very insecure cryptosystems including WEP [18]. Two major weaknesses were found in RC4's key scheduling algorithm (KSA). The first being the existence of a large class of weak keys, and the second being related key vulnerability. One of WEP's biggest downfalls is that its secret keys are relatively shorter than other security protocols' keys — typically, 40 bits due to US Government restrictions on the export of cryptographic technology at the time the protocol was drafted. This key length was too short and made brute force attacks [14].

III. RC4A Stream Cipher

RC4A, an RC4 family stream cipher algorithm, developed by S. Paul and B. Preneel [13] which attempts to increase security without decreasing efficiency. Their approach essentially takes two RC4 instances and crosses information between them. RC4A stream cipher works in two phases, KSA (Key Scheduling Algorithm) phase and PRGA (Pseudo Random number Generation Algorithm) phase. During PRGA two successive output byte are generated. The goal behind RC4A was to increase security primarily by increasing the internal complexity of the algorithm [9]. RC4A is made through improvement on the RC4, i.e., providing 2 S arrays (S1 and S2) that are independent from each other and, so that RC4A should not have bias in consecutive output byte. RC4A uses three counter i, j1, and j2. Variable j1 and j2 are introduced corresponding to S1 and S2. In KSA for RC4A, like KSA of RC4, the array S1 is initialized, using the secret key K, Keystream, WK, are generated from the array S1 like PRGA (Pseudo Random number Generation Algorithm) of RC4. Then, like S1, the array S2 is initialized using WK. Unlike RC4 in PRGA of RC4A two successive output byte are generated. All the arithmetic operations are computed modulo N (N=256) [13].

A. Algorithm

KSA (K)
RC4_KSA(K, S1)
For i = 0 …l – 1
WK[i] = RC4_PRGA(S1)
RC4_KSA(WK, S2)

PRGA (S1, S2)
Initialization:
i = 0
j1 = j2 = 0
Generation loop:
i = i + 1
j1 = j1 + S1[i]
Swap(S1[i], S1[j1])
Output z1 = S2[S1[i] + S1[j1]]
j2 = j2 + S2[i]
Swap(S2[i], S2[j2])
Output z2 = S1[S2[i] + S2[j2]

IV. Hardware Architecture

Fig. 1 illustrates the functional block diagram of the hardware proposed in this paper. The hardware consists of Controller, Datapath, Storage Unit and Key Unit. Controller provides essential signals to operate Datapath, Storage Unit and Key Unit. Datapath unit is responsible for the key set up and key stream generation. As per received signal from controller it generates required signal for operation of storage and key unit. With the data from storage and key unit it executes arithmetical and swapping operation. And then generate the output. The storage unit was used to contain memory elements for S1, S2 known as S1 box, S2 box. Storage unit dealt with S1 box and S2 box. Storage Unit consists of six Rams. The Key Unit was used to contain memory elements for K array known as K box to store key of variable length from 8 bit to 512 bit. For this function one RAM was used in the hardware.

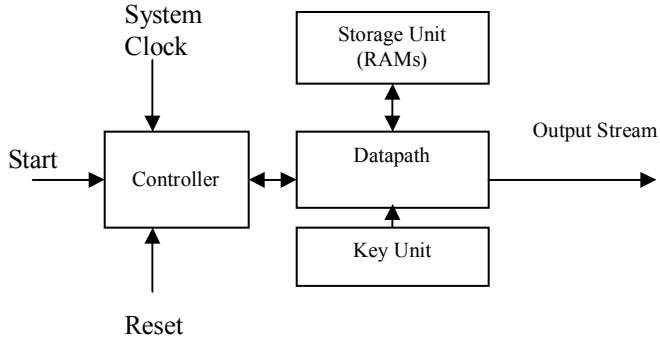


Fig. 1: Functional block diagram of the proposed hardware

V. Design Methodology

A. Design Cycle

Design cycle for our work consists of the following steps:

- Verilog implementation of the algorithm of RC4A stream cipher.
- Verifying the algorithm on Register-Transfer-level (RTL)
- Synthesis and logic optimization.
- Place and Route for specific device.

B. Software tools:

The entire design was described using Verilog HDL language. Our choice has the advantages to be portable on all circuit design platforms. For implementing proposed algorithm on FPGA device, Quartus II 5.0 development software have been used which is fully integrated package for creating for logic design for Altera FPGA. Altera Quartus II EDA tool provides a complete, multiplatform design environment for system-on-a-programmable-chip (SOPC) design. It supports system level design, FPGA, and CPLD (Complex programmable logic device) design, synthesis, place and rout, verification and device programming. Each stage of design flow can be invoked from the GUI of the Quartus II.

C. Hardware tools:

Altera Nios development kit is a board mounted with EPXAI device from the APEX family. The part number of the FPGA chip mounted on the Nios development kit used in this thesis is EP20K200EFC484-2X. It provides 8,696 registers; 106496 memory bits; and 200,000 typical gates. It contains an embedded array to implement memory functions and logic array to implement general logic functions. It has 8,320 logic elements grouped into 52 Mega Logic Array Block(LAB) structures, each of which consists of 16 logic Labs and one Embedded System Block(ESB).The ESB provides 2,048 programmable bits that can be configured as product term logic, look up table based logic, or various types of memory. The kit has other peripherals such as external SDRAM (static Dynamic Ram), SDRAM controller, watchdog timer and UART (Universal Asynchronous Receiver and Transmitter). The kit is an ideal platform for system prototyping, emulation, hardware and software development or other special requirements. It provides a flexible, powerful debug and development environment to support the development of systems using APEX devices.

VI. Hardware Implementation Results

The whole design was analyzed & synthesized by using Altera FPGA device (part number-EP20K200EFC484-2X). Synthesis results for the proposed hardware are shown in Table 1. Timing of various stages of the proposed hardware is shown in Table 2.

Table 1: FPGA Implementation Results

FPGA Device: APEX 20KE (part number-EP20K200EFC484-2X)		
Area allocation	Used/Available	Utilization
Total logic elements	480/8,320	5%
Total pins	146/376	38%
Total memory bits	10,240/106,496	9%

Table 2: Timing Results of various stages

Stage	required time	Clock cycle
Initialization	7.7 µs	256
KSA	61.3 µs	2042
PRGA	23 µs	766

Throughput of the hardware proposed can be defined as, *Throughput* = N * clock frequency. Where N is the number of bits produced in every clock cycle [7].So, Throughput of proposed hardware is 177.98 Mbits/s where the clock frequency is 33.33 MHz and N=(512x8)/766 = 5.34.

From Table 2 it is observed that proposed hardware implementation of RC4A require high time and clock cycle for KSA. It is reasonable because to increase security internal complexity of the algorithm and number of variables involved in each output byte was increased. Large number of variable increases the set up time as it involves more arithmetic operation [13].

Comparison of the results of performance (in term of throughput) and consumed area (in terms of FPGA CLB slices) for the proposed hardware and RC4 stream cipher are presented in the Table 3. For hardware implementation, of RC4 of [10] XILINX 2V250FG256 FPGA device and of RC4 of [11] XILINX XC4000E-4013EEPQ208-2 FPGA device was used. RC4 in [4] was implemented in software. From Table 3 it is observed that for the performance in term of throughput proposed hardware implementation outperforms the implementation of RC4 in [4, 10, and 11]. But it consumes relatively more area. To increase security of the RC4A primarily internal complexity of the algorithm was increased. By increasing the number of variables involved in each output the size of predictive state was increased, reducing biases. The cost is a large increase in memory requirements as well as set up time [9].

Table 3: Performance comparison of RC4A and RC4

Cipher	FPGA Device	Area (Slices)	Freq. (MHz)	Throughput (Mbps)
RC4A (Proposed)	EP20K 200EFC 484-2X	480	33.33	177.98
RC4 [10]	2V250 FG256	140	60.8	120.8
RC4 [11]	XC4000-4013E EPQ208-2	225	17.8	17.6
RC4[4] Soft-ware	-	-	150	160

The proposed design is able to support variable key lengths from 8 bit to 512 bit. The RC4 implementation in [8] is able to support variable key lengths from 8-bit to 128-bit only. The other designs [6, 11 and 12] that supports merely fixed key lengths. Therefore, it can be seen that proposed design has more option of using key compare to other RC4 stream cipher implementations and longer key will be useful for the secure use of WLAN

Unlike RC4 stream cipher, it generates two output streams at a time, whereas RC4 generate only one output stream. So, at the same time proposed hardware implementation can generate almost double number of keys with compare to RC4. Thus, it is possible to use any of keystream which increase the unpredictability of the keystream as well as security. Keystream of proposed hardware implementation has passed the all the described statistical tests listed in [3] which prove the randomness of the keystream of the proposed hardware implementation.

VII. Conclusion

Hardware implementation of RC4A stream cipher for WLAN security is presented in this paper. Hardware performance of RC4A is compared with that of RC4 which is adopted by WEP. Proposed implementation achieved better data throughput than that of RC4. Longer, variable key length and two output streams make proposed hardware more secure than RC4. Unlike RC4, RC4A considered as a secure cipher. Therefore, it is a more suitable alternative for long term privacy. Nevertheless, the cost of this improved security is longer encryption time and high consumed area.

In the light of the presented hardware implementation results, it can be said that proposed hardware be able to flexible solution for WLAN security.

References

[1] B.O. Hara and A.Petrick, *IEEE 802.11 Handbook: A Designers Companion*, IEEE Press, New York , 1999.

[2]B. Zoltak, "VMPC One-Way Function and Stream Cipher," Fast Software Encryption, FSE 2004, LNCS 3017,210-225,Springer-Verlag, 2004.

[3]B. Preneel et al., *NESSIE Security Report,* Version 2.0, IST-1999-12324, 2003.

[4]B. Schneier and D. Whiting, "Fast Software Encryption: Designing Encryption Algorithms for Optimal Software Speed on the Intel Pentium Processor," Fast Software Encryption workshop (FSE97), LNCS, Springer-Verlag, Haifa, Israel, Vol. 1267, 242-259,20-22 January 1997.

[5] D.Nagamalai, B.C.Dhinakaran, P. Sasikala, S.H.Lee and J.K. Lee, "Security Threats and Countermeasures in WLAN," Lecture Notes in Computer Science(LNCS),Volume 3837/2005 ,Springer, Berlin / Heidelberg,168-182,2005

[6]K.H Tsoi, K.H Lee and P.H.W Leong, "A Massively Parallel RC4 Key Search Engine," Proc. of the 10th Annual IEEE Symposium on Field-Programmable Custom Computing Machines ,California,13-21,2002.

[7] L.Batina, J.Lano, N.Mentens, B.Preneel, I. Verbauwhede, and S. B. Örs, "Energy, Performance, Area versus Security Trade-offs for Stream Ciphers," In ECRYPT Workshop, SASC - The State of the Art of Stream Ciphers, Brugge, Belgium,302-310,2004.

[8] M.Galanis, P. Kitsos, G. Kostopoulos, N. Sklavos, and C. Goutis, "Hardware Implementation Of The RC4 Stream Cipher," In Proceedings of 46th IEEE Midwest Symposium on Circuits & Systems, Cairo, Egypt, 2003.

[9] M.E. Mckague, Design and analysis of RC4 like stream cipher, Masters Thesis, University of waterloo, Canada, 2005.

[10]M. Galanis, P. Kitsos, G. Kostopoulos, N. Sklavos and C. Goutis, "Comparison of the Hardware Implementation of Stream Cipher," The international Arab Journal of Information Technology, Vol. 2, No.4, 2005.

[11] P. Hamalainen, M. Hännikäinen,T. Hamalainen, and J. Saar, "Hardware Implementation of the Improved WEP and RC4 Encryption Algorithms for Wireless Terminals," In Proceedings European Signal Processing Conference (EUSIPCO),Tampere, Finland,2289-2292,2000.

[12]P. D. Kundarewich, S. J.E. Wilton and A. J. Hu, "A CPLD-Based RC4 Cracking System," The 1999 Canadian Conference on Electrical and Computer Engineering Canada, 1999.

[13]S. Paul, and B. Preneel, "A New Weakness in the RC4 Keystream Generator," Fast Software Encryption, FSE 2004, LNCS 3017, 245-259, Springer-Verlag, 2004.

[14]T.Hassinen, "Overview of WLAN security," Seminar on Network Security, Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology,2006.

[15]T. Arich and M. Eleuldj "Hardware implementations of the data encryption standard," Microelectronics, the 14th International Conference on 2002 – ICM, Page(s):100 – 103, December 11-13, 2002.

[16]W. Stallings, "IEEE 802.11: Moving closer to Practical wireless LANs," IT Pro, 152, 17-23, 2001.

[17] W.C. Hsieh, Y.H. Chiu and C.C. Lo, "An Interference-Based Prevention Mechanism Against WEP Attack For 802.1 IB Network, IFIP International Federation for Information Processing," Volume 165/2005 Springer, Boston, 27-138, 2006.

[18] http://en.wikipedia.org/wiki/RC4_(cipher)

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A New Approach to Sort Unicode Bengali Text

Md. Ahsanur Rahman, and Md. Abdus Sattar

Department of CSE, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
E-mail: ah39san@yahoo.com

Abstract - Character order in Unicode for Bengali is different from the sorting order suggested by the governing authority. As a result, simple letter by letter comparison doesn't yield correct order of Bengali words. The presence of modifier characters in Bengali made the situation more complicated. The objective of our study is to adapt the suggested collation order for Unicode represented Bengali text while achieving maximum possible efficiency. Here we propose an algorithm for this purpose. The proposed algorithm is applicable to any chosen sorting order. Also it compares words in O(1) time, irrespective of their lengths. Thus complexity of sorting texts is always O(n log n).

I. Introduction

Bengali is the sixth most extensively spoken language in the world [1]. Bangla Academy [5] is the organization looking after the standardization, development and publicity of Bengali. On the other hand, Unicode is the universally used coding system that provides a unique number for every character irrespective of the platform, program and language [2]. However, the problem is that the Unicode sort order does not follow the sorting order suggested by Bangla Academy. This dissimilarity, along with, the use of conjunctive characters and modifiers in Bengali has restricted the straightforward sorting of Bengali. In this paper we shall discuss about these problems and provide solutions to them.

II. Problems in Bengali Sorting

Specifically, the problems associated with sorting Bengali words are as follows:

a. Unlike English, pair to pair sorting doesn't yield actual ordering for Bengali words [3,4]. Because the sorting order suggested by Bangla Academy differs from Unicode ordering for some letters. For e.g. in Unicode ং ≺ অ [6], but according to the Bangla Academy Dictionary অ ≺ ং [7]

b. There are frequent uses of conjunctive characters which are formed by placing a link character (generally ্) between two consonants. For e.g. ক + ্ + ত = ক্ত , ঞ + ্ + জ = ঞ্জ While comparing two words, the link character must be ignored.

c. In Bengali, vowels and consonant modifiers are placed before and/or after the base letter. For e.g. ে + ক = কে, ক + া = কা etc. Since Bengali sorting is primarily based on base letter values, modifiers must always be placed after the base letter, to obtain proper order.

III. Previous Works

Many research works have been done to solve the problem of sorting Bengali texts. Some of the most prominent ones among them are described below:

a. Palit and Sattar [8] proposed to type the words as they are spelled (this is called phonetic typing [9]). This solves the 3rd problem specified above. To solve the 1st problem, each letter is mapped to a unique ASCII value so that they follow the order suggested by Bangla Academy i.e.

Vowels < Consonants < Vowel modifiers

This approach has at least two shortcomings:

(i) It doesn't use Unicode – which questions its universal acceptability.

(ii) It is not flexible enough to be applied for any desired sorting order.

b. The method proposed by Emrul and Masroor [3] was somewhat similar to the method described in [10] and was based on an ancillary map where the proper collation weights of Bengali Unicode characters and their types were specified. The method shifts left biased modifiers to the right of the base letter at first which solves the 3rd problem. Blank modifiers are inserted after base letters having no modifiers. Then the word_value for each word is calculated where

$$\text{word\_value} = \sum_{i=0}^{n-1} a_i d^{-i}$$

Here n = number of letters in the word, a_i = collation weight of i'th letter in the word and d = 100. The word values are then sorted by any standard sorting technique to obtain the sorted list of words.

This approach has the following shortcomings:

a. Storage is wasted for storing blank modifiers.

b. The approach is practically infeasible in presence of long words in the input list. To prove this, the

CPP code emulating the encoding technique of [3] to encode the word কর্মচারী is given below:

```
int s[N] = {25,0,54,63,51,0,30,3,54,7}; // it represents the
                    //list of collation weights for the word কর্মচারী
double w=0.0,e=1.0;
for (int i=0; i< N; i++) {
    w += s[i]/e; e = e*100;
}
```

It generates word_value, w = 25.005463510030036600 for কর্মচারী instead of generating the result presented in [3] (25.005463510030035407). Apparently it happens due to round off error. This error may cause wrong output. For e.g. the word কর্মচালু also produce same word_value as কর্মচারী So the word pair (কর্মচালু, কর্মচারী) would remain same after sort using [3], though in actual order কর্মচারী ≺ কর্মচালু

IV. Proposed Algorithm

Let's rename base letter as B, modifier as M and link character as L. Since according to [7]:

$$B \prec BM \prec BLB \text{ (conjunctive)}$$

we assign each letter a unique collation weight so that the weights follow the suggested order (to beat the 1st problem). That is:

(Weight of B) < (Weight of BM) < (Weight of BL)

To solve the 3rd problem we use phonetically typed input. Inputs typed in other schemes must be converted in such a way that the modifiers always take place after the base letters (as in [3]), before applying the algorithm.

The 2nd problem was beaten by the step 4.1.3 of the algorithm.

Step 5 of the algorithm minimizes storage requirement by converting weight strings into bit strings.

The proposed algorithm is given below:

1. Read desired orders of base letters (vowels or consonants) and modifiers from the user
2. Create a collation weight table (CWT) where collation weight for a modifier = (Its position in the given order) and collation weight for a base letter = 12 * (Its position in the given order)
3. Read input word list
4. For each word in the list
   - 4.1 For each letter in the word
     - 4.1.1 Find the collation weight (c) of that letter from the CWT
     - 4.1.2 If the letter is a base letter then Append c to the weight string w
     - 4.1.3 Else Add c with the last element of w
5. Let, maximum length of weight string = L and
   K = (L ≤ 6) ? 64 : ((L ≤ 12) ? 128 : 192)
   Convert each weight string to K bits long bit string where each weight occupies 10 bits from MSB
6. Sort the bit strings by a stable sort technique
7. Remap each bit string to corresponding word

V. Correctness Proof

Claim 1: The algorithm encodes the letters in such a way that the weights given to them retain same order as of the letters themselves.

Proof: It suffices to prove that, For each pair of letters w_i and w_j if $w_i \prec w_j$ in given order then $wt(w_i) < wt(w_j)$ where wt(x) = weight of letter x

The proof is based on the following observations:

(i) Bangla word = $(B \circ N?)^+$

(ii) According to [7] : $B_i \prec (B_i \circ N_j) \prec (B_i \circ N_{j+k}) \prec B_{i+1}$

Where B = a base letter, N = a non-base letter, B_i = i 'th base letter and N_j = j 'th non-base letter in the suggested ordering.

So we have to prove that:

$wt(B_i) < wt(B_i \circ N_j) < wt(B_i \circ N_{j+k}) < wt(B_{i+1})$

i.e., $(12*i) < (12*i + j) < (12*i + j + k) < (12*(i+1))$

i.e., $0 < j < j + k < 12$

The 1st 2 relations are obvious. So we have to prove the last one: $j + k < 12$. Since there are 11 non-base letters in Bengali, index (h) of N_h can be maximum 11 i.e. $h < 12$ so $j+k < 12$.

Claim 2: Comparing bit strings in stead of weight strings doesn't change the output that would be generated if just weight strings would be compared.

Proof: Since the maximum weight in a weight string = 12*(maximum index of base letter) + (index of link character) = 12*50 + 11 = 611, we need at least 10 bits to represent each weight. According to [11] the longest Bengali word is অঘটনঘটনপটিয়সী which would generate a weight string of length 11 i.e. we have to use at least 110 bits to represent a bit string. For safety we used up to 192 bits. So no truncation error will occur in step 5. So comparing bit strings in stead of weight strings won't change the output of the algorithm.

Since claim 2 is correct we can say that our algorithm is based on comparing weight strings. Since claim 1 is also correct the output generated by this algorithm conforms to the given collation order.

VI. Experimental Result

We implemented the proposed algorithm using JAVA and used the program to sort the word pair (ক্ষতি, কৈ) using the suggested ordering. The suggested ordering for base letters is:

অ ,আ, ই, ঈ, উ, ঊ, ঋ, এ, ঐ, ও, ঔ, ং, ঃ, ঁ, ক, খ, গ, ঘ, ঙ, চ, ছ, জ, ঝ, ঞ, ট, ঠ, ড, ড়, ঢ, ঢ়, ণ, ত, থ, দ , ধ, ন, প, ফ, ব, ভ, ম, য, য়, র, ল, শ, ষ, স, হ

Other letters are ordered as: া, ি, ী, ু, ূ, ৃ, ে, ৈ, ো, ৌ, ্

According to this ordering the weights for the letters ি, ৈ, ্, ক, ত, ষ are shown in Table 1.

Table 1: Weights of different letters

Letter (x)	Weight(x)	Letter (x)	Weight(x)
ি	2	ক	15*12=180
ৈ	8	ত	33*12=396
্	11	ষ	48*12=576

Table 2 shows the weight strings as well as bit strings for the words ক্ষতি and কৈ computed by our program.

Table 2: Weight Strings and bit strings for ক্ষতি and কৈ

Word	Weight string	bit string (in hex)
ক্ষতি	{191, 576, 398}	2FE4063800000000H
কৈ	{188}	2F00000000000000H

So after sorting the bit strings and decoding them to corresponding words we get the pair (কৈ, ক্ষতি) which is correctly sorted.

VII. Conclusions

Because our algorithm uses integer numbers rather than floating point numbers, it is applicable in presence of even the largest word in Bengali. Besides, it ensures efficient sorting and enables the user to change the collation order. Also we succeed to find many mistakes in the dictionary [5,7] using this program. For e.g.: ক্রীতদাস ≺ ক্রীতক in dictionary. But our program gives: ক্রীতদাস ≺ ক্রীতক. In future we shall try to improve the time and space complexity of the algorithm.

References

[1] http://www.worldlanguage.com/Languages/ Bengali.html

[2] The Unicode Consortium, *The Unicode Standard,* Version 4.0, Reading, MA, Addison-Wesley, 2003.

[3] Shah Md. Emrul Islam and Muhammad Masroor Ali, "An Approach to Sort Unicode Bengali Text Using Ancillary Maps", Asian Journal of information Technology, 4(10):890-894. © Grace publications, 2005.

[4] Md. Shahidur Rahman and Md. Zafar Iqbal, "Bangla Sorting Algorithm: A linguistic approach", in Proceeding of ICCIT (International conference on Computer and Information Technology), Dhaka, Bangladesh, pp.204-208, 1998.

[5] Hoque M.E., S. Lahidi and S.Sarker, *Baboharik Bangla Ovidhan*, Bangla Academy, 1992

[6] http://www.unicode.org/charts/PDF/U0980.pdf

[7] M. Ali, M. Moniruzzaman, J. Tareque, *Bangla - English Dictionary*, Reprinted 1st edition, published by Bangla Academy, 1994.

[8] Rajesh Palit, Md. Abdus Sattar, "Representation of Bangla Characters in the Computer Systems", Bangladesh Journal of Computer and Information Technology, Vol. 7, No. 1, December, 1999.

[9] http://www.ekushey.org/?page/phonetic_bangla_typing

[10] Md. Sharif Uddin, Rahat Khan, A.B.M. Tariqul Islam, S.M. Rafizul Haque, "A New Approach in Computer Representation of Bangla Words and Bangla Sorting Algorithm", in Proceeding of National Conference on Computer Processing of Bangla, Dhaka, Bangladesh, 2005.

[11] http://www.banglaprosar.org.au/

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A 7-Segment Display for Bangla, English and Other Indian Numerals

M Midul Islam, Mohammad Kabir Hossain, Khondker Shajadul Hasan, Abul L Haque

Department of Computer Science and Engineering, North South University
12, Kemal Ataturk Avenue, Banani, Dhaka-1213, Bangladesh
E-mail: midul_007@yahoo.com, {mkhossain, shajadul, ahaque}@northsouth.edu

Abstract - 'Seven segment display' is a very popular device that displays Indo-Arabic numerals by turning on or off the individual segments in it. Currently this display device is used to represent the numerals of almost all European languages. However Indian languages are lacking of a proper segmented display device to display its numerals. Over the past few years many research works have been done with Bangla language which result many proposals to display Bangla numerals. However we found very few significant works in displaying the numerals of other Indian languages. Also in the proposed devices for displaying Bangla numerals we found the number of segments are 16, 11, 10, 9, 8, 14 (dual 7 segments) etc. In this paper we are proposing a unified seven segment display device which is capable of displaying Bangla, English and most of the Indian languages' numerals that include Assamese, Manipuri, Hindi, Nepali, Marathi, Gujarati, Punjabi, Kannada, Telugu, Sikkim, Bhutan etc. Our proposed display device has the same simplicity that of English with an acceptable level of accuracy and uniformity.

Keywords: - Seven segment display, Unified display device, Bangla and English numerals, Indo-Arabic numerals.

I. Introduction

In different electrical and electronics devices and in some work places displaying numeric digits is extremely necessary. For many years a very popular seven segment display is being used in different places to serve this purpose. Unfortunately this display device can display only the English digits. So in spite of the necessity of showing the digits in local languages, sometimes it does not become possible. Many researchers have come forward to display Bangla numerals in dot matrix and segmented display [1-7]. However dot matrix representation would require a large number of dots to be manipulated at a time which would make the underlying logic circuitry complex and hence inefficient. So researchers moved to the segmented display where the same problem of dot matrix representation must be considered as well. For that, the number of segments must be kept as minimum as possible. Also in the multi-lingual countries it is highly desirable that one single display system must display a numeric figure in multiple languages where one of the languages should be English. For that a unified display is needed which is capable of displaying the numerals of multiple languages including English. In this paper we have proposed a unified segmented display technique which requires only seven segments like the popular seven segment display of English. This display device shows not only Bangla and English numerals but also numerals of many other languages used in the Indian sub continent. This display device is unique as it has the lowest number of segments ever proposed at the same time it is multi lingual covering at least thirteen languages of Indian sub continent.

II. Review of Other Existing Models

In early days Bangla numerals would represented in 3x8 dot matrix display [2] which would require manipulating 24 logic functions. First ever segmented display of Bangla numerals is found in [3] which used two 7 seven segment English display together. In [4] Ahmed Yousuf Saber et al were first to proposed a single display unit of 11-segment to display Bangla numerals. But their device suffers from too many segments and also some of segments have many fold curves which is very difficult to manufacture. In [5,6,7] we see other proposals of 9, 10, 16 segments all of them suffer of having too many segments. In [8] we find a 13-segment display for devanagari numerals which are used in Hindi, Nepali and Marathi language. A major break through is found in [1] where Fazle Rabbi, Mohammad Kabir Hossain and Monzur Ahmed proposed an eight-segment display of both Bangla and English numerals. Their proposed model uses not only lowest number of segments but also it can display the numerals of two languages although the visual quality of some of the digits are not that much good. In [2] Shanawaz Islam et al proposed another 8 segment display which is someway derived from [1] after 45^0 rotation of the segments and two opposite bend curves converted to L shaped and two others made straight. Although it gives better display quality for some of the digits but lost the capability of showing the English digits. Also the display quality of Bangla 3 and 8 is not up to the mark in this model.

III. The Proposed Seven Segment Display

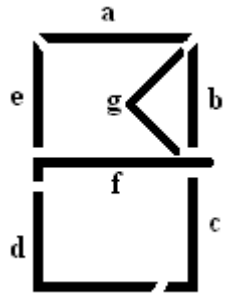


Fig. 1 The proposed 7-segment display

Fig. 1 shows our proposed model of the seven segment display. The segments are designated with English letters a, b, c, d, e, f and g. The segments are non-overlapping and non-bending. Among them a, b and e are straight while c, d, f and g are made L shaped to improve the visual quality of the digits. With the help of these seven segments we successfully displayed all the numerals of Bangla, English, Hindi, Assamese, Nepali, Marathi, Gujarati, Manipuri, Punjabi, Telugu, Kannada, Sikkim, Bhutan languages that came from eight different language families.

IV. Representation of the Numerals in Different Language

In the table 1 we showed how the segments of our proposed model will be turned on to display the numerals. In the left most column of the table the language family is shown and next column shows the name of the languages that belong to that family. Using table 1 we can easily find the truth tables for each segment, functional representation of each segment and finally the logic circuit of individual family using different methods [10].

Table 1 Representation of different numerals in proposed model

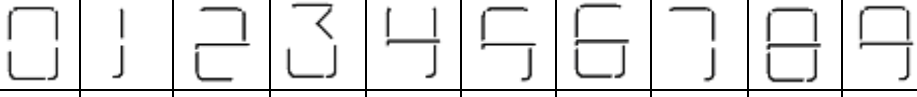
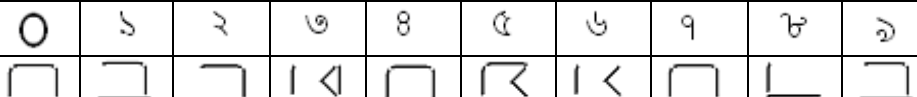
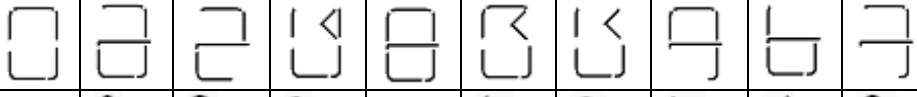
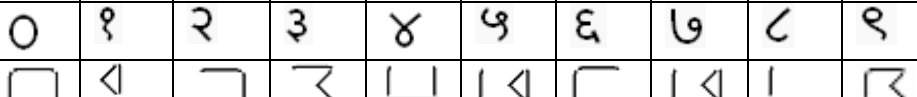
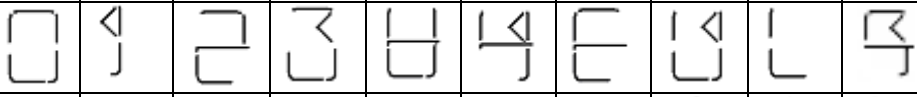
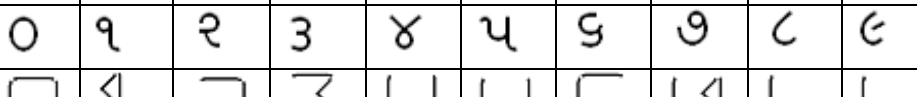
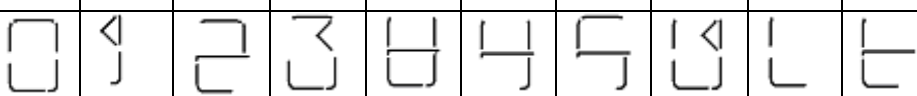
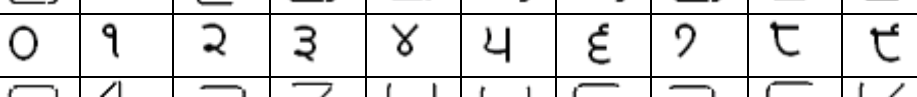
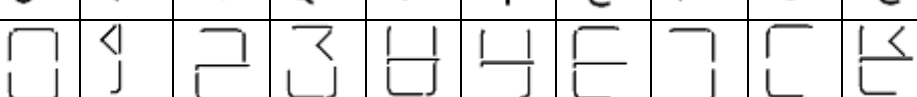
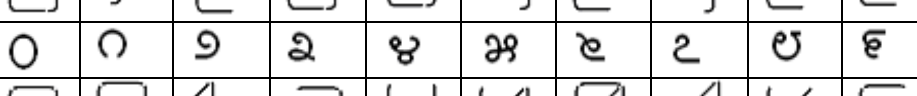
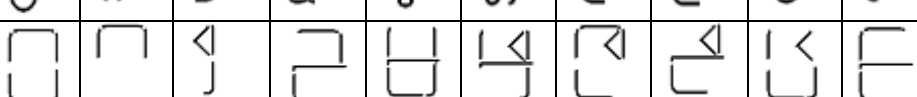
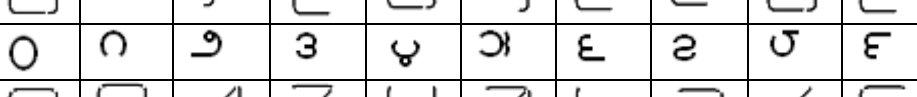
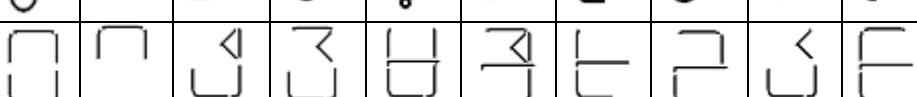
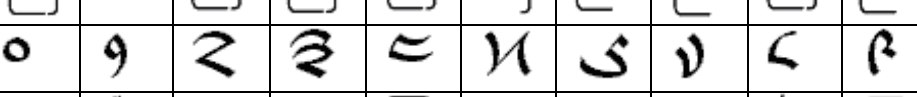
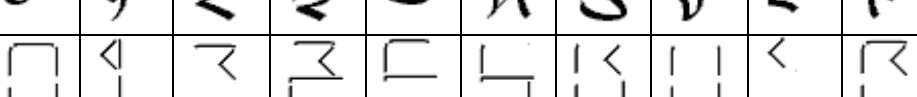
Language family	Language used in	0	1	2	3	4	5	6	7	8	9
European	English, others										
Eastern Nagari	Bangla, Assamese, Manipuri										
											
Devanagari	Hindi, Nepali, Marathi										
											
Gujarati	Gujarati										
											
Gurumukhi	Punjabi										
											
Kannada	Kannada										
											
Telugu	Telugu										
											
Lepcha	Sikkim, Bhutan										
											

Table 2 Combination of different segments in each numeral

Language family	Language used in	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
European	English, others	a,b,c, d,e	b,c	a,b, f,d	a,g, c,d	e,f, b,c	a,c, e,f	a,c,d e,f	a,b,c	a,b,c d,e,f	a,b,c e,f
Eastern Nagari	Bangla, Assamese, Manipuri										
		a,b,c, d,e	a,b,c d,f	a,b, f,d	b,c, d,e,g	a,b,c d,e,f	a,c,d e,g	c,d e,g	a,b,c e,f	c,d, e,f	a,b, c,f
Devanagari	Hindi, Nepali, Marathi										
		a,b,c, d,e	b,c,g	a,b, f,d	a,g, c,d	b,c, d,e,f	e,f, b,c,g	a,d e,f	b,c, d,e,g	d,e	a,c,e, f,g
Gujarati	Gujarati										
		a,b,c, d,e	b,c,g	a,b, f,d	a,g, c,d	b,c, d,e,f	e,f, b,c	a,c, e,f	b,c, d,e,g	d,e	d,e, f
Gurumukhi	Punjabi										
		a,b,c, d,e	b,c,g	a,b, f,d	a,g, c,d	b,c, d,e,f	e,f, b,c	a,d e,f	a,b,c	a,d,e	d,e, f,g
Kannada	Kannada										
		a,b,c, d,e	a,b,e	b,c,g	a,b, f,d	b,c, d,e,f	b,c,ef, g	a,b,d, e,g	b,d,fg	c,d e,g	a,d e,f
Telugu	Telugu										
		a,b,c, d,e	a,b,e	b,c, d,g	a,g, c,d	b,c, d,e,f	a,b,c f,g	d,e, f	a,b, f,d	c,d,g	a,d e,f
Lepcha	Sikkim, Bhutan										
		a,b,c, d,e	b,c,g	a,g	a,d, f,g	a,e,f	c,e,f	c,d e,g	b,c,de	g	a,d,eg

Table 2 shows the corresponding combination of the segments for each numeral. Tables 3 to 10 show the sum of product (SoP) functions of each segment from a to g for each of the language family.

Table 3 SoP function for European family

Segment	Sum of product function
a	Σ(0,2,3,5,6,7,8,9)
b	Σ(0,1,2,4,7,8,9)
c	Σ(0,1,3,4,5,6,7,8,9)
d	Σ(0,3,6,8)
e	Σ(0,4,5,6,8,9)
f	Σ(2,4,5,6,8,9)
g	Σ(3)

Table 4 SoP function for Easter Nagari family

Segment	Sum of product function
a	Σ(0,1,2,4,5,7,9)
b	Σ(0,1,2,3,4,7,9)
c	Σ(0,1,3,5,6,7,8,9)
d	Σ(0,1,3,4,5,6,8)
e	Σ(0,3,4,5,6,7,8)
f	Σ(1,2,4,7,8,9)
g	Σ(3,5,6)

Table 5 SoP function for Devanagari family

Segment	Sum of product function
a	Σ(0,2,3,6,9)
b	Σ(0,1,2,4,5,7)
c	Σ(0,1,3,4,5,7,9)
d	Σ(0,2,3,4,6,7,8)
e	Σ(0,4,5,6,7,8,9)
f	Σ(2,4,5,6,9)
g	Σ(1,3,5,7,9)

Table 6 SoP function for Gujarati family

Segment	Sum of product function
a	Σ(0,2,3,6)
b	Σ(0,1,2,4,5,7)
c	Σ(0,1,3,4,5,6,7)
d	Σ(0,2,3,4,7,8,9)
e	Σ(0,4,5,6,7,8,9)
f	Σ(2,4,5,6,9)
g	Σ(1,3,7)

Table 7 SoP function for Gurumukhi family

Segment	Sum of product function
a	Σ(0,2,3,6,7,8)
b	Σ(0,1,2,4,5,7)
c	Σ(0,1,3,4,5,7)
d	Σ(0,2,3,4,6,8,9)
e	Σ(0,4,5,6,8,9)
f	Σ(2,4,5,6,9)
g	Σ(1,3,9)

Table 8 SoP function for Kannada family

Segment	Sum of product function
a	Σ(0,1,3,6,9)
b	Σ(0,1,2,3,4,5,6,7)
c	Σ(0,2,4,5,8)
d	Σ(0,3,4,6,7,8,9)
e	Σ(0,1,4,5,6,8,9)
f	Σ(3,4,5,7,9)
g	Σ(2,5,6,7,8)

Table 9 Sop function for Telugu family

Segment	Sum of product function
a	Σ(0,1,2,5,7,9)
b	Σ(0,1,2,4,5,7)
c	Σ(0,2,3,4,5,8)
d	Σ(0,2,3,4,6,7,8,9)
e	Σ(0,1,4,6,9)
f	Σ(4,5,6,7,9)
g	Σ(2,3,5,8)

Table 10 Sop function for Lepcha family

Segment	Sum of product function
a	Σ(0,2,3,4,9)
b	Σ(0,1,7)
c	Σ(0,1,5,6,7)
d	Σ(0,3,6,7,9)
e	Σ(0,4,5,6,7,9)
f	Σ(3,4,5)
g	Σ(1,2,3,6,8)

V. Functional Design of the Model

The BCD to seven segment decoder of our proposed model can be designed with help of the tables 3 to 10. We can implement the functions for a to g for each language family. The circuit will take BCD input and will give us 8 sets of functions for a to g. Now we will use 7 multiplexers of 8×1 type (TTL 74151 IC) [11] to select one set of functions of a to g for a particular language family. Three common selector bits will be connected to all the multiplexers to select those functions. Now at any time the display mode can be changed to switch to different language family by changing the selector bits. The fig 2 shows the block diagram design of our proposed model.

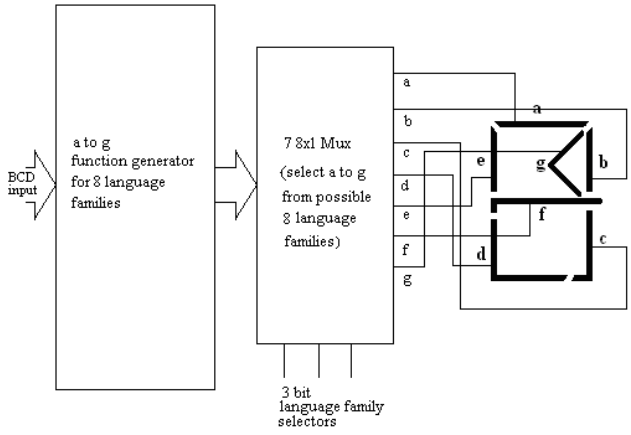


Fig. 2 The block diagram of the decoder of the proposed model

VI. Advantages in the Proposed Model

The seven segment display model that we proposed is a novel one as this is the first of its kind to display the numerals of Bangla, English and many other Indian languages using as few as seven segments. Also in the display model there is no curved segment and no segment overlapping that makes the device easy to manufacture. The proposed decoder of our model shown in fig 2 is also efficient since the language can be easily selected with help of 3 bit selectors.

VII. Conclusion

To the best of our knowledge this is the only model that displays numerals of so many languages in one device with only seven segments. We hope that this display device will become popular in Indian sub continent and will be used successfully.

References

[1] Fazle Rabbi, Mohammad Kabir Hossain and Monzur Ahmed, "An 8-segment display for both English and Bangla digits," in Proceedings of 6th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2003, pp. 338-341

[2] Shanawaz Islam, M. G. Rabiul Alam, Md. Nizam Uddin, "An 8-segment display for simple and accurate representation of Bangla numerals," in Proceedings of 4th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2006, pp. 185-188

[3] Gahangir Hossain and A.H.M. Ashfak Habib, "Designing Twin 7-segment Display for Bangla Numeric Characters", Asian Journal of Information Technology 3(1): 56-68, 2004, Grace Publications Network, 2004.

[4] Ahmed Yousuf Saber, Mamun Al Morshed Chowdhury, Suman Ahmed, Chowdhury Mofizur Rahman, "Designing 11-segment Display for Bangla Digits," in the Proc. 5th ICCIT, Dhaka, Bangladesh, 2002, pp. 237-240.

[5] N. Mahmud and Md. M Rahman Khan, "Designing 9-segment display for Bangla digits," in Proceedings of 3rd International Conference on Electrical, Electronics and Computer Engineering (ICEECE), Dhaka, Bangladesh, 2003.

[6] Md. Abul Kalam Azad, Rezwana Sharmeen, Shabbir Ahmed and S. M. Kamruzzaman, "A unique 10 segment display for Bangla Numerals," in the Proc. 8th ICCIT, Dhaka, Bangladesh, 2005.

[7] A.K.M Khaled Ahsan Talukder and K. Roy "Development of 16-segment multilingual Display Driver," in Proceedings of 3rd International Conference on Electrical, Electronics and Computer Engineering (ICEECE), Dhaka, Bangladesh, 2003.

[8] Lau, K. T. Jawalekar, S. R, "A display system for Hindi numerics", IEEE Transactions on Consumer Electronics, vol 34, issue 2, pp. 352-356, May 1988.

[9] "Indian numerals – Wikipedia, the free encyclopedia", to be found at http://en.wikipedia.org/wiki/Hindi_numerals

[10] M. Morris Mano, "Digital Design", 3rd edition, Pearson Education Inc. 2004

[11] TTL logic Gates, Data Handbook.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Proposal for a Generic Multimedia Framework with a view to facilitate easy middleware development

Sachin P. Kamat

Lead Engineer, System LSI Division, Samsung India Software Operations Pvt. Ltd.
66/1,Bagmane Tech Park, C.V.Raman Nagar, Bangalore, India. 560093.
E-mail: sachin.kamat@samsung.com

This paper proposes a generic multimedia framework which is application independent that will facilitate developers to integrate codec libraries (both software and hardware) efficiently with the standard multimedia applications. This will reduce the development time. It will also relieve the application providers of shipping the codec libraries along with their players unless it's a very proprietary codec. There will be only one copy of a particular codec in any electronic device with multimedia support. Thus a lot of space is saved especially in embedded applications where memory requirement is a constraint. The applications can also utilize the hardware acceleration features provided by the device thereby giving enhanced performance and reducing the CPU overhead.

I. Introduction

Multimedia is a very fascinating technology that finds place in everyday life. More and more devices especially handhelds are coming up with multimedia support. Multimedia in this context refers to the device's ability to support image, audio and video. Support, here, refers to the capability to record and play (display) images, audio and video, streaming and non-streaming, content using either hardware or software codecs with the help of multimedia applications.

Multimedia plays an important role in handheld devices like PDAs, mobile phones, smart phones, etc. These devices come with different real time operating systems. Many multimedia applications come as a package with these operating systems or they can be installed subsequently. Examples of such applications include multimedia player, camcorder, video telephony, video on demand, etc.

There is however one major constraint with any embedded system, i.e., the memory. The memory should be used judiciously and conservatively. Moreover handhelds these days come with hardware codecs and accelerators built as a part of the chip. These hardware codecs increase the performance of multimedia data processing tremendously. Hence they should be made use of by the applications. Several frameworks like [1] have been proposed by various bodies but there is no unified framework catering to various types of applications like playing, recording, streaming, etc.

The paper is organized as follows. Issues with the existing multimedia system are discussed in section II. Section III discusses the proposed objectives. Some information about the various multimedia elements is discussed in section IV. Section V discusses the proposed framework and conclusions are given in section VI.

II. Issues with the Existing Multimedia System

This section lists the core issues that are prevalent in the present systems. They are as follows.

1. Every application (player) capable of playing multimedia data comes with its own library of codecs which it can handle. The more the players, the more is the redundancy in the codecs because the new player which can decode the same file format as the already existing one will once again contain its own copy of that particular codec. This is a burden on the memory especially in embedded systems because of duplication of codecs.
2. The applications cannot make use of the hardware capabilities provided by the devices since there is no common interface.
3. Time to market is one of the major factors that determine business potential in today's competitive world. When it comes to middle ware development, the integration of codecs is not a generic process but an application dependent one. Hence the codecs are as such not portable and requires the study of application for which they need to be integrated. This is time consuming and non-portable process.
4. Besides, the time taken to develop an application is also more.
5. The application cannot be upgraded easily.

The table below lists the sizes of some popular decoder libraries highly optimized for an ARM 11 based system. From the table it is clear that the memory consumed by them is considerable.

Table 1 Library sizes of popular decoders.

Decoder	Library Size (in KB)
AAC	140
MP3	74
H.264	188
MPEG-4	165

III. The Proposed Objectives

The paper proposes the following objectives.

1. To make the multimedia application independent of the codec library, i.e., there will be only one standard common library in a system and any multimedia application will use the codecs only from that library. This will save a lot of memory.
2. To make the codec library generic.
3. To allow the applications to make use of hardware codecs and accelerators.
4. Development time is reduced for both the applications as well as for the systems.
5. The library will become portable since it is generic.
6. Upgrading the existing applications with newer codecs is easy and single shot for all applications (provided the applications are scalable).

IV. Information about Multimedia Elements

This section describes some basic concepts about multimedia data and applications. The basic components of multimedia are the multimedia data and the multimedia application apart from the multimedia hardware.

The multimedia data can be either streaming or non-streaming. Non-streaming data is the static file residing in the memory of the device. Any multimedia data is composed mainly of two parts:

- Stream information
- Multimedia data generally in encoded form

The file information is the file header or the metadata which contains information specific to the data like file format, length, bit rate, number of channels, etc. The data is the actual multimedia data. Any multimedia application is generally made up of the following components:

1. Input reader
2. Information extractor known as parser
3. Stream demultiplexer (in case of decoding applications)
4. Encoder/decoder
5. Multiplexer (in case of encoding application)
6. Output renderer/writer

However, some applications may contain in addition to the above some other specific components for doing custom operations. The order in which these parts are used depends on the application. In some applications they follow a sequence, in some they are interleaved and some others they are just data dependent. Each application has got its own set of application programming interfaces (APIs) to interact with the codecs. This makes the process complex and application specific.

The codecs are the main components that do the actual encoding and decoding of the data. They adhere to the adopted standards and take in certain set of inputs and give out certain set of bit streams conforming to the governing standards.

The commonly available applications for different operating systems follow their own format to play a file. Some media applications assign a unique Id to the file based on the format of the file and then based on the Id do further processing and rendering. Some media applications pass the header information to all the codec header analyzers and based on some factor determine the format and do the other processing.

V. The Proposed Framework

This section describes the proposed framework in detail. The technique proposed here aims to make the codec library independent of the application. Therefore a common minimum set of APIs is proposed.
The flow of data or the structure of application for the encoding and for decoding is explained separately as follows.

A. Decoding

Decoding is the process of obtaining the original raw samples from the encoded one. The input to the application will be the encoded data either from a static file residing on some memory or from a network stream. This encoded data will contain two parts as mentioned above- the stream header and the encoded multimedia data. The header will contain information that will help the decoder to decode the stream properly. The application reads this information and sends it to the decoder library. The schematic representation of the framework for the decoding process is shown in Fig.1 below.

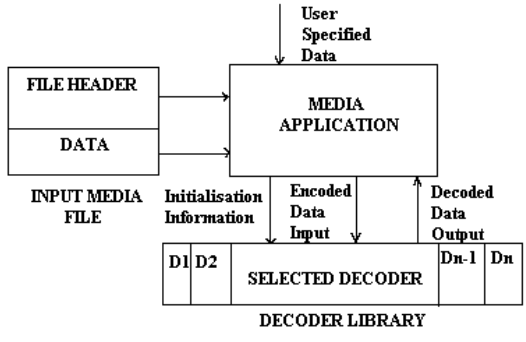


Fig. 1 Block schematic of the framework for decoding process.

The application interaction with the input file should be only in three contexts. Firstly, to determine the file format in order to load the appropriate decoder from the library. Secondly, to read the decoding information to send it to the decoder and thirdly, to read the encoded data itself.

The moment the application receives a file or a stream for processing, it scans the header. It obtains the file format from the header data. If the application supports that file format then it reads further and extracts all the header information. Some information will be sent to the decoder and some other, that is metadata will be used to display file information like title, artist, etc to the user. Some of the file information can also be user configurable like the number of channels, etc. The next phase is to load the respective decoder. Once the appropriate decoder is loaded, it needs to be initialized and used.

The application should interact with the decoder only in three contexts - to initialize the decoder, to decode the data and to de-initialize it. Initialization involves initializing the various modules of the decoder, allocating the buffers, etc and also initializing the various parameters like bit rate, sampling frequency, number of channels, etc. This information is sent to the decoder by the application from the header. Once the decoder is initialized, it is ready to decode the data. The decoder takes in encoded samples and sends out raw decoded samples. The application then has to only send the encoded data from the input file or stream to the decoder. The decoder decodes this data as per the set parameters and sends the decoded output to the application for post processing and rendering. The de-initialization involves freeing up of buffers, resetting certain parameters, etc.

B. Encoding

Encoding is the process of coding and/or compressing the original raw samples. The input file to the application will be the raw data file. Here all encoding parameters like bit rate, complexity, sampling frequency, etc are set by the user through the application. This data will be sent to the encoder. The encoder to be used is also user defined. The user can also insert some file specific metadata into the encoded file. This encoded file will contain two parts as mentioned above- the header and the data. The header will contain information that will help the decoder to decode the file properly. The schematic representation of the framework for the encoding process is shown in Fig. 2 below.

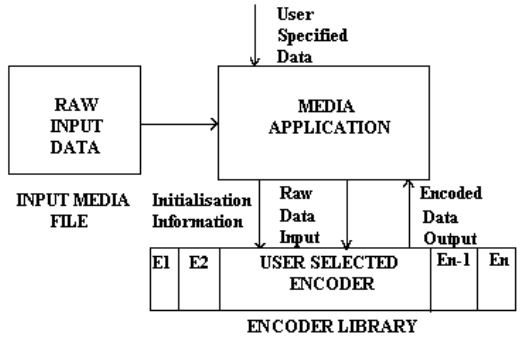


Fig. 2 Block schematic of the framework for encoding process.

The application interaction with the input file here could be only in the context of reading the input samples. The next phase is to load the respective encoder from the encoder library. Once the appropriate encoder is loaded, it needs to be initialized and used.

Like with the decoder, the application has to interface the encoder only in three contexts. First time to initialize the encoder, then to encode the data and finally to de-initialize it. Initialization involves initializing the various modules of the encoder, the buffers, etc and also initializing the various parameters like bit rate, sampling frequency, number of channels, etc. This information is sent to the encoder by the application. Once the encoder is initialized it is ready to encode the data. The encoder takes in raw samples and sends out encoded samples. The application then has to send only the raw data from the input file to the encoder. The encoder encodes this data as per the set parameters and sends the encoded output to the application which does the post processing and either transmits the data over the network or stores it on memory.

VI. Conclusions

It is clear from the above discussion that one can have a generic application independent interface for the codecs, both hardware and software. All other filtering and rendering could be done by the application either before (pre-processing) or after (post-processing) the en/decoding. These are however application specific and can be proprietary. Thus there can be a level of abstraction here. However it is better to define common interfaces for pre-processing and post processing tasks also in order to make use of the hardware accelerators provided by the devices. To a middleware developer, the application should be a black box with only the standard interfaces. The application developers too need not have to bother about designing different interfaces. Thus lot of development time can be saved; system resources like memory can be conserved, hardware utilization can be maximized thereby gaining better performance and software resources can be re-used.

References

[1] J.Bormans, J.Gelissen and A.Perkis, "MPEG-21: The 21st century multimedia framework," in Signal Processing Magazine, IEEE, vol. 20, pp. 53-62, March 2003.

[2] J.Goossen and T.Henriksson, "Media Player Software in a System of Subsystems," in Proc. Seventh Working IEEE/IFIP Conference on Software Architecture, 2008. WICSA 2008, pp. 61-70, 18-21 February, 2008.

[3] S.Sethuraman, S.Parameswaran, D.Tamia, A.Kulkarni and M. Singhal, "Multi-format media player/recorder software design methodology for programmable processors with hardware accelerators," in Proc. International Conference on Consumer Electronics, 2005. ICCE. 2005 Digest of Technical Papers, pp. 137-138, 8-12 January 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Segmentation of Printed Bangla Characters Using Structural Properties of Bangla Script

Mohammad Isbat Sakib Chowdhury, Barnali Dey and Md. Saifur Rahman

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: isbat.sakib@gmail.com, barnalidey.eee@gmail.com, saifur@eee.buet.ac.bd

Abstract – Some difficulties arise in segmentation of Bangla characters because of their unique structural properties, such as, having connectors adjoining two characters and vowel modifiers modifying either the top portion or the bottom portion of a character. This paper presents some ways to overcome these problems by taking into account these unique properties and thereby increases the accuracy of character segmentation of printed Bangla scripts that leads to greater accuracy of optical character recognition applications.

I. Introduction

Segmentation is the preliminary step of optical character recognition for any language. Unlike languages having roman alpha-numeric characters, segmentation of Bangla characters, however, is difficult due to having some unique structural properties [1, 2]. These properties make it impractical to segment characters by vertical scanning alone [3]. Currently there are four main methods of segmentation [4]:

- The first one is the "classical" approach that consists of methods that partition the input image into sub-images, which are then classified. The operation of attempting to decompose the image into classifiable units is called "dissection."
- The second class of methods avoids dissection, and segments the image either explicitly, by classification of pre-specified windows, or implicitly by classification of subsets of spatial features collected from the image as a whole.
- The third strategy is a hybrid of the first two, employing dissection together with recombination rules to define potential segments, but uses classification to select from the range of admissible segmentation possibilities offered by these sub-images.
- Finally, holistic approaches that avoid segmentation by recognizing entire character strings as units, are described.

This paper shows the adoption of second class of methods to solve the problems of segmenting Bangla characters. It is a feature based technique where the features of the Bangla script have been used.

II. Basic Features of Bangla Script

The writing style of Bangla is from left to right with 11 vowel and 39 consonant characters (Fig. 2.1). These 50 characters may be called as 'basic characters'. The concept of upper/lower case is absent in Bangla. From Fig. 2.1 it is noted that most of the characters have a horizontal line at the upper part. These characters are connected to the neighbouring characters within a word using this horizontal line, called **matra**. Out of 50 basic characters, 32 characters have matra.

অ	আ	ই	ঈ	উ	ঊ
ঋ	এ	ঐ	ও	ঔ	

(a)

ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ
ট	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন
প	ফ	ব	ভ	ম	য	র	ল	শ	ষ
স	হ	ড়	ঢ়	য়	ৎ	ং	ঃ	ঁ	

(b)

Fig. 2.1. Basic Bangla characters: (a) Vowels (b) Consonants

In Bangla script, sometimes a vowel takes a modified shape depending on its position within a word. If the first character of the word is a vowel then it retains its basic shape. Generally a vowel followed by a consonant, takes a modified shape and is placed at the left or right or both or bottom of the consonant (Fig. 2.2).

Vowel	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
Modified	া	ি	ী	ু	ূ	ৃ	ে	ৈ	ো	ৌ
ক+vowel	কা	কি	কী	কু	কূ	কৃ	কে	কৈ	কো	কৌ

Fig. 2.2. Examples of modified shapes of vowels

In Bangla script, each line can typically be divided into three strips: top, core and bottom [5].
A **line** containing two words is shown in Fig. 2.3, where these three strips are illustrated.

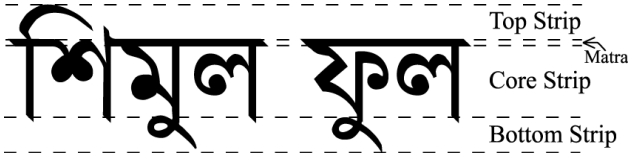


Fig. 2.3. Top, core and bottom strip of a line

The top strip and core strip are always separated by the **matra**. The top strip contains the top modifiers, and bottom strip contains bottom modifiers. In a Bangla word, top and bottom strips are present only when the word contains certain vowel or consonant characters or modifiers.

III. Overview of Segmentation Process

A. Detecting Average Line Height

When a new image is to be processed, the first step taken is the detection of average line height. This is accomplished by roughly detecting all the lines and their heights in the image without considering any special case. All such heights are summed up for all lines in the image and finally this sum is divided by the number of lines found. Thus the average line height is obtained.

B. Line Segmentation

Detecting line parameters:
For line segmentation, the whole image is scanned horizontally starting from its top left corner. When three black pixels are found in a row, the corresponding row index (**starting row index**) indicates the starting row of the first line. The scanning continues and when a completely empty row is found, the preceding row index (**ending row index**) indicates the end of the first line. Thus a line is specified by two parameters - starting row index and ending row index.

Special case:
If the top strip of a line contains only '◌ে' or '◌ৈ' and nothing else, an error occurs in detecting the ending row index of that line. There is a small gap between the **matra** and these characters in top strip. This gap is erroneously considered as the gap between two consecutive lines. Therefore, a portion of the original line containing '◌ে' or '◌ৈ' is considered as a single line. Thus two lines will be detected instead of one. To eliminate this problem, a threshold value equal to (average line height)/7 is used. When the gap between two consecutive lines is not greater than this threshold value, the aforesaid special case is considered. The line is further scanned horizontally to find the correct ending row index. In this way, the actual line is detected properly.

The special case is illustrated in Fig. 3.1.

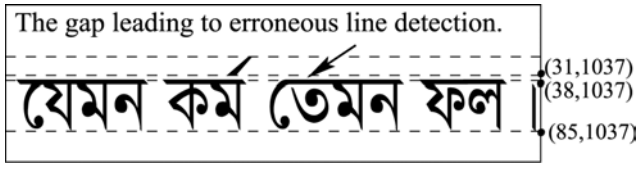


Fig. 3.1. Special case in line detection

Here, due to the gap between the '◌ে' and the **matra** of 'ম', the ending row index of the line might be mistakenly detected as 31. Thus a line will be considered here erroneously containing only the character '◌ে' and the starting row index of the next line will be found as 38. But here, 38 – 31 = 7, which is less than the threshold value of (average line height)/7 for this document. Thus 31 is not considered as the ending row index of the first line, rather scanning continues and when the ending row index 85 of the next line is found, it is considered as the correct ending row index. Thus the checking of the threshold value has eliminated the error.

C. Matra and Bottom Strip Detection

After a line is detected, the words within that line have to be detected. But before that process, the **matra** and the starting row index of the bottom strip of the line are detected.

Detecting Matra:
First, horizontal scanning is performed from the starting row of a line to its middle row over the entire column range to find the row that contains the maximum number of black pixels. Let this number be denoted as **MaxPixel** and the corresponding row be denoted as **hPosition**. Setting **hPosition** as center, the adjacent 10 rows at each side of **hPosition** are traversed, and the continuous rows are found whose number of black pixels are all greater than 0.8***MaxPixel**. The number of these continuous rows is the stroke length of this line which is marked as the **StrokeLength**. The uppermost row of these continuous rows is the starting row of the **matra** whereas the lowermost row is the ending row of the **matra** and is denoted as **EHeader**.

Detecting Bottom strip:
The bottom strip is at the lower portion of the line. For most of the Bangla characters, the lowest point lies at the beginning of the bottom strip. In fact, the starting row of the bottom strip can be said to be the horizontal line, which consists of most of the lowest points of the characters in a line. For this, each of the character is considered and the lowest point of it is obtained. The average of all the lowest points gives the starting row of the bottom strip.

At first, a vertical scanning is performed from row **r1** to row **r2**, to roughly detect the characters in the line without considering any special cases, where
r1 = **EHeader** + (ending row index of the line – **Eheader**) *2/3
r2 = ending row index of the line
Thus it is seen that only the portion lower than the middle of the line is scanned. When a black pixel is found, it is assumed to be starting column index of a character. In a similar way, a vertical scanning is performed over the same rows as mentioned above to find a blank column. When it is found, it is denoted as the ending column index of that character.

The point (i, j) of the character is checked if it is blank where i = **EHeader** + (ending row index of the line – **Eheader**) *2/3 and j = (ending column index – starting column index)/2 . If that point is blank, each point of that character is checked if it is blank beginning from the starting column index and the row corresponding to r1. When a single non-blank point is found, the next step starts to find the lowest point of this character.

Lets suppose the black pixel we found earlier is (i, j). Then (i+1, j) is checked if it is blank. If it is blank, then (i, j+1) is checked if it is blank. Again if this point is also blank, (i, j-1) is checked. If this pixel is also found to be blank, the first pixel (i, j) is the lowest point of the character. In any of these steps, if a black pixel is found, that point is denoted as (i, j) and the process continues. Thus, in this recursive manner, the lowest point of the character is obtained.

All the lowest points, i.e. the lowest row index of all the characters in the line are summed up and the result is divided by the number of characters found. Thus the average lowest point is found and this indicates the starting row of the bottom strip.

D. Word Segmentation

Starting column index: For the first word of a line, the line is scanned vertically starting from the very first column of the image i.e. the first column of the line. When three black pixels are found in a column, it is indicated as the starting column index of the first word.

Ending column index: The search of ending column index starts only after finding the starting column index. Ending column index is found by scanning vertically. A blank column should indicate the ending column index of the current word, but it is not always the case. While detecting the ending column index, an error occurs due to presence of characters that have no **matra** (e.g. এ, ঐ, ও, ঔ, ঙ, ঞ, ৎ, ং, ঃ) and in some cases, characters having half-**matra** (e.g. খ, গ, ণ, থ, দ, শ). These characters are not connected through **matra** within the word. Therefore, a small number of empty columns may be found from vertical scanning. This will cause an error in detecting the ending column index.

For example, the word ‘ এবং’ contains ‘এ’ and ‘ং’ which have no **matra**. So, the gap between ‘এ’, ‘ব’ and ‘ব’, ‘ং’ may lead to an erroneous ending column index for this word. Fig. 3.2 illustrates this.

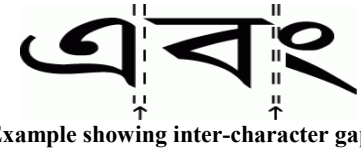


Fig. 3.2 Example showing inter-character gaps that may lead to errors in word detection

A threshold value equal to 1/8 of the line height is used to eliminate this problem. Each time a blank column is found, further vertical scanning is performed over a range equal to this threshold value. If any black pixel is obtained, it indicates that the word has not yet ended. The scanning continues. When no more black pixels are found by performing the vertical scan over the specified threshold range after finding a blank column, it indicates the correct end of the word.

E. Character Segmentation

Search for top segments:
To detect the top segments, vertical scan is performed over all the columns of the word i.e. from the starting column to the ending column of the word within the top strip. But all the rows of the top strip within the mentioned range of columns are not scanned because it is unnecessary to scan all of them. Only those rows, within the first half of the top strip are scanned.

Search for bottom segments:
To detect the bottom segments, vertical scan is performed over all the columns of the word i.e. from the starting column to the ending column of the word within the bottom strip. But all the rows of the bottom strip within the mentioned range of the columns are not scanned because it is unnecessary to scan all of them. Only those rows, within **rbottom1** and **rbottom2** are scanned, where **rbottom1** = row corresponding to the beginning of the bottom strip + **StrokeLength** and **rbottom2** = ending row of the word.

Search for core segments:
Let **r1** = row immediately after the **matra**. To detect the core segments, vertical scan is performed over all the columns of the word i.e. from the starting column to the ending column of the word and within the range of rows **rstart** to **rstop**, where **rstart** = **r1** + **StrokeLength** and **rstop** = row immediately before the starting row of the bottom strip. If two black pixels are found in a column, the corresponding column (**score1**) indicates the starting column of the 1st core segment of the word. Further scanning is performed to find the end of this segment. When a blank column is found, the column immediately before this blank column (**ecore1**) indicates the end of the 1st core segment. Thus normally a core segment is represented by two coordinates: (**r1, score1**) and (**rstop, ecore1**). But there are exceptions in this case also where a vertical scanning in search of a blank column is not sufficient to detect the end of a core segment. This is due

to the fact that the next segment starts before the preceding segment has yet ended i.e. overlapping occurs in some cases. The approach to this problem is further discussed.

Let, the starting column of the next core segment (2nd core segment) of the word = **score2**.

- If **ecore1 < score2**,
  i.e. ending column of 1st core segment < starting column of 2nd core segment, the 1st core segment can be represented by a rectangular matrix specified by two coordinates: (**r1, score1**) and (**rstop, ecore1**) as mentioned earlier. In this case, no portion of the 2nd core segment is overlapped with the rectangular matrix of the 1st core segment.

- If **ecore1 > score2**,
  i.e. ending column of 1st core segment > starting column of 2nd core segment, it indicates that a portion of the 2nd core segment is overlapped with the matrix of the 1st one. For example, in the word 'কে', the 1st core segment is 'ে' while the 2nd core segment is 'ক'. Here the 1st core segment ('ে') cannot be represented by a rectangular matrix specified by the coordinates: (**r1, score1**) and (**rstop, ecore1**) because as a portion of the 2nd core segment ('ক') remains inside that matrix, the ending column **ecore1** cannot be found by just scanning vertically (shown in Fig. 3.3). If only vertical scanning is used, a blank column from **rstart** to **rstop** will be found after 'ক' ends and the two separate segments 'ে' and 'ক' will be detected as one segment. So, there will be an error in detecting the core segments.

Fig. 3.3 Special case of character segmentation where two characters are not separable by vertical scanning

To solve this problem, when **ecore1** > **score2**
i.e. ending column of 1st core segment > starting column of 2nd core segment, the 1st core segment is represented by three coordinates. The steps involved are discussed below.

Step1. After finding the starting column index of the 1st core segment, vertical scanning is performed to find the ending column index. Now during this scanning whenever a black pixel is detected, it is checked whether the pixel's row index (**rp1**) falls within the range **rmid** to **rend**, where **rmid** = row corresponding to the middle of the core strip of the word and **rend** = ending row of the core strip – **StrokeLength**.

If the row index of the detected black pixel does not fall within this range, vertical scan continues until a blank column is found or until a black pixel is detected whose row index is within the range **rmid** to **rend.**

Step2. Let, the column index corresponding to the black pixel found in specified range = **cp1** and its row index is **rp1** as mentioned earlier. Now, a vertical scan is performed in search of a blank column over the following specified range:

Range of columns: Starting column index = **cp1** + 1 and ending column index = **cp1** + **StrokeLength**.
Range of rows: Starting row index = **rp1** and ending row index = **rstop** (i.e. ending row of the core strip).

If no blank column is found from this vertical scan, the black pixel found at (**rp1, cp1**) is not considered again and further scan is performed to detect a blank column or another black pixel in the range mentioned in Step1. Then Step2 is continued until a blank column is found. When a blank column is found either in Step1 or Step2, the column preceding this blank column indicates the end of the 1st core segment.

Let, the coordinate of the black pixel detected in Step1 within the specified range is (**rpi, cpi**) and the blank column found (within the range **rpi** to **rstop)** in Step2 has a column index = **ccore3**. Then the 1st core segment can be represented by three coordinates: (**rcore1, ccore1**), (**rcore2, ccore2**) and (**rcore3, ccore3**) where,

rcore1 = **r1** = starting row index of the core strip of the word,
ccore1 = **score1** = starting column index of this segment,
rcore2 = **rpi** = row index corresponding to the black pixel detected in Step1,
ccore2 = **cpi** = column index corresponding to the black pixel detected in Step1,
rcore3 = **rstop** = ending row index of the core strip of the word,
ccore3 = column index of the blank column found in Step2.

Now, the 1st core segment can be represented by a rectangular matrix specified by the coordinates (**rcore1, ccore1**) and (**rcore3, ccore3**). But this matrix may contain some black pixels of the next core segment due to the overlap as explained before. To eliminate these unwanted black pixels, all the pixel contents of the sub-matrix defined by (**rcore1, ccore2**) and (**rcore2, ccore3**) are made white. Thus finally the segment 'ে' is represented by a matrix specified by the coordinates (**rcore1, ccore1**) and (**rcore3, ccore3**) that contains no overlapping parts of the next segment. This is shown in Fig. 3.4.

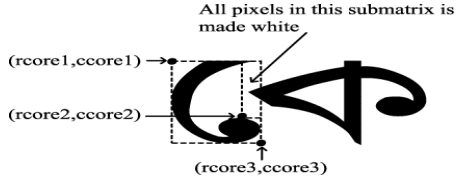


Fig. 3.4 Process to solve the problem of inseparable characters by the use of three coordinates

Now when a segment such as '  ' is detected, the detection of the next segment also requires special techniques. For example, in the word '  ', the 1st segment is specified by three coordinates (**rcore1, ccore1**), (**rcore2, ccore2**) and (**rcore3, ccore3**) as shown in Fig. 3.4. To find the starting column index of '  ', a vertical scan is performed over the column **ccore3** + 1 from the row **rstart** to **rstop.** If this column is blank, it indicates that there is no overlapping part. So vertical scanning continues from **ccore3** + 2 to the end column of the word to find the starting column index of the next segment. But if that column is not blank, it indicates there is an overlapping part as in the word ''. So, now vertical scanning again starts from **ccore2** to **ccore2** + 2***StrokeLength** and the range of rows is from **rstart** to **rcore2** – **StrokeLength**. If a black pixel is detected, it indicates the starting column index of the next segment, say, **ccore4.** The ending column index is found by the techniques discussed earlier. Let's suppose the ending column index is **ccore5.** Then the segment '' is specified by three coordinates (**rcore1, ccore4**), (**rcore2, ccore3**) and (**rcore3, ccore5**). Thus, this segment can also be represented by a rectangular matrix specified by (**rcore1, ccore4**) and (**rcore3, ccore5**) and in the same way as '', all the black pixels in the sub-matrix specified by (**rcore2, ccore4**) and (**rcore3, ccore3**) have to be made white. It is illustrated in Fig. 3.5.

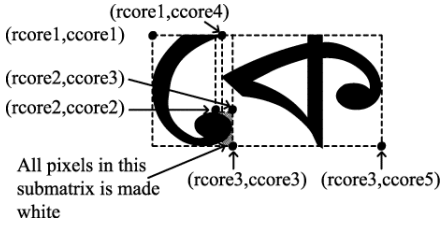


Fig. 3.5 Continuation of the special case of three coordinates representation to the next character

When the characters '', '', '' are considered for segmentation, a special technique has to be applied to obtain the correct ending column index. Since vertical scan begins after the starting row of the core strip by an amount equal to the stroke length (**rstart = r1 + StrokeLength**), a blank column is found over this range of rows before the correct ending of the character, actually around the middle part. So, the character is segmented into two parts, i.e. detected as two separate segments. To detect these two segments as a single character, the following technique is applied.

Let, a blank column (**cb**) is found over the specified range of rows which will indicate a wrong ending column index. Now a horizontal scan is performed from the beginning of the core strip (**c1**) up to the blank column just found (**cb**) and the range of rows is from **r1** + (**rstop – r1**)*3/4 to **rstop**. If a specific range of rows is found to be blank, again vertical scan is performed to obtain the next segment which is actually the remaining part of the character. When the next segment is found, the two segments are considered to be two parts of the same character. To represent these two parts together as a single character, two coordinates (**rcore1, ccore1**) and (**rcore2, ccore2**) are used.

For example, the segmentation of the character '' is illustrated in Fig. 3.6.

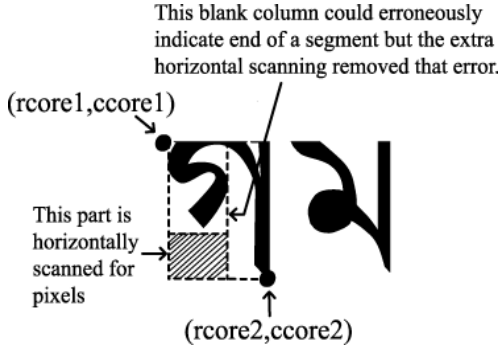


Fig. 3.6 Example showing the case where a single character is segmented into two parts and the process to solve it

IV. Conclusion

The segmentation techniques described here overcome most of the problems incurred while performing character segmentation of printed Bangla characters. The techniques are also fast because they avoid unnecessary scanning. Since character segmentation is the primary step of optical character recognition applications, these techniques while employed in such applications can greatly increase the recognition accuracy of printed Bangla scripts.

References

[1] Veena Bansal and R. M. K. Sinha, "A Devanagari OCR and A Brief Overview of OCR Research for Indian scripts", Proceedings of STRANS '01, IIT Kanpur, 2001.

[2] B. Choudhuri, U. Pal and M. Mitra, "Automatic Recognition of Printed Oriya Script", in Proc. of Sixth Int. Conf. on Document Analysis and Recognition, pp.795-799, 2001.

[3] B. Choudhuri and U. Pal, "A Complete Printed Bangla OCR System", Pattern Recognition, vol. 31, no. 5, pp. 531-549, 1998.

[4] R.G. Casey and E. Lecolinet, "Strategies in character segmentation: a survey", in Third International Conference on Document Analysis and Recognition (ICDAR '95), pp. 1028, Volume 2, 1995.

[5] Ahmed Asif Chowdhury, Ejaj Ahmed, Shameem Ahmed, Shohrab Hossain and Chowdhury Mofizur Rahman, "Optical Character Recognition of Bangla Characters using neural network: A better approach", in 2nd International Conference on Electrical Engineering (ICEE 2002.), Khulna, Bangladesh.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Bangla Numeral Recognition Engine (BNRE)

Mohammed Moshiul Hoque[1], Md. Rezaul Karim[1], Md. Gahangir Hossain[1], Md. Shamsul Arefin[1], Md. Monjur-Ul-Hasan[1]
[1]Department of Computer Science & Engineering, Chittagong University of Engineering & Technology (CUET)
e-mail: moshiul_240@cuet.ac.bd, jony_005@yahoo.com, ghcsecuet@yahoo.com, sarefin@cuet.ac.bd, mhrasel@gmail.com

Abstract: Numeral recognition is the process to classify the given character according to the predefined character class. This paper proposed a methodology for recognizing Bangla handwritten numerals which are based on fuzzy logic theory due to its low computational requirement. Every numeral is segmented and several features are extracted for each segment. In this paper, we use unique fuzzy rule base for each numeral. We have tested our engine for Bangla numerals considering various writing style and got more than 80% recognition accuracy.

I. Introduction

On-line numeral recognition involves the real time recognition of numeral as the writes them. Process of numeral recognition takes place within very short period of time. In fuzzy logic approaches, some global and local or geometric features are used. There are a lot of works on Bangla handwritten numeral recognition techniques in the world at present. One of these works has been presented in [1] where online Bangla numeric characters are recognized by automatically fuzzy linguistic rules. Bangla letters are recognized by self organizing mapping method has been presented in [2], other technique has been presented in [3] where online Bangla alphabetic characters are recognized by extraction of meaningful fuzzy rules. Automatic generation of fuzzy rule base for online handwriting recognition process has been presented in [4]. Handwriting recognition is a challenge with On-line character recognition is the development of a system that can recognize these characters in real-time. So for humans the possibility to communicate with the computer via handwriting is a tremendous enhancement of the man-machine interface. With the increasing the interest of computer applications, modern society needs the handwritten text into computable readable form. Therefore, handwriting recognition is a very interesting input method. The main objective of this paper is to recognize the Bengali numeral using fuzzy logic considering various writing styles of different users. For recognition, we have to use fuzzy rule-base for each character and extract features. The main advantages of fuzzy logic approaches are that it requires small amount of memory space and accuracy is very high.

II. Bangla Numeral Recognition Engine (BNRE)

The core of the recognition engine is the knowledgebase that is in the form of fuzzy rules. The outline of the BNRE can be depicted as Fig. 1. The input of a new numeral has to pass following processing steps: segmentation, fuzzy features extraction, learning, and recognition. Input data is segmented according to angle difference. The numbers of features can be categorized into global, geometric and position features in the fuzzy features extraction step. Fuzzy features are distorted into database and stored into rule base in the next steps.

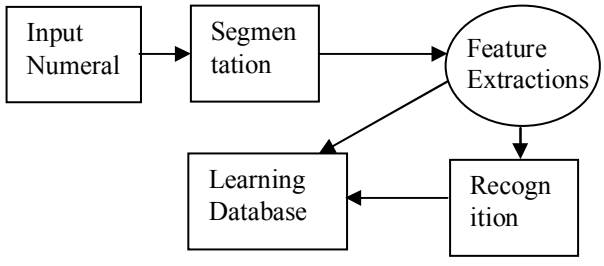


Fig. 1 Architecture of the BNRE

III. Segmentation

Each numeral divided into a several segments. The segmentation is based on the movement and angle differences between first four point connected line and consecutive second four point connected line. A segmented Bangla numeral seven is depicted in Fig. 2.

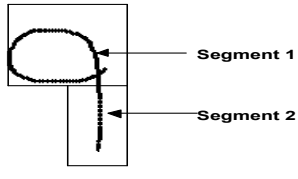


Fig. 2: Segmented Bangla numeral seven.

If the angle difference is more that 90^0 then recognized a new segment. When electronic pen or mouse is up recognized also a new segment. Segmentation steps are described in the following:

Step 1: Initialize a new segment when pen down occur. Set coordinate = current coordinate;
Step 2: Store Coordinate and increase no_of_point;
Step 3: Set Coordinate = next coordinate and go to step 2 until pen up or abrupt change in direction with no_of_point >4;
Step 4: If pen up store coordinate and set coordinate to 0;
Step 5: Go to step 1;
Step 6: If abrupt change in direction with no_of_point>4 set no_of_point to 0;
Step 7: Initialize new segment;

Step 8: Go to step 2.

IV. Features Extraction

Fuzzy features play an important role in the character recognition. Features are divided into three categories: global features, positional features and geometric features.

A. Global Features

Global features are common to all types' features. Global features belong to whole character. Some global features are given in Equations (1)-(4).

$$\min X^{seg(n)} = Min(x_i) \quad (1)$$

$$\max X^{seg(n)} = Max(x_i) \quad (2)$$

$$\min Y^{seg(n)} = Min(y_i) \quad (3)$$

$$\max Y^{seg(n)} = Max(y_i) \quad (4)$$

The coordinate of the center points of each segments are calculated by the Equations (5)-(6).

$$centreX^{seg(n)} = (\min X^{seg(n)} + \max X^{seg(n)})/2 \quad (5)$$

$$centreY^{seg(n)} = (\min Y^{seg(n)} + \max Y^{seg(n)})/2 \quad (6)$$

These global features are necessary for calculating the both positional features and geometric features.

B. Positional Features

The positional feature determines the relative position of identified segment with respect to universe of discourse. If the individual segments have been identified, the next step is the determination of the center of the identified segment. The universe of discourse is divided into two linguistic variables which are *Vertical position (VP)* and *Horizontal Position (HP)*. The *Vertical Position* is divided into eight linguistic terms: {Nearly Top (NT), Top Centre (TC), Top (T), Middle (M), Bottom Centre (BC), Bottom (B), Nearly Bottom (NB), Centre (C)} and the *Horizontal position* is divided into seven linguistic terms: {Left (L), left corner (LC), nearly left (NL), Center (C), Right (R), Nearly Right (NR), Right Centre (RC)}. Then the relative position of the identified segment is expressed by the Equations (7)-(8).

$$\mu_{HP} = \frac{\left(centreX^{seg(n)} - \min X^{seg(n)}\right)}{\left(\max Xseg(n) - \min X^{seg(n)}\right)} \quad (7)$$

$$\mu_{VP} = \frac{\left(centreY^{seg(n)} - \min Y^{seg(n)}\right)}{\left(\max Y^{seg(n)} - \min Y^{seg(n)}\right)} \quad (8)$$

Then this membership value is compared with predefined linguistic terms for horizontal and vertical position.

C. Geometric Features

Geometric features are extracted per segment and divided into two main categories which are straight line and arc. The fuzzy values involved these features are *arcness* and *straightness* [4].

C.1. Straightness Determination

The straightness ($\mu_{Straightness}$) of a given segment is calculated by fitting a straight line with minimum least squares error. The membership function for straightness of the every segment are calculated by the Equation (9).

$$\mu_{Straightness} = \frac{\left[D_{P(0)P(N)}\right]}{\sum_{K=1}^{N} D_{P(K)P(K+1)}} \quad (9)$$

Where $D_{P(K)P(K+1)}$, is the straight-line distance between point K and point $(K+1)$ on the nth segment. The number of element in the segment is depicted by N. If $(\mu_{Straightness})$ is greater than 0.6 then the segment is straight line and otherwise it is an arc.

C.2. Arcness Determination

The ratio of the distance between end-points and total arc length determined the arcness ($\mu_{Arcness}$) of a particular segment. The membership function for arcness is given according to Equation (10).

$$\mu_{Arcness} = 1 - \mu_{Straightness} \quad (10)$$

If given segment is determined to be an arc then categorized the arc into one of the five types such as: A-shape (A), U-shape (U), C-shape (C), D-shape (D) and O-shape (O). These categories are distinguish by using the angle of rotation, angle of slope joining end points, the measure of arcness, relative length and area covered by the segment [5]. Different types of arc can be defined according to the Equations (11)-(15).

The equation of the A-like curve is

$$\mu_A = \min(1, \sum I_y / n), \quad \text{where } I_y = 1 \text{ if } y > (y_S + y_e)/2 \text{ ; Otherwise 0} \quad (11)$$

The equation of the U-like curve is

$$\mu_U = \min(1, \sum I_y / n), \quad \text{where } I_y = 1 \text{ if } y < (y_S + y_e)/2 \text{ ; Otherwise 0} \quad (12)$$

The equation of the C-like curve is

$$\mu_C = \min(1, \sum I_x / n), \text{ where } I_x = 1 \text{ if } x < (x_S + x_e)/2 \text{ ; Otherwise 0} \quad (13)$$

The equation of the D-like curve is

$$\mu_D = \min(1, \sum I_x / n), \text{ where } I_x = 1 \text{ if } x > (x_s + x_e)/2 \text{ ; Otherwise 0} \quad (14)$$

The equation of the O-like curve is

$$\mu_O = \left(\sum_{K=1}^{N} D_{P(K)P(K+1)}\right)/(2 * 3.1416 * r) \quad (15)$$

Where r is the radius of the curve, x_S, x_e is the start point and end point of segment respectively in the X-axis and y_S, y_e is the start point and end point of segment respectively in the Y-axis. The other types of O-shape arc and corresponding equation and picture are given in Table 1.

Table 1: Different types of O-Like curve

Shape	Name	Function
	O-shape Top	$\mu_{OT} = \min(\mu_O, \mu_U)$
	O-shape Bottom	$\mu_{OB} = \min(\mu_O, \mu_A)$
	O-shape Left	$\mu_{OL} = \min(\mu_O, \mu_D)$
	O-shape Right	$\mu_{OR} = \min(\mu_O, \mu_C)$

After calculating all fuzzy features, the next step is to mapping these values to linguistic term according to predefined range of Table 2.

Table 2: Linguistic terms and corresponding range

Linguistic terms	Meaning	Range
Z	Zero	≤ 0
VVL	Very Very Low	> 0 ~≤ 0.12
VL	Very Low	> 0.12 ~≤ 0.24
L	Low	> 0.24 ~≤ 0.36
M	Medium	> 0.36 ~≤ 0.48
H	High	> 0.48 ~≤ 0.60
VH	Very High	> 0.60 ~≤ 0.72
VVH	Very Very High	> 0.72 ~≤ 0.84
E	Excellent	> 0.84

V. Fuzzy Rule-Base Generation

The rule base is depends on the number of segment contained in the given numeral. If the numeral contains n segments then the number of rule base are n!.

A. Algorithm for Rule Generation

For each segment different variation is collected for the character. Variation of a segment means which has the same segment serial but different position in the universe of discourse. For each variation mean the important geometric feature are calculated using important features analysis. For finding the important global feature same segment number is used as the criteria. The details of rule generation algorithm have been described in [1]. Fuzzy rule is unique for every character. The fuzzy rule is built up according to important features and a small fragment of fuzzy rules for Bengali numerals 0, 1 and 5 are presented in Table 3.

Table 3: Fuzzy Rules for numerals ০, ১ and ৫.

Numerals	Segment No.	Fuzzy Rules
০	1	If ((seg.(1).μ_{ARC} =VVH) OR ((seg.(1).μ_{ARC} =E)) AND ((seg.(1).μ_{STR} =VVL) OR (seg.(1).μ_{STR} =Z)) THEN Output=20;
১	2	If (seg.(1).μ_{ARC} =M) AND (seg.(1).μ_{VP} =BC) AND (seg.(1).μ_{DL} =VH) AND (seg.(1).μ_{CL} =VL) AND (seg.(2).μ_{ARC} =VH) AND ((seg.(2). μ_{OL} =E) OR(seg.(2). μ_{OL} =VVH)) THEN Output=21;
৫	5	If (seg.(1).μ_{STR} =VVH) AND (seg.(1).μ_{HP} =L) AND (seg.(1).μ_{HL} =VVH) ((seg.(2).μ_{STR} =H) OR (seg.(2).μ_{STR} =VH)) AND (seg.(2).μ_{VP} =LC) AND (seg.(2).μ_{PS} =M) AND(seg.(2).μ_{SLEN} =VH) AND ((seg.(3).μ_{ARC} =M) OR (seg.(3).μ_{DL} =Z)) AND (seg.(3).μ_{CL} =VVH) AND (seg.(3).μ_{HP} =NB) AND (seg.(4).μ_{ARC} =VH) AND (seg.(4).μ_{CL} =M) AND (seg.(4).μ_{DL} =VL) AND (seg.(5).μ_{ARC} =H) AND (seg.(6).μ_{UL} =VVH) AND ((seg.(5).μ_{AL} =VVL) OR (seg.(5).μ_{AL} =Z)) THEN Output=25;

VI. Learning Mode

In learning mode, user draws the numeral in the drawing pad. The overall learning process describes in the following:

Step 1: Numeral segmentation.

Step 2: Feature extraction. The important features extraction is described by the tabular form in the following.

Step 3: Data store. Features stored mechanism into databases in the following manner.

```
Set rs = New ADOBD. Recordset
If segmentno = 1 Then
Rs.Open "INSERT INTO seg1table
ElseIf segmentno = 2 Then
Rs. Open "INSERT INTO seg2table
ElseIF segmentno = 3 Then
Rs. Open "INSERT INTO seg3table
Else
Rs.Open "INSERT INTO others
```

VII. Recognition Mode

The overall recognition process can be described in the following way:

Step 1: Recognition mode segmentation is same as the learning mode segmentation.

Step 2: Recognition mode feature extraction is same as the learning mode extraction.

Step 3: Store mechanism of data into array are described according to following the code.

```
Dim SL AS INTEGER
SL = 1
If MVL<= 0 Then
    Ch1 (SL, 5) = "Z"
    ElseIf MVL > 0 and MVL <=0.125 Then
    Ch1 (Sl, 5) = "VVL"
    ElseIf MVL > 0.125 and MVL <=0.25 Then
    Ch1 (SL, 5) = "VL"
```

Step 4: In comparison stage, array data are compared with the Rule base data.

Step 5: Highest value is calculated from step 4 as percentage.

Step 6: Recognized the numeral.

VIII. Experimental Results

In experimentation, we have 10 Bangla numerals (zero to nine) each consisting of five different samples of each numeral and collected from hundreds different users. We have tested our engine for total of five thousands different sample numerals for Bangla numeral segmentation and to evaluate the different types of features that are necessary for recognizing the numerals.

A. Segmentation Results

The segmentation result of Bangla numeral 1 and 5 are presented in Table 4.

Table 4: Segmentation results for numeric ১ and ৫.

Numerals	Segment Number					
	1	2	3	4	5	6
১	N	N	Y	Y	N	N
৫	N	N	Y	Y	Y	N

In Table 4, Y means 'yes' and N means 'no'. Every numeral can have many segments and it depends on writing style of different users.

B. Feature Extraction Results

Several features are extracted for each character according to segmentation basis. Number of Segment and features can be varied with numeral writing styles. Table (5)-(7) represents the positional features (HP: horizontal and VP: vertical), straightness features (ST: straightness, VL: vertical line, HL: horizontal line, PS: positive slanted, NS: negative slanted, HLEN: horizontal length, VLEN: vertical length, and SLEN: slanted length) and arcness features (ARC: arcness, CL: C-like, DL: D-like, AL: A-like, UL: U-like, OL: O-like, OLL: O-like left, OLR: O-like right, OLB: O-like bottom and OLT: O-like top.) for numeral one (১) respectively. Table 8 shows the recognition percentages for Bangla numeral ৫ (five) with two different writing styles and Table 9 portrays the overall recognition accuracy for ten Bangla numerals each consisting of five different samples which are collected from different users.

Table 5: Positional feature for numeral ১ (one).

Numeral	Segment serial	μ_{HP}	μ_{VP}
১	1	L	BC
১	2	LC	C

Table 6: Straightness feature for numeral ১ (one).

Segment	μ_{ST}	μ_{VL}	μ_{HL}	μ_{PS}	μ_{NS}	μ_{HLEN}	μ_{VLEN}	μ_{SLEN}
1	Z	-	-	-	-	-	-	-
2	VL	-	-	-	-	-	-	-

Table 7: Arcness feature for numeral ১ (one).

Segment	μ_{ARC}	μ_{CL}	μ_{DL}	μ_{AL}	μ_{UL}	μ_{OL}	μ_{OLL}	μ_{OLR}	μ_{OLB}	μ_{OLT}
1	M	VL	VH	L	VL	VH	H	VL	Z	M
2	VH	L	Z	VH	H	E	M	M	L	L

Table 8: Recognition result for numeral ৫ (five)

Numeral	No. of segment	First recognized		Second recognized		Original
	3	5	74.11%	6	13.89%	5
	4	5	61.78%	0	39.5%	5

Table 9: Performance evaluation

Number of numeral	Number of Samples/numeral	Total number of numerals	Recognition accuracy
10	500	5000	82.099%

IX. Conclusion

Handwritten numeral analysis involves the computational identification of numerals and other written in handwritten script. The goal of this work is to develop an idea and to make module to recognized online handwritten Bangla numerals regardless of various writing style using fuzzy logic because of low computational requirements and ease of implementation. We have used fuzzy logic for features extraction. Fuzzy features are mapped into some predefined linguistic variables. All the features value is stored into temporary database, then central database. For recognizing a numeral, fuzzy rule are calculated, this fuzzy rule is compared with rule base and the character that contains highest percentage value is recognized character. The recognition rate of the proposed system is about more that 80% and speed are very well. The accuracy of the proposed system depends on the writing styles. A comprehensive approach of neural network and fuzzy logic may help to achieve the higher accuracy. This work may be accomplished in computer vision research for example, tracking the vehicle in a road by recognizing its number plate in its back side or font side. This will be helpful for road traffic management, to count number of particular vehicle and their types.

References

[1] M. M. Islam, S. M. F. Rahman, M. M. Hoque, "Online Bengali Handwritten Recognition with Automatically Generated Fuzzy Linguistics Rules", Proc. of International Conference on Computer and Information Technology (ICCIT), pp. Bangladesh, 2005

[2] M. Badrudoza, "Recognition of Bengali Handwritten Letters Using self- Organizing Maps (SOM)", Proc. of International Conference on Computer and Information Technology (ICCIT), pp. 355-359, Bangladesh, 2003.

[3] S. M F. Rahman, M. M. Islam, M. M. Hoque, "Extraction of Meaningful Fuzzy Features for Bangla Online Handwritten", Proc. of International Conference on Computer and Information Technology (ICCIT), Bangladesh, 2005

[4] A. Malaviya, L. Peters, "Extracting meaningful handwriting features with fuzzy aggregation method," Proc. of International Conference on Document Analysis and Recognition, pp. 841-844, Canada, 1995.

[5] A. Malaviya, H. Surmann, L. Peters, "Automatic generation of fuzzy rule base for online handwriting recognition," Proc. of EUFIT, pp. 1060-1065, 1994.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Application of Artificial Neural Network in Social Computing in the Context of Third World Countries

Md. Shamsuzzoha Bayzid, Anindya Iqbal, Chowdhury Sayeed Hyder and Mohammad Tanvir Irfan

Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
E-mail: shams.bayzid@gmail.com, anindya_iqbal@yahoo.com, deeyas017@yahoo.com, mtirfan@cse.buet.ac.bd

Abstract - **In the last decade, applications associated with artificial neural network (ANN) has been gaining popularity in both the academic research and practitioner's sectors. But unfortunately in underdeveloped countries this versatile tool has not yet been used. Here we consider some momentous sectors and explore the applicability of ANN in the context of third world countries. Here we explore the design of feed forward neural network for (1) assisting micro credit institutions to select appropriate locations to set up branches and (2) determining HIV risk of a locality. The simulation procedure and results are discussed accordingly.**

I. Introduction

Social problems vary form country to country. So researchers in the field of social science approach their native problems in their own ways. Huge amount of data analysis is involved in many problems of such kind. Finding different patterns from this data, inexact or hierarchical matching of patterns and making future predictions for intelligent decision making are the real challenges. For analysis of huge amount of data, application of computational tools like Artificial Neural Network, Association rules, Decision trees, Cluster Analysis are widely used.

The application of artificial neural network has been studied extensively in various sectors. Application in the field of finance is one of the most important sectors for the researchers and practitioners purpose for the last few decades. Many neural network applications are related to financial decision making. Hawley, Johnson and Raina [1] and Refenes [2] provide an overview of neural network models used in the field of finance and investment. Different forecasting about financial activities like currency exchange rate, bank failures, stock indices, bankruptcy, credit scoring have been analyzed greatly [3, 4, 5, 6, 7]. Neural network is not practiced only in the financial field. Researchers of various sectors have made efficient use of neural network. Image processing is such a field that can be cited here [8].

As the nature of most of the social problems varies widely from country to country, use of computational tools depends on the appropriate implementation considering specific aspects of those problems. That is why, social researchers in third world countries cannot use the tools developed in developed countries where this sort of application of computational methodologies are very common. Here lies the motivation of research on the part of computer scientists in underdeveloped countries. They have to come forward to apply computational analysis to solve the social research issues. From this motivation we have chosen two problems with unique aspects in the context of third world countries. The problems are:

1. Decision making on branch setup of micro-credit organizations
2. HIV Risk determination of a locality

Here we have used feed-forward Artificial Neural Network based on multi-layer perceptron (MLP) as our computational tool. This is a well known and extensively used tool which has inherent capability to discover hidden patterns in known (training) data and make almost accurate prediction later on in case of unknown (test) data. Construction of neural network architecture and determination of other parameters are discussed in the methodology section. Feature vector selection was very carefully done with consultation of domain experts and related research papers. The difficulty lies with availability of data. For world-wide recognized problems researchers find data from repositories offered by world famous research organizations and universities. As we are dealing with exclusively our local problems, data is not available as this sort of practice (use of computational tools for analysing social problems) is not well-known. But someone has to start. We have generated artificial data and developed the model which will hopefully encourage our organizations concerned to accumulate real-world data, use our model and find real world decisions.

The rest of the paper is organized as follows. In Section II we present primary definitions and concepts of artificial neural network. In Section III we describe our problem solving approach. Section IV deals with the implementation details of microcredit branch setup and Section V deals with HIV/AIDS risk determination problems including feature vectors selection. Finally we conclude in Section VI.

II. Preliminaries

In this section we discuss some definitions and related topics.

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network

function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are used, in this supervised learning, to train a network.

Backpropagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by the user. Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function with a finite number of discontinuities.

Standard backpropagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule, in which the network weights are moved along the negative of the gradient of the performance function. The term backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks. There are a number of variations on the basic algorithm that are based on other standard optimization techniques, such as conjugate gradient and Newton methods.

Properly trained backpropagation networks tend to give reasonable answers when presented with inputs that they have never seen. Typically, a new input leads to an output similar to the correct output for input vectors used in training that are similar to the new input being presented. This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs.

Backpropagation training functions are to train feedforward neural networks to solve specific problems. There are generally four steps in the training process: assemble the training data, create the network object, train the network, and simulate the network response to new inputs.

III. Methodology

In this section we discuss the methodology we use to solve the problems.

A. Feature vector selection and interpretation of output

Consultation of domain experts and study of relevant research work help us deciding feature vectors, that is important attributes of the problem which are likely to influence a decision. Some of these selected feature vectors may be directly expressed in numeric form. Most of them will be converted to a numeric value that significantly expresses its relative position. For a simple example, excellent may be represented by 1, very good by 0.8, good by 0.6, moderate by 0.5, bad by 0.3 and so on. Similarly a bit more categorical representations like 1 for very rich, .6 for solvent, etc.

The outcome of known data or later on predicted outcome also needs to be interpreted in numeric form. In case of micro-credit branch setup the choice of a place will be highly recommended if setting up the branch raises average income of target income-group people. This is the criterion of choosing a place among many alternatives. The higher raise is predicted, the higher the possibility of choosing this place. For a problem to predict risk of HIV infection in a locality may also be represented in numeric form, i.e., 0.8 to 1 will denote high risk; below 0.4 will denote low risk.

B. ANN architecture and necessary parameters determination

Construction of input layer depends on number of selected feature vectors. Empirical study is made to define hidden layer architecture which gives accurate result on training and validation data and also will work well on huge amount of data in consideration of convergence time. The desired output format defines output layer architecture.

C. Training, test, and validation process

The neural network model used in this study is feedforward neural network based on multi layer perceptron (MLP). MLP has been studied extensively by the researchers for a long time and has been accepted world wide. It is easy and efficient. These factors influenced us to adopt this method.

The design of ANN proceeds through several phases. First, the inputs are analyzed for discovering important attributes of the domain that are used as feature vectors. Then suitable network architecture is chosen and a significant training data set is used to train the network. After the network is being trained, a significant data set called validation data set is simulated with the network to check whether the network is worthy or not.

After selecting the relevant feature vectors our job was to select appropriate network architecture to optimize our objectives. The approach used in these implementations in order to construct neural network is totally empirical, and is not based on theoretical evidence. Indeed, the design of an MLP is by no means an exact science and no complete theoric explanation exists to obtain the optimal architecture of the MLP. For these reasons we had to rely on simulating different architecture and picking the best one. We search over various network architectures involving different number of layers and neurons and selected the configuration that incurred the least amount

of mean square error with an eye on the training time of the network.

Our final task is to train and validate the network with relevant data set. There is no such repository in third world countries and the institutions do not seem to be eager to publish relevant data for lack of infrastructural support and technical orientation. So we randomly generated feature vectors for training, validation and testing purpose.

We generate data by a computer program written in C language to train and validate the network. We fed the training data to train the network. Then we simulate the trained network with the validation data and check whether the accuracy of the network can achieve a significant level of accuracy. In case of failure, we train the network again with modified number of neurons. We go through this way recursively until the trained network can achieve the preset accuracy level. Then our network is fit for assisting the decision making of microcredit branch set up in a specified locality. Fig.1 demonstrates the whole process.

IV. Implementation

In this section we find out the feature vectors that are appropriate in the context of third world countries and give experimental results. First, we focus on the problem of microcredit branch setup problem.

A. Micro credit Branch Setup

In this section we reveal the most important attributes that can bias the decision of branch setup of microcredit institutions and suggest an appropriate feedforward neural network.

In recent years, the development community came to view microcredit as an increasingly important tool for poverty alleviation and economic empowerment. Microcredit institutions have been grown rapidly throughout the world. This huge success of microcredit programs often tend to overshadow the occasional failures of microcredit programs and these failures are now becoming too frequent to overlook. So an objective study of the suitability of microcredit program for a specific locality has attracted quite a lot of attraction among the researchers. And there has been dazzling revelations too. Our goal is to choose a place from alternative options to set up a new branch of a micro-credit organization which will maximize the goal of having positive impact on the income of target people. At the same time other inevitable issues like the risk of failure to collect repayment have to be considered. Hence, an objective study of the suitability of microcredit program for a specific locality has considerable impact in future in terms of repayment and actual effect on poverty level. Based on comprehensive consideration of the factors we have decided the feature vectors as described below:

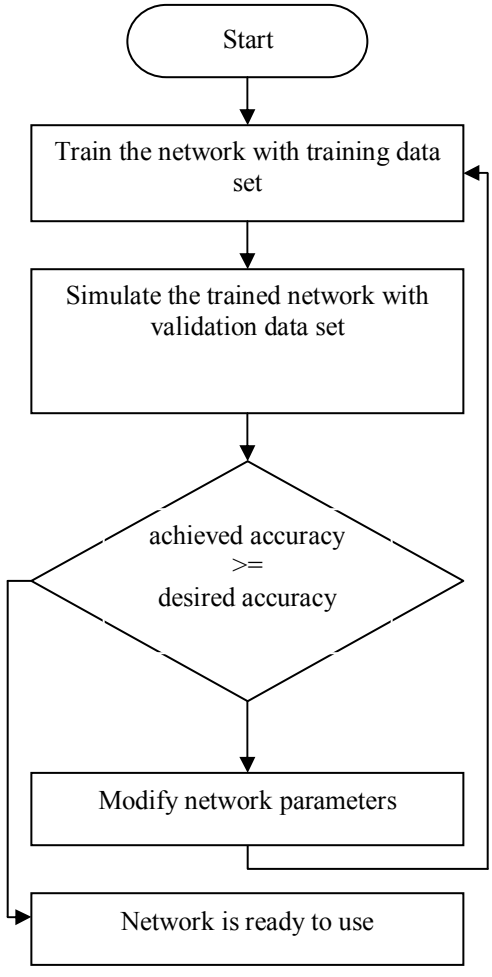


Fig. 1 Flowchart of the network building process.

- **Target population:**

There is no point in setting up a microcredit branch in the middle of desert for the simple reason that there will be no one to give credit to. So the target population is an important feature and indeed many microcredit institutions specify a minimum population to start a branch in their regulations. For example, 'Pride' (a Tanzanian company) specifies this minimum limit as 100000.

- **Average years of schooling:**

We have considered the average years of schooling of both the target population and the heads of families. The first one is obvious; certainly someone with a significant amount of schooling will know what to do with the money. A research on microcredit programs in Bangladesh shows that another year of schooling increases non agricultural assets by around 45% among microcredit customers [9]. The need of schooling of family is not obvious. But again research shows that this also has a positive impact. Additional year of schooling of household head increases asset accumulation by 7% [9].

- **Contribution of enterpreneurial activities towards GDP:**

Experience shows that most of the microcredit induced investments are in manufacturing of skill based economic activity. A recent work in this regard states, "Societies with little debt problem but strong enterpreneurial culture may thus be a better environment for offering untied microcredit than societies where entrepreneurial activities need decisive encouragement" [10]. Indeed, a region where enterpreneurship is common as opposed to predominantly agri-societies microcredit can fuel development more effectively.

- **Infrastructure:**

It may seem very selfish that microcredit programs are often commercial organizations making profit. At least they have to reach the break even point. So they want to have good communication link, electricity and other facilities. A relevant work reveals that "With the presence of electricity, asset accumulation increases by almost 42 percent" [11]. Also the presence of other financial institutions facilitates the transfer of cash.

- **Rate of unemployment:**

Extreme poverty forces people to consume credit for personal uses.

Critics usually respond that offering credit to very poor community lead them into debt and deepen their economic dependence. In such cases credit is never recovered and the program is doomed. So an extraordinary high rate of unemployment is a discouraging factor. However the unemployment rate of women is a controversial factor and should be measured as a function of the percentage of women who are currently in economic activities or willing to do so. So a door to door survey is needed to measure this factor.

- **Par Capita Income:**

For the same reason stated in the previous point, lending in extremely poor areas can be catastrophic for the institutions. Again estimating the par capita income of women target group presents a problem and should be estimated as a function of household income.

- **Natural Calamity:**

People located in areas prone to natural calamities are tricky since creditors may fail to repay the loans as they have to face natural brutality and struggle hard to get them over. So from the perspective of loan repayment such areas are penalized in points.

- **Distance to nearest health centre:**

This feature may seem bizarre at first; at least it did it to us. But statistics shows that there is strong correlation between health services and economic performance of the creditors. A relevant research shows that among the village characteristics, presence of rural health centre considerably increases women's performance, which is not surprising as better health definitely increases productivity of work force in general. Noticeable thing is the magnitude of increase. Women on the average accumulate 94% more assets than those women who do not have any rural health centre present in their village [11].

Table 1 Demonstration of actual output and simulated output.

Feature Vectors									Actual Output	Simulated Output
f1	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*		
5	4	4	1	1	2	1	1	4	0	0
5	1	1	5	3	3	2	5	5	1	1
5	1	4	1	4	1	1	3	2	0	0
3	4	4	2	5	5	2	5	1	1	1
1	3	5	3	4	5	1	4	4	1	1
2	1	3	4	3	5	2	2	2	0	1
3	5	1	5	5	4	2	2	5	1	1
5	1	1	1	2	1	1	5	5	0	0

- **Existence of training facilities:**

Skilled population can utilize credit more effectively. So number of training centre in the locality is a direct indicator of enterpreneurship and influences setting up of the branch of microcredit institution positively.

B. Data and Result analysis

The features of an existing branch are known, although here we have assumed those for lack of real data. These are numerically represented as we have discussed in the methodology section. Table 1 demonstrates the result of a working example. Here *f1, f2, f3, ..., f9* are 9 feature vectors associated with the decision making about microcredit branch setup with 5 as the most influencing value whereas 1 is the least. Actual output indicates whether the microcredit branch should be set up or not with the value of 1 or 0 respectively. This output is generated by a computer program which manipulates some decision making heuristics. The training data set, validation data set and test data set are also generated by the same program. Now we simulate the same data by the trained feed forward neural network. And the simulated output in the table indicates the output generated from this network. Here, among 8 decision making situation, our network fails only in one situation resulting into 87.5% efficiency.

The experimental result is convincing enough to adapt in the real life. Our procedure can predict whether a branch should be set up or not with accuracy rate of around 90%. We trained the network with different simulated data set and all the time we found our network good enough to serve our purpose.

Now we proceed to our second problem .

V. HIV/AIDS Risk Determination

In this section, we reveal the motivation behind the need of HIV/AIDS risk determination for a particular locality and find out the factors that have to be considered to measure the risk from the perspective of third world countries. There is variation in different relevant factors even among the underdeveloped countries. So, we have

chosen Bangladesh, one belonging to this group, to analyze and determine the feature vectors.

HIV/AIDS has been a devastating threat to the world for the past few decades. Researchers around the world are storming their brain to control this epidemic. With prevalence rate of less than 1%, HIV/AIDS in Bangladesh [UNAIDS report] may not look like a major threat. Yet nothing can be further from truth. In a population of over 130 million, a mere 1% rise would mean an addition of more than a million to the numbers. The first case of HIV in Bangladesh was detected in 1989 and according to a 2004 UN study, HIV infections have tripled in the following six years. UNAIDS estimated that 11000 adults and children were living with HIV at the end of 2005 [UNAIDS global AIDS report, 2006]. So, definitely Bangladesh is on the brink of the epidemic and being a country where poverty, illiteracy and poor health are rife, the spread of AIDS presents a daunting challenge.

Since Bangladesh is densely populated and not much well organized, it is difficult to monitor the situation of the country as a whole. Moreover, unlike developed countries, Bangladesh lacks the scientific laboratories, research facilities, equipment, and medical personnel to deal with an AIDS epidemic. So, what we can do is to evaluate the risk of different location individually and take effective measure according to the vulnerability of a particular location. In this paper, we reveal a scientific methodology to evaluate the degree to which a particular locality is vulnerable to HIV/AIDS risk. And in this point, we unveil the tremendous capability of feed forward network to serve our purpose.

A. Feature Vectors

We consider different factors that can influence the risk determination of a region and settle down with some attributes that we think most relevant in the context of Bangladesh. In the rest of this section we elaborate each feature and describe why we think them important.

- **Access to sex workers:**

Sex work is central to this epidemic that is primarily spread by unprotected heterosexual intercourse. It is also a feature of all cultures, encompassing a wide range of people and behaviours. Sex work can involve men and transgender people, as well as women. People who are engaged in selling sex obviously have multiple sex partners and are therefore highly vulnerable to several Sexual Transmission Diseases (STDs/STI) and HIV/AIDS infection. Sex workers in Bangladesh have a higher client turn-over rate than in any other south Asian country, and consistent condom use during paid sex is rare (depending on the region, 0–12% of sex workers said that they used condoms with new clients), reports UNAIDS update of 2005. There are over 105,000 sex workers, both female and male, in the country. Brothel-based female sex workers reportedly see around 18 clients per week, while street-based and hotel-based workers see an average of 17 and 44 clients per week respectively [UNAIDS report]. So the existence of brothel as well as number of street and hotel sex workers is some factors that we should consider and evaluate properly.

- **Intravenous drug users:**

Intravenous drug users (IDUs) are open to the high risk of HIV/AIDS through the repeated use of same needle. A national survey data indicates that HIV incidence among IDUs jumped from 1.8% in 2001 to more than 4% in 2004. In one Dhaka “hotspot” the prevalence has jumped to 9%. A survey in Central Bangladesh revealed that more than 70% IDUs routinely share needles. This is comparable to levels in countries that are experiencing a concentrated and growing HIV epidemic. Illegal sale of blood by IDUs increases the threat of tainting the national blood supply. So it is not the case that only they themselves are at high risk, rather they endanger their society as well.

- **Homosexuality:**

This is the most common cause of spreading HIV in high income western country like USA. Statistics reveals that homosexuality among men (MSM) is responsible for 62% risk in USA (Fact Sheet - HIV/AIDS among Hispanics in the United States). There has been little research on the role of sex between men in our country's HIV epidemic. But still we have to be careful since sex between men is highly stigmatised in Bangladesh and is not openly talked about, making it easy for people to underestimate how commonly it occurs. So a tough and challenging survey is needed to measure this factor.

- **Migrant workers:**

Large numbers of people have moved around within Bangladesh, to neighbouring countries or overseas, in order to work. In many cases, migration does not change an individual's sexual behaviour, but leads them to take their established sexual behaviour to areas where there is a higher prevalence of HIV [12]. Long working hours, isolation from their family and movement between areas may increase the likelihood that an individual will become involved in casual sexual relationships, which in turn may increase the risk of HIV transmission. So regions with large number of migrant workers are vulnerable to the HIV/AIDS risk and hence penalized in points.

- **Number of truck drivers:**

This feature may seem strange first but in fact it has strong correlation with HIV/AIDS risk determination. Bangladesh has a large road network, involving thousands of drivers and helpers. Truck drivers spend long periods of time away from home, and it is common practice for them to have relations with sex workers while on the road. 24-34% of truck drivers of India, according to various survey results, have been reported to be engaged in sex with commercial sex workers [13]. Though there has not been such survey in Bangladesh, it can be stated without any fear of being wrong that situation is not much better in our country. A research work in this regard

states, "There is no entertainment. It is day-in-day-out driving... When they stop, they drink, dine and have sex with women. Then they transfer HIV from urban to rural settings" [14].

- **Medical malpractice:**

Bangladesh, being a poor country, is still lagging far behind from quality medical practices. Except some reputed hospitals and clinics, the proper screening of blood and restriction on repeated use of same syringe is not in practice. There has been no proper step to restrict professional blood donors who are mostly drug addicted. HIV prevalence among pregnant women attending antenatal clinics is also a major factor to be considered. Thus medical malpractice can clear the way for AIDS pandemic to explode in Bangladesh. So, we should carefully evaluate the medical standard of a region while measuring its vulnerability to HIV/AIDS risk.

- **Tourist spot:**

A lot of people from home and abroad gather in a tourist spot. Tourists often get involved in extra marital sex. And utilizing this fact, some dishonest hotel businessmen provide their clients with professional sex workers. These sex workers are open to the clients from both home and abroad. Hence, the condition of the HIV/AIDS epidemic may be aggravated in a tourist spot. And certainly a region which has some tourism values has to be penalized while measuring this feature vector.

- **Dependency of women:**

Most of the women in Bangladesh are not economically self dependent rather they depend on their husbands. This leads to a men dominant society where women are obsessed with the thought that they have to obey their husbands anyway. And unfortunately, women have no right to protest against being affected from their husbands who are in bad practice like unsafe sex with sex workers, homosexuality and drug addiction. The decision of using condom while encountering sex is entirely depends on the whims of their husbands. Although these things may seem very strange to the modern generation but the actual picture of most of the rural areas where illiteracy and poverty are rife, is no better than that.

- **Social factors:**

Here, we take the rate of literacy and religious sentiment into account while considering social condition as a feature that can contribute a lot to the HIV/AIDS vulnerability. Still in the twenty first century, high illiteracy rate is a trademark of third world countries. Ultimately unawareness along with superstition and stigma remains a problem in many areas. Many people do not know how AIDS spreads let alone how to prevent it. Awareness can restrict people from unsafe sex or at lest it can grow eagerness of using condom. It can also fuel the use of disposable syringe.

With the advent of sky culture and globalization, the young generation is motivated by the western culture that unfortunately makes the bind with the religion very loose. They vastly accept extra marital sex and many of them are sexually multilated. So we have to consider these social values very carefully.

B. Data and Result Analysis

The approach of data and result analysis for this case is exactly same as described in the Section IV (B).

VI. Conclusion

The model we have need to be applied with real data in the scenario of third world countries which may be done by people from organizations concerned by generating data set considering the features. The accuracy of the model can also be compared with other techniques like graph partitioning which we consider to be our future work. We hope that our effort will introduce the prospect and applicability of computational aids in the field of social science to solve the indigenous problems.

References

[1] D. Hawley, J. Johnson and D.Raina, "Artificial Neural Systems: A New Tool for Financial Decision Making," Financial Analysts Journal, pp. 63-72, 1990.

[2] A. P. Refenes, *Neural Networks in the Capital Markets,* Wiley, 1995.

[3] M. T. Leung, A. Chen and H. Daouk, "Forecasting Exchange Rates using General Regression Neural Networks".

[4] K. Y. Tam and M. Y. Kiang, "Managerial Applications of Neural Networks: The Case of Bank Failures," Management Science, pp. 926-947, 1992.

[5] M. T. Leung, H. Daouk and A. Chen, "Forecasting Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models".

[6] Z. Yang, "Probabilistic Neural Networks in Bankruptcy Prediction," Journal of Business Research, pp. 67-74.

[7] H. Jensen, "Using Neural Networks for Credit Scoring," Managerial Finance, pp. 15-26, 1992.

[8] A. K. M. Rahman and C. M. Rahman, "A New Approach for Compressing Images using Neural Network," in Proc. CIMCA, 2003.

[9] H. I. Latifee, "Microcredit and Poverty Reduction," in Proc. International Conference on Poverty Reduction Through Microcredit, 2003.

[10] R. Faridi, "Microcredit Programs in Bangladesh, Assessing Performance of Participation," 2004.

[11] M. J. A. Chowdhury, "Evaluating the Impact of Microcredit on Poverty in Bangladesh: A Panel Data Approach"

[12] L. Gelmon, K. Singh, "Sexual networking and HIV risk in migrant workers in India ," in Proc. International Conference of AIDS, pp. 13-18, 2006 [5].

[13] P. Chandrasekaran, G. Dallabetta, " Containing HIV/AIDS in India: The Unfinished Agenda," The Lancet Infectious Diseases, vol. 6, no. 8, pp. 508-521, 2006.

[14] A. Christensen, "Truckers Carry Dangerous Cargo ," Global Health Council, 2002.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Comparison of Fuzzy Queries on Fuzzy Database and Classical Database

A.H.M. Sajedul Hoque[1], Md. Sadek Ali[2], Md. Aktaruzzaman[3] Sujit Kumer Mondol[4], and Dr. Babul Islam[5]

Dept. of Computer Science & Engineering, Dept. Of Information & Communication Engineering, Islamic University Kushtia-Bangladesh
E-mail: (sajidiuk,sadek_ice,mazaman_iuk,sujit_iu,babul_cst)@yahoo.com

Abstract - In this paper we have designed a sample fuzzy database and we have applied both classical and fuzzy query on this fuzzy database. We have also shown that the time cost of fuzzy query on classical database (DB) is the same as the classical query on classical DB. But the time cost of the fuzzy query on sample fuzzy database has been reduced.

I. Introduction

A database-management system (DBMS) consists of a collection of interrelated data and a set of programs to access those data. The primary goal of a DBMS is to provide an environment that is both convenient and efficient for people to use in retrieving and storing information. A number of operations are performed on a DBMS. Searching is an important operation among those. A significant amount of time is needed for searching data from a DBMS. As the size of a DB increases, the searching time also increases. A number of algorithms have been developed to improve the performance of searching using query. But those algorithms have been developed only for classical DB. The traditional DBMS cannot manipulate incomplete, imprecise and vague data such as very high, about 30, etc. properly. To overcome this problem, FDBMS (Fuzzy Database Management System) has been introduced. The primary focus of fuzzy logic is on natural language, where reasoning with imprecise propositions approximates is rather typical. As the size of DB is increasing day by day, programmers are intending to reduce the time complexity to access data from a large database [1]. The performance of a query is influenced by the structure of data and size of a DB. Large database may consist of millions of data and it costs significant amount of time to find any particular record from that database. The search time may be reduced by indexing database through the B-tree algorithm [2]. We have eliminated lack of expressiveness and also reduced searching time by designing fuzzy database and using fuzzy query on it, which is the exploration of this paper.

The rest of the paper is organized as follows. In section II describes the used sample fuzzy relational database and membership functions over a fuzzy attribute. Section III describes the construction procedures of classical and fuzzy database. Section IV describes the measurements of query cost of classical and fuzzy query over classical database and also shows the cost of fuzzy query over fuzzy database. In section V shows the results and discussion. Finally, section VI concludes the paper.

II. Fuzzy Database and Fuzzy Query

The database which contains incomplete, imprecise and vague data is called fuzzy database. Fuzzy database is founded on fuzzy logic and fuzzy set. There are two feasible ways to incorporate fuzziness in DBMS [3]: One is making fuzzy queries to the classical databases and the other is adding fuzzy information to the system. Among several data models, relational data model is the most useful and powerful model [2]. In this paper, the fuzzy database based on relational data model, called fuzzy relational database, enhances the relational model by modelling imprecise in data and/or query. A fuzzy relation is represented by a fuzzy or crisp attributes. Crisp attribute have precise data, such as Tk.7980 balance and fuzzy attribute consists of imprecise data. There are four possible types of fuzzy attributes [4]: Type 1, Type 2, Type 3 and Type 4. Among these, only Type 1, which has precise data and linguistic labels over them, is explored to build fuzzy database in this paper. The schema of used relation in this paper is as follows:

Account_schema=(account_no,branch_name,balance),
where all attributes are crisp for classical database. For fuzzy database, the only fuzzy attribute is balance and others are crisp. In both cases, account_no. is a primary key. An example of a sample account relation is shown in table1

Table 1 Account Relation

account_no.	branch_name	balance
103	Dhaka	7890
104	Narial	4500
107	Bagura	8500
120	Bagura	8100
121	Kushtia	6500
108	Kushtia	2500
110	Dhaka	3250
101	Narial	15000

We use three linguistic terms over balance attribute. They are "**High**", "**Moderate**", and "**Low**". The membership functions over fuzzy attribute, **balance**, are shown in fig 1 [5] [6].

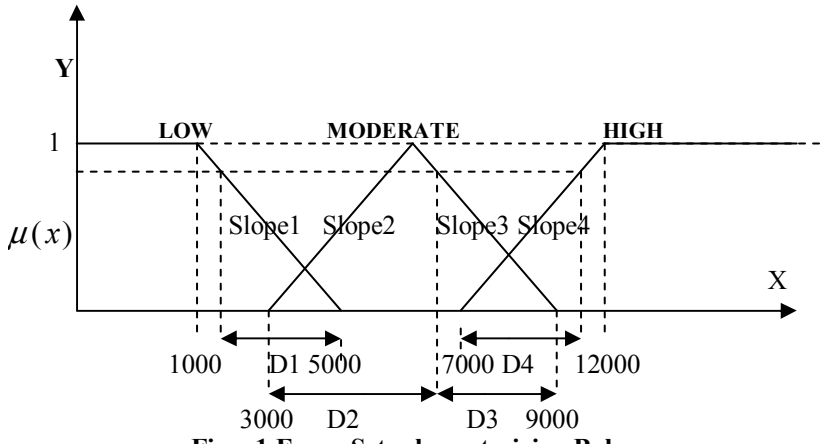


Fig. 1 Fuzzy Sets characterizing Balance.

Degree of membership of **x** for low

- Compute D1=5000-**x**
- If(D1<=0), then Degree of membership=0 (1)
  else if (x<=1000), then Degree of Membership=1
  else Degree of membership = $Slope1 * D1$

Degree of membership of **x** for moderate

- Compute D2=**x**-3000, D3=9000-**x**
- If (D2<=0) or (D3<=0), then Degree of membership=0 (2)
  else Degree of membership=$\min \begin{pmatrix} D2 * Slolpe2 \\ D3 * Slope3 \\ 1 \end{pmatrix}$

Degree of membership of **x** for high

- Compute D4=**x**-7000
- If (D4<=0), then Degree of membership=0
  else if (x>=12000), then Degree of membership=1 (3)
  else Degree of membership= $Slope4 * D4$

The query statement, used to retrieve data from database, which involves imprecise information is called fuzzy query. We only explore those queries over nonkey [2] attribute, **balance**. Then, multiple records may be retrieved through these queries.

III. Building Classical and Fuzzy Database

To build a classical database, we construct an index file based on B-tree algorithm during building a data file of desired relation [2]. The internal structure of index file of sample database after applying B-tree algorithm is shown in fig-2.

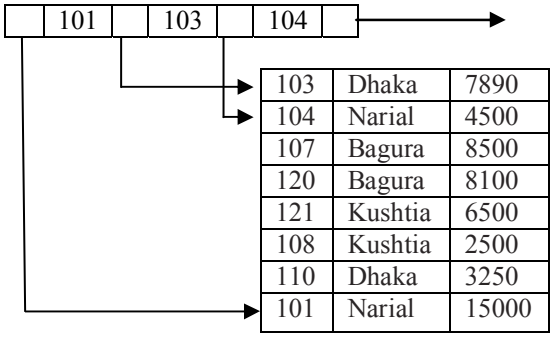


Fig. 2 The internal structure of classical database.

Here every pointer of leaf node point to only one record of account data files and the leaf node contains four pointers.

To build fuzzy database, we build an individual index file for each linguistic term of fuzzy attribute of that database. So, in this paper we build three index files, based on B-tree index structure, for each of those linguistic terms, named low index file, moderate index file, and high index file. The search-key [2] of those index files consist of primary key, account_no, with fuzzy value of that balance. These index files are built during building data file for account. To build account fuzzy database, we follow the following steps for each record:

1. Insert the desired record in account data file after testing validity of primary key of that record and set, **ad,** with the address of the current record where it is inserted in account data file.
2. Compute the fuzzy value of balance of current record for **low, moderate and high** through the formula 1, 2, 3 respectively.
3. The search-key is made by packing account_no. and fuzzy value for each linguistic term, except zero fuzzy values.
4. If fuzzy value for **high** is not zero, then
   - The search-key is inserted into **high index file**. Let it is inserted into ith position of a leaf node for **high index file**.
   - In p_i, the address of current record, **ad**, is assigned.
5. Repeat step 4 for linguistic terms moderate and low.

For example, the current record is (103, Dhaka, 7890). This record is inserted into account data file and the address of that record in account data file is taken. The membership degree for high fuzzy set is calculated as: p1= 7000, p2= 12000, D4= 7890-7000= 890, Slope1=.0002, $\mu_{High}(7890) = 890 * .0002 = .178$, which is greater than 0. So, the search-key containing account_no., 103, and fuzzy value, .178, for balance7890 is inserted into high index file. The membership degree of 7890 for moderate fuzzy set is

calculated as: p1= 3000, p2= 9000, D2= 7890-3000= 4890, D3= 9000-7890= 1110, Slope1=.0003, Slope2 =.0003, $\mu_{moderate}(7890) = \min\begin{pmatrix}1.467\\ .333\\ 1\end{pmatrix}$ =.333, which is greater than 0. So, he search-key containing account_no., 103, and fuzzy value .333 for balance 7890 is inserted into moderate index file. Similarly, for low fuzzy set: p1= 5000, p2= 3000, D1= 5000-7890 = -2890, Since D1<0, $\mu_{low}(7890)$=0. So, the search-key for balance 7890 is not inserted into low index file, and so on. The internal structure of the low, moderate and high index files for leaf node with 4 pointers are shown in figure 3, 4 and 5 respectively.

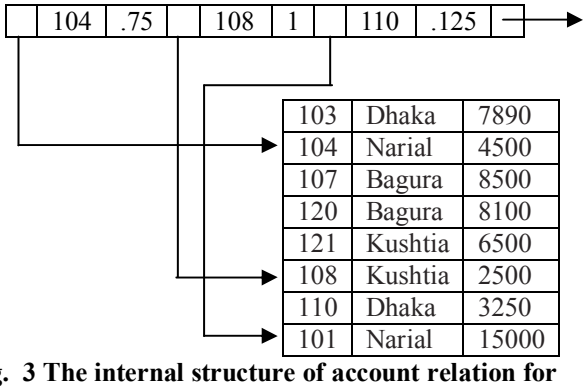


Fig. 3 The internal structure of account relation for low linguistic term

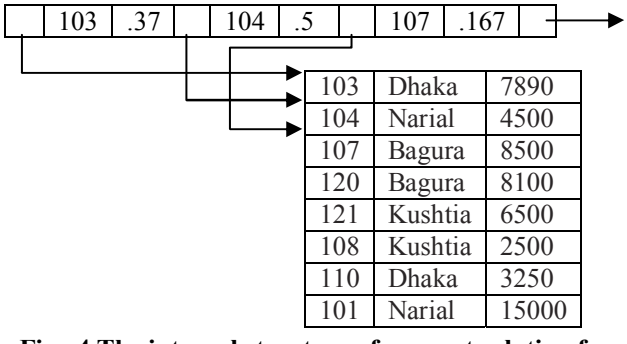


Fig. 4 The internal structure of account relation for moderate linguistic term

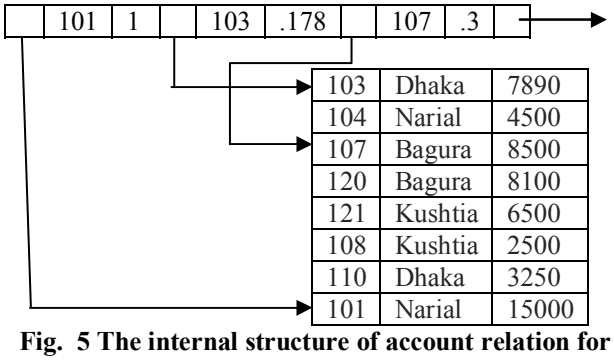


Fig. 5 The internal structure of account relation for high linguistic term

IV. Measurements of Query Cost

The query cost is measured in terms of disk accesses, CPU time to execute a query and the cost of communication [2]. In large database systems, only the number of disk accesses is considered because it is slower than memory operation. In this paper, we have measured the query cost of classical and fuzzy query over classical database and also measured the cost of fuzzy query over fuzzy database.

A. Query Cost of Classical Query over Classical Database

In a query processing, we traverse a path in the tree from the root to some leaf node. If there are k search-key values in the file, at most $\left\lceil \log_{\lceil n/2 \rceil}(k) \right\rceil$ nodes are to be accessed [2]. Typically a node is made to be the same size as a disk block, which is typically 4kb. The cost of the query operation for single record is represented in terms of I/O operations which are equal to the height the tree plus one I/O to fetch for the record; each of these I/O operations requires a seek and a block transfer. Thus, the cost is $\left(\left\lceil \log_{\lceil n/2 \rceil}(k) \right\rceil + 1\right) * (t_S + t_T)$, where t_S, t_T are seek time and block transfer time respectively [2]. If the query produces, **m,** records as output, and then the cost will be $\left(\left\lceil \log_{\lceil n/2 \rceil}(k) \right\rceil + m\right) * (t_S + t_T)$. Typical values for high end disks today would be $t_S = 4$ milliseconds and $t_T = 0.1$ milliseconds. For example, the classical query Q.1 is "find all account numbers whose balance is greater than 7000". The output of this query statement is shown in table 2.

Table 2 Output of Query Statement Q.1

Account_no.	Branch_name	Balance
103	Dhaka	7890
107	Bagura	8550
120	Bagura	8100
101	Narial	15000

In our example, k=8, n=4, m=4. So, a lookup query requires only $\left\lceil \log_{\lceil 4/2 \rceil}(8) \right\rceil = 3$ nodes or blocks to be accessed. So, the query cost of above query statement is $(3+4)*(4+0.1) = 28.07$ milliseconds.

B. Query Cost of Fuzzy Query over Classical Database

The necessary steps to retrieve a record from classical database by using fuzzy queries, with equality on non-key, are as follows:

1. Set C with root node of index file of database which is defined in query statement.
2. Repeat steps 3 and 4 while C is not leaf node.
3. Find the smallest search key K_i of node C.
4. Set C with the node pointed to by P_i.
5. By using P_i extract the record from database.
6. Calculate the fuzzy value in fuzzy set which is defined as linguistic term in fuzzy query statement.

7. If the fuzzy value is greater than zero or greater than or equal to the threshold, then select the record. Otherwise not.
8. Set i=i+1
9. Repeat steps 5, 6, 7 and 8 until the end of the index file.

For example, the fuzzy query Q.2 is "find all account numbers whose balance is high". So, the output of the desired query Q.2 is shown in table-3.

Table 3 Output of Desired Query Q.2

Account_no.	Branch_name	Balance	$\mu_{High}(balance)$
103	Dhaka	7890	.178
107	Bagura	8550	.3
120	Bagura	8100	.22
101	Narial	15000	1

Thus, the query cost of above fuzzy query statement Q.2 is $(3+4)*(4+0.1)=28.07$ milliseconds, which is the same as the classical query on classical database. The difference between the linguistic term of traditional logic and fuzzy logic is that linguistic term of traditional logic does not provide the facility of level of membership function, but the linguistic term of fuzzy logic does. For example, the fuzzy query Q.3 is "finds all account numbers whose balance is high with threshold .3". So, the output of the desired query is shown in table-4.

Table 4 Output of Desired Query Q.3

Account_no.	Branch_name	Balance	$\mu_{High}(balance)$
107	Bagura	8550	.3
101	Narial	15000	1

Thus, we can say that although fuzzy query on classical database increases the expressiveness of human expression, there is no effect on searching time.

C. Query Cost of Fuzzy Query over Fuzzy Database

The necessary steps to retrieve records from fuzzy database by using fuzzy query are as follows:

1. Set C with root node of desired index file which is defined as the linguistic term in query statement.
2. Repeat steps 3 and 4 while C is not leaf node.
3. Find the smallest search key K_i of node C.
4. Set C with the node pointed to by P_i.
5. By using P_i extract the record from database.
6. Test the record by using selection condition whether it is the desired record.
7. If it satisfies the selection condition, then it is the desired record. Otherwise not.
8. Set i=i+1;
9. Repeat steps 5, 6, 7 and 8 until the end of the index file.

For example, the fuzzy query Q.2 is "find all account numbers whose balance is high". So, the output of the desired query is same as table-3.

The fuzzy value is calculated during building the fuzzy database and is stored in the index file named linguistic term. So, extra time to compute fuzzy value is not killed during search any record from the database. In our example, the number of search-key of **high index file** is 4, i.e. k=4, n=4. So a lookup fuzzy query requires only $\left\lceil \log_{\lceil 4/2 \rceil}(4) \right\rceil = 2$ nodes or blocks to be accessed. Since the result contains four records, i.e. m=4, thus the query cost of above fuzzy query statement Q.2 on fuzzy database is $(2+4)*(4+0.1)=24.06$ milliseconds, which is less than Q.2 for fuzzy query on classical database.

V. Results and Discussion

The number of search keys in index file may be reduced by constructing several index files corresponding to the linguistic terms. That is, if the number of distinct records in database is n, then for fuzzy database, each index files contain either n or less than n records, whereas for classical database, the index file contains absolutely n records. The number of search keys in index file of fuzzy database depends on the number of linguistic terms over fuzzy attributes of fuzzy database. So, the searching time of one record from that fuzzy database by using fuzzy query is also reduced. In this paper, the used linguistic terms are three: **low**, **moderate** and **high**. The query costs to retrieve one record from classical/fuzzy databases with different size by using classical/fuzzy query are shown in table-5, where the number of pointers in each node, i.e. **n**, is 20.

Table 5 Experimental Results

For Classical Query		For Fuzzy Query					
		Low		Moderate		High	
K	Query cost (ms)	K	Query cost (ms)	K	Query cost (ms)	K	Query cost (ms)
10000	20.5	7000	19.86	4000	18.87	2000	17.63
		6000	19.6	4500	19.07	3500	18.63
		4000	18.9	5000	19.3	3000	18.4
		2000	17.63	4555	19.1	6500	19.73
		1500	17.12	6000	19.6	5500	19.43
50000	23.4	10000	20.5	30000	22.5	15000	21.2
		5000	19.7	25000	22.13	25000	22.13
		22000	21.9	17000	21.44	13000	20.97
		31453	22.54	24316	22.08	9761	20.46
		19673	21.7	11435	20.74	29431	22.42

Observing the above table, we can conclude that the query cost to find one record from classical database using classical/fuzzy query is less than the cost to find one record from fuzzy database using fuzzy query. For example the searching time using classical query of a record against the classical database with 50000 records is 23.4 ms, whereas for fuzzy query and fuzzy database, that time will be less than 23.4. This time is so effective when the database is so large and the node size is large.

VI. Conclusion

We have seen in fuzzy database the query cost though the searching time is decreased, the memory space is increased because of increasing index file as increasing the linguistic terms. We will extend this work by using fuzzy queries with fuzzy quantifiers, such as **very**, **very, very** in future. By using the fuzzy query on fuzzy database concepts, we will extend traditional database management system which has some intelligence.

References

[1] Seymour Lipschutz, *Theory and Problems of Data Structures,* Mcgraw Hill Education.1999.

[2] Abraham Silberschatz, Henry F. Korth, S. Sudarshan; *Database System Concepts*, Mcgraw Hill Education. 4th edition,2002.

[3] Liberios Vokorokos, Anton Balaz, Norbert Adam, *Parallelism in fuzzy databases*, Teachmedia, 5th Edition, 2006.

[4] S. Rajasekaran and G.A. Vijayalakshmi Pai*, Neural Networks, Fuzzy Logic and Genetic Algorithm-Synthesis and Applications,* Prentice Hall India, 3rd Edition,2003

[5] Al Stevens, *C Database Development*, MIS Press (Pitman) ,2nd Edition, 1991.

[6] Jose Galindo, Angelica Urrutia, Mario Piattini, *Fuzzy Databases Modeling, Design and Implementation.*John Wiley & Sons. Inc.4th Edition, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Solution to the Security Issues of an E-Government Procurement System

Md. Sadiqul Islam, Sanjoy Dey, Gourab Kundu, A.S.M. Latiful Hoque
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000
{sadique96, sanjoydey33}@gmail.com, {gourabkundu, asmlatifulhoque}@cse.buet.ac.bd

Abstract

To make the e-government procurement process totally secure, protecting critical information at database level is absolutely vital along with ensuring secure data transaction over the network. With the assumption that network security in data transaction is ensured, this paper focuses on how to store, process and retrieve data so that critical information remains secured and cannot be viewed by anyone including procuring entities, vendors and even the database administrator. Here, the concept of encryption is put into practice to ensure that the total database management process is secure and fair. Compression is also used together with encryption to keep the volume of encrypted data manageable. A theoretical as well as an implementation oriented treatment to the use of encryption and compression has been presented in this paper with a view to providing security and space-efficiency for the e-government procurement system.

Keywords: Database Security, E-Government Procurement, Information Security

[1]

1 INTRODUCTION

Data transaction is an integral part of a procurement system, whether it is the traditional way of procurement or it is electronic procurement. Providing proper security to the relevant data in all steps of data transaction is a key to the success of an e-government procurement system. In the traditional government procurement process, the procuring entity publishes a call for tender and the interested legitimate vendors submit their quotes to a sealed box within a particular period of time. On a particular day, the sealed box is opened in presence of all the vendors bidding for the project. Then the quotes are evaluated and the winning vendor is announced. This paper focuses on the implementation of the sealed box by presenting a framework which will ensure that quotes submitted by the vendors are as secure as they can be in a sealed box, that is the bidding information cannot be viewed by other vendors, the procuring entity and the database administrator until the bids are legally opened on a particular date and time.

In this paper the development of an e-government procurement system has been considered in the context of a developing country. But this approach can be applied to any government procurement where security and fair competition are major concerns. The requirement of making procurement database secure, arises from the fact that government procurement in a developing country is quite different from that of conventional e-government procurement systems for a number of reasons [7]. In developed country it requires a lower level of security, confidentiality and competitions whereas e-GP of developing country requires fully secured and confidential system at a real-time environment. The system must ensure the unbiased competition among the bidders. So the e-GP of a developing country requires a technology that must have a one-to-one mapping to the existing Government Procurement system. Also the government laws for e-government procurement in different Asian countries impose strict security requirements on data transaction in the whole process [6].

The rest of the paper is organized as follows: Section 2 contains the state of the art on e-government procurement. Section 3 describes an outline for the development of an e-government procurement system. A discussion about the security concerns in the current process is presented in section 4. Section 5 discusses the possible encryption and compression techniques. Section 6 presents different aspects of implementing the secure box. A solution for Personal Identification Number (PIN) recovery is described in section 7 and section 8 is the conclusion.

[1]The corresponding author E-mail: sadique96@gmail.com

2 STATE OF THE ART

Most of the existing works focus on the analysis of the basic process, scope and financial benefits of e-government procurement. Some excellent high level descriptions of government procurement protocols can be found in [10]. A high level description of implementation and maintenance of e-government procurement in Andhra Pradesh of India has been outlined in [4]. While all these works highlight on the high level description of the e-government procurement process, none of them goes into any technical detail of implementation aspects.

Considerable works have been done on designing conceptual frameworks for e-government procurement. A three tier architecture and access control and authorization techniques [17] for e-government procurement have been proposed. The importance of authentication, authorization and security measures in public procurement has been underlined in [9]. The study presented in National Electronic Commerce Coordinating Council [3] emphasizes the importance of database security and prevention of data leakage for a successful procurement. While these works outline some of the technical details of implementation, no further guideline about implementing them has been provided.

Some works, although not great in number, identify the security for data transaction over the network as a key feature for secure e-government procurement. A rigorous analysis on using Public Key Infrastructure (PKI) to provide authentication and authorization to ensure secure data transaction has been presented in [8].

But ensuring secured data transaction over the network provides only a part of the solution. Even if non-repudiated data is stored in the database there is a high possibility that critical information can be obtained from the database by using elementary retrieval methods. Particularly the database administrator having supreme privileges has the access to all the secret data. So some framework for hiding the stored information from the database administrator has to be there. A three tier client-server with an encryption based database protection has been outlined [7]. But no detail on how to implement the encryption based scheme has been provided there. Current condition of e-government procurement in Italy has been presented [13], but with a statement that provides full security to the procurement database is to be achieved. The Website of Government Procurement of India [1] specifies some uses of encryption in providing total security to the database, but the technical details of e-GP is not yet disclosed. This paper attempts to present a detailed methodology, covering all the aspects, for using encryption to ensure proper security and privacy of critical procurement information.

3 AN OUTLINE FOR THE IMPLEMENTATION OF AN E-GOVERNMENT PROCUREMENT PROCESS

The standard for a sealed bid e-government procurement, jointly accepted by world trade organization, world bank and Asian development bank as stated [12] can be described as follows :-

a) All the interested procuring entities and vendors must be registered with the system.
b) The beginning of procurement is marked by the event that the procuring entity publishes a call for tenders. This is published through the web based system and the call remains valid for a particular period of time.
c) The interested vendor will buy the tender documents for that particular procurement through this system.
d) The vendors wishing to participate in the bid will prepare their quotes and submit it to the system. Once a quote is submitted, it cannot be viewed or modified by anyone. Even the procuring entity will not be able to see any quote before the particular date and time of opening the bids.
e) On the particular date and time of opening the tenders, the procuring entity is given access to all the submitted quotes for further evaluation.

We have considered a centralized system for the proposed procurement database. The data model satisfying the requirements outlined in a - e has been given in Figure 1.

4 SECURITY CONCERNS IN THE CURRENT PROCESS

Within the current process framework of e-government procurement, unwanted and unauthorized access to critical information can occur at two levels:

a) During data transaction: Secure Sockets Layer (SSL) has been a very reliable protocol for the prevention of data hacking and unauthorized modification of data while it is being transferred over a network.The official specification of SSL 3.0 is outlined in [2]. VeriSign White Paper[14] has outlined how SSL certifications can be obtained and used by a commercial organization.
b) Database Security: In a successful procurement process, when any vendor purchases any document of a tender, his purchase record with his identity is stored but the record must be kept protected so that no other vendor or the procuring entity or the database administrator can have

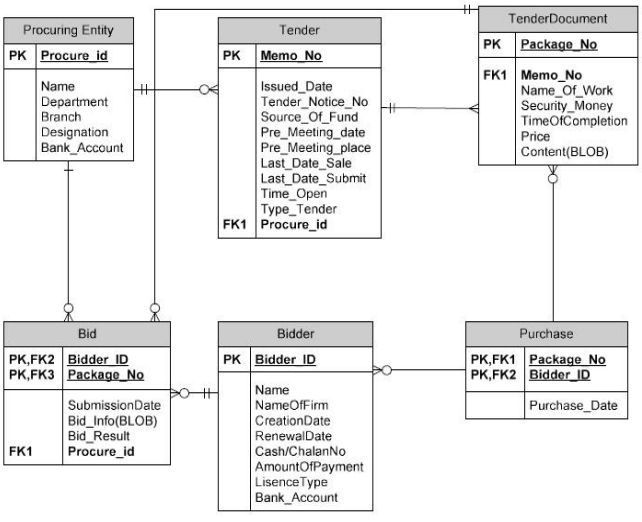


Figure 1. Data model of an E-government procurement System

access to it. In traditional government procurement it is not needed to hide the purchase information from the procuring entity, only the bid & the identity of bidder are hidden from the authority. But in e-government procurement it is needed because if it is not encrypted then the system contains bidder id and package no both in encrypted and decrypted form. Thus the identity of the bidder cannot be hidden properly and may retrieve illegally. Moreover, contents of all bids(bidder info.,quotes etc.) must be protected from everyone including the database administrator until opening time of all bids.

Now, if the data is stored in its plain text, there is a high degree of risk that unwanted access to critical information may not be prevented. In fact at a higher level of the system, total control over user access can be accomplished by using Authentication mechanisms. But this cannot prevent any attempt to retrieve vital information from the central database by executing basic data retrieval methods (i.e. queries) at the database level. Authorization mechanism may provide part of the solution. But this cannot ensure the fact that the database administrator has no access to the private data of the users. So to keep the private information hidden even from the database administrator, data will be encrypted before being stored in the database.

So these security requirements, lead to the introduction of the idea of a 'Secure Box' (Figure 2). It will be working as a layer between the users and the central procurement database to make the data transaction secure. It will encrypt critical user information before storing them in the database and also decrypt the critical information in a secure way while retrieving them on requests made by the appropriate user entitled to access that information.

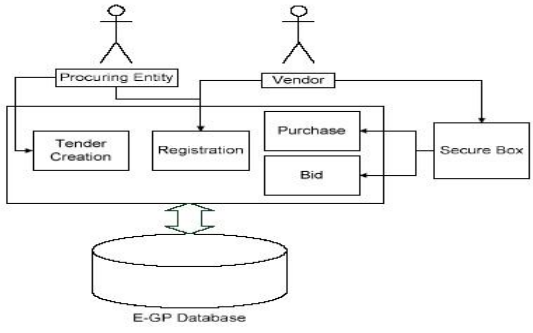


Figure 2. Security Framework

5 ENCRYPTION AND COMPRESSION TECHNIQUES

5.1 Encryption

Over a long period of time, two schemes, namely symmetric key encryption and public key encryption are widely used and have been considered to be very successful.

The main focus of our work is on securing data in the database. And providing a pair of public and private key is computationally more expensive than generating a single key for symmetric key encryption. So in case of designing this framework, symmetric key encryption will be used. Advanced Data Encryption (AES), has been the industrial standard for maintaining database security for commercial purposes. Here the data will be encrypted by Rijndael Algorithm [5]. The key length has been set to be 256 bit.

5.2 Compression

Encryption of the entire bid file is time consuming and computationally expensive. Moreover, all the information in the bid file need not be encrypted; only some portion containing quotation and bidder information need to be encrypted. Also as the proposed system is a centralized e-government procurement system so the amount of bid information is huge. So if compression can be combined with encryption to store the bid information in the bid repository then it will save computation expansion and storage cost.

For compression of the bid file, Huffman Algorithm will be used. Huffman tree is constructed based on the frequency

of characters in a file. In case, two quotations from different vendors are same, there may be a problem since the Huffman trees generated in both cases may be of the same format. So we choose to set the frequency of digits according to the PIN of the vendor so that two same quotations from different vendors will always generate different Huffman trees. In our implementation, the frequency of each digit in the bid file was set to be PIN mod 1000.

5.3 Combining Encryption and Compression

To achieve security and space efficiency, we choose to use both encryption and compression. So at first Huffman tree is constructed, the compressed bid file is created and then the Huffman tree is encrypted and added to the compressed bid file.

6 HOW THE ENCRYPTION SCHEME WORKS

6.1 Protecting the purchase records

The encryption scheme will work as follows:

1) Each vendor, upon buying a tender document, will be given a random string generated by the system as his Personal Identification Number (PIN). This PIN will be used in all the subsequent encryptions where the information of this particular vendor related to that particular package has to be encrypted. One thing is important; this PIN will be generated in a per-document basis. Different PINs will be used for different tender documents and every vendor buying a document package will have a unique PIN. So no vendor will have a constant PIN for every procurement.

The justification for using different PINs for every tender document is that, using a fixed key for every tender document may face a critical problem. The tender document information is likely to contain some well known text and encrypting it every time with the same PIN may provide enough clues about how a certain piece of text is being encrypted. So it leaves a possibility that someone wanting to access the vital information may make an attempt to recognize the pattern of the PIN. On the other hand if the system generates a random key (PIN) depending on the system time, MAC address, vendor and tender document information for every document for a particular vendor and use it for encryption then the scope for recognizing the pattern of a PIN by observing the encrypted data is absolutely minimum.

2) Then the vendor id will be encrypted and will be entered in the Purchase Table along with the package information of the tender document. But here one issue has to be taken into consideration while storing the package information. If the package information is stored without encryption, there is a risk that the system administrator has a chance to retrieve submission time and date of the corresponding package from web log file. Then if the record of the vendors, who have accessed the system within the period of time between submission of that package and the deadline for submitting quotes, is combined with the information about the submission time of a tender package, it can help to identify a number of probable vendors who might have accessed that tender document. The solution to these problem is again encryption. The package information itself can be encrypted with the PIN of the respective vendor before being stored in the purchase record. This leaves hardly any clue in the database about which vendor has purchased which tender document. But at the same time, it generates a problem, that is, in the real life scenario; the procuring entity has the knowledge of how many tender documents it has sold which is quite necessary. The work around this problem is to create a new field in the tender document record which will simply keep a count the number of vendors purchasing that document.

When any vendor submits a quote, the purchase record has to be used to verify whether he has purchased the required documents for that procurement. As the vendor has to mention against which package he is submitting his quote, the vendor has to provide the correct package information. The vendor id along with that package information, before being stored in the bid submission table, will again be encrypted by the PIN. Thus the vendor id and the package information in bid submission table have the same encrypted form as they have in the purchase table. So establishing a relationship between these two tables is not handicapped by this encryption scheme.

6.2 Making sure that no vendor can see or edit his quote after dropping it

The intending vendor will be asked for PIN of the tender document when he wants to drop a bid. The secure box then encrypt vendor id and package no using the PIN. If it finds an entry in the bid table it will inform the vendor that he has already submit the bid and cannot see the quote before openning time.

6.3 About the retrieval of the bid information

The quotes from different vendors will be made available to the procuring entity on the particular date and time of opening the quotes. For this purpose, when the time for opening the quotes comes, a notification will be made to all the competing vendors to submit their respective PINs for that particular procurement to the e-government procurement system within a particular period of time.

The bidders will logon to the system when the time of opening the tender box arrives. Then the procuring entity will ping them to provide the corresponding PIN number from that user. As the bidder is logged on at that time, the procuring entity can easily retract the bid from tender box. Here logging on of the bidders resembles much like the physical presence of actual system. The status of the bidder as logged on can be easily implemented by checking out the session of the corresponding bidder. The procuring entity then can see all the bidders in a page and can perform decryptions.

6.4 PROCESS DESCRIPTIONS

`Algorithm I: Tender document Purchase`
1. Get the vendor information over SSL.
2. Select the row in the Vendor table using vendor ID and check validity. If no row is selected or license is expired go to step 8.
3. Generate a random number using system time, MAC address, Vendor ID and Tender document package information [16]. Assign it as a PIN to that vendor for that particular package.
4. Encrypt the document ID and vendor ID by PIN as the key using the algorithm presented in the Encryption Algorithm to be used of Section V.
5. Record the purchase by inserting encrypted user ID, encrypted document ID in the Purchase Record.
6. Download tender document from tender repository.
7. If more copies of documents are requested go to step (2).
8. Stop purchase.

`Algorithm II: Tender document Submission or bid`
1. Vendors log in the website using user ID and password over SSL.
2. Submit general information about tender over SSL.
3. Encrypt the vendor ID, document ID by PIN as the key using the algorithm presented in the Section V.
4. Select the row using encrypted vendor ID and document ID in the Purchase Record. If no row is selected go to step (8).
5. Upload the tender Doc in the tender box over SSL protocol.
6. Encrypt the upload stream of the tender document by this PIN using the algorithm presented in the Section V.
7. Store the encrypted stream as BLOB into Bid table with encrypted user ID and document ID.
8. Stop bid submission.

`Algorithm III: Opening Tender Box`
1. Check whether it is time for opening the bids officially by watching system time. If not, stay in this step.
2. Collect user ID and PIN from all competing bidders over SSL using any of the strategies discussed in Section VI (D).
3. Encrypt the user ID with this PIN as key using the algorithm presented in Section V.
4. Select the row using encrypted vendor ID and document ID in the Purchase table and decrypt this with the PIN using the algorithm presented in Section V.
5. Select the row using encrypted vendor ID and document ID in the Bid table and decrypt this row with the PIN using the algorithm presented in the appendix.
6. Match the decrypt vendor ID and document ID generated in step 4 and 5. If these are not same stop opening. This ensures the referential constraints of Purchase and Bid tables.
7. Find the encrypted tender document from the bid table and decrypt the tender document using the PIN.
8. System prints the general information and financial offers.
9. System verifies the general information.
10. Repeat step 2 to step 9 for all bidders.
11. System prepares the comparative statements of all financial offers.

7 A SOLUTION OF PIN RECOVERY SCHEME

Encryption with a distinct PIN for each vendor is at the core of providing secure data transaction. The database stores no clues whatsoever about the PIN that is used for each user. And as we see, without providing the correct key, the decryption function will never generates the original data, it is vital that the user can provide his PIN correctly. Although it is expected that in an event of such magnitude, the user will be careful enough to keep track of his PIN, it is always humane to forget any information. In that case, there has to be a recovery scheme for the user so that he can retrieve his PIN. For pin recovery, we have considered the following scheme.

To enable a recovery scheme, a separate record called 'Recovery' will be kept. The idea behind this recovery scheme is to use a security question. When a vendor purchases documentation, his user id and package no for

that document is stored in the Purchase record. In the Recovery record, information in three fields is stored. These are - user id, security question and PIN.

`Algorithm I (Storing PIN information):`

1. During purchase of document the vendor is given a choice to select from a list of security questions. The vendor chooses a security question and answers the question.
2. The user id of the vendor is encrypted by the corresponding package number of his/her purchase document.
3. The security question is encrypted using user id.
4. The PIN of the vendor is encrypted by a key which is a function of user id, package number, password, and answer to the security question. This function can be a simple Boolean operation. But it must be a one to one function so that its inverse function is available.
5. Then the encrypted user id, security question and PIN is stored in Recovery table.

`Algorithm II (Retrieving PIN information):`

1. The vendor has to log in and inform the system about the package number of that tender document.
2. The system can encrypt the user id with that package number and then can search the first entry of the Recovery record for a match.
3. The system then decrypts the security question with this decrypted user id and then asks vendor the security question and gets the answer.
4. If the correct answer is given, then the correct key will be generated by the function described in the Step 4 of previous algorithm.
5. By the key generated in step 5, the PIN can be recovered from the third column of 'Recovery' table.
6. The system will separately encrypt both the user id and the package number with that PIN and then look up the Purchase Record to see whether there is any such record which contains the combination of the encrypted user id and package number just formed previously. If any such record exists, the PIN is correctly retrieved and delivered to the vendor. However, if there is no such record, then the PIN is not correctly recovered and the vendor is again prompted to give a correct answer to his security question.

8 CONCLUSION

In this paper, a framework for providing proper security to critical information in e-GP procurement database has been presented. To implement the tender box, which is the most critical part of the system, an encryption based Private Information Retrieval (PIR) is used. Also a recovery procedure for recovering the PIN of a vendor on which the encryption scheme is established has been proposed.

References

[1] The website of Government Procurement of India. http://www.eprocurement.gov.in.

[2] Specification of SSL version 3.0. http://wp.netscape.com/eng/ssl3/ssl-toc.html, 1996.

[3] E-Government Procurement Policy Issues. In *National Electronic Commerce Coordinating Council*, United States Of America, December, 2000.

[4] K. Bikshapathi and N. K. C. Reddy. Setting Up Implementation Operation and Maintenance of E-government procurement Exchange for The Government of Andhra Pradesh. In *International Conference on E-Government*, Hyderabad, India, July, 2005.

[5] J. Daeman and V. Rijman. AES Proposal: The Rijndael Block Cipher. In *AES Algorithm Submission*, http://csrc.nist.gov/CryptoToolkit/aes, September 1999.

[6] A. Guenou. Legal Aspects of E-Government Procurement. In *International Conference on E-government procurement*, Seoul, Republic of Korea, June,2005.

[7] A. Hoque, M. Hasan, M. Parvez, G. Hossain, and M. Shah. E-Government Procurement of Developing Countries: Problems and Prospects. In *International Conference on ICT for The Muslim World*, Kualalampur, Malaysia, November,2006.

[8] C. M. Jang. PKI Based Secure E-Government Procurement: Digital Signature and Public Key Infrastructure. In *International Conference on E-Government Procurement*, Seoul, Republic of Korea, June,2005.

[9] M. Kamoto. E Bidding System for Public Procurement in Japan. In *International Conference on E-Government Procurement*, Seoul, Republic of Korea, June,2005.

[10] R. Kramer. Protocols for Government Procurement of Software Assets. In *International Conference on E-Government Procurement*, Seoul, Republic of Korea, June,2005.

[11] Y. J. Lee. Standardization of E-Government Procurement: Business Requirements, Current Status and Proposals. In *International Conference on E-Government Procurement*, Seoul, Republic of Korea, June,2005.

[12] P. Magrini. Transparency in Public Procurement: The Italian Perspective. In *1st High Level Seminar on E-Government Procurement*, Naples, Italy, January, 2006.

[13] V. W. Paper. How to Offer The Strongest SSL Encryption. http://www.verisign.com/static/016585.pdf.

[14] W. Stallings. SSL: Foundation for Web Security. *The Internet Protocol Journal*, 1, June,1998.

[15] R. Tausworthe. Random Numbers Generated by Linear Recurrence Modulo 2. *Math. Comput*, 19:201–209, 1965.

[16] S. J. Won. National e-government procurement Service Experience. In *International Conference on E-Government Procurement*, Seoul, Republic of Korea, June,2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Voltage Mode Control of Single Phase Boost Inverter

Ainul Anam Shahjamal Khan

Department of Electrical and Electronic Engineering,
Chittagong University of Engineering and Technology,
Chittagong-4349, Bangladesh
E-mail: khanshahjamal@yahoo.com

Kazi Mujibur Rahman

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology,
Dhaka, Bangladesh
E-mail: kmr@eee.buet.ac.bd

Abstract – Boost inverter is able to generate an ac voltage whose peak value can be larger or smaller than the dc input voltage in a single stage. The boost inverter consists of two individual boost dc-dc converters. Each of the boost dc-dc converters has to be controlled in variable operating point condition. This paper proposes two voltage mode controllers. First one is conventional feedback controller based on small signal model. Second one is a simplified voltage mode controller based on Feedforward Pulse Width Modulation (FF-PWM) and simple voltage feedback loop to control the output voltage of the boost inverter. The second controller gave better result. The controller also ensures output voltage to meet IEEE Std 519 voltage harmonic limits. Unlike most previously proposed controllers, the proposed method is based only on output voltage feedback, making unnecessary the inductor current measurement. The control method is verified by means of simulations.

I. Introduction

Boost DC-AC inverter [l] is a novel converter, whose main advantage is to achieve an output voltage higher or lower than the input one. Other advantages are the quality of output voltage sine wave and reduced number of switches i.e. only four switches required.

This property is not found in the traditional full bridge inverter, which produces an instantaneous ac output voltage always lower than the input dc voltage. The power stage of boost inverter consists of two current bi-directional dc-dc boost converters and the load is connected differentially across them as in Fig. 1. Each converter produce a dc-biased sinusoidal waveform as Fig. 2. The modulation of each converter is 180 degrees out of phase with respect to the other, which maximizes the voltage excursion across the load.

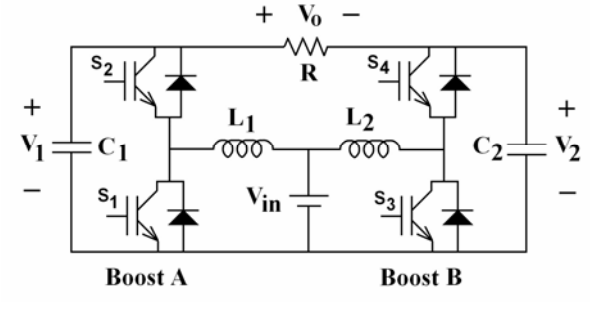


Fig. 1 The boost inverter.

The DC and small-signal performance of a boost DC to AC converter can be determined by substituting the circuit by PWM switch model [2,3] and analyzing the resultant linear circuit. The converter can be modeled using the models of the current-bidirectional switch based Phase-leg averaging technique [4].

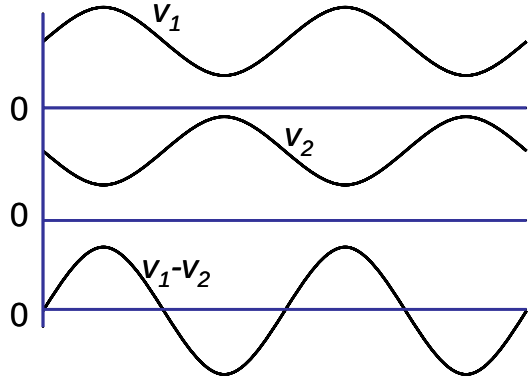


Fig. 2 Typical waveforms of the basic approach to achieve dc–ac conversion, with boost characteristics.

Several control techniques have been proposed to control the individual boost of the boost inverter [1, 5-7]. The sliding mode controller [1] involves complex theory, variable switching frequency and also involves current mode control in addition to voltage mode. The effect of supply voltage variation on output voltage is not reported. A controller based on Energy Shaping theory [5] is proposed which is a complex one. An adaptive control [6] designed for the boost inverter in order to cope with RL load. This involves very complex non-linear control theory. A double-loop regulation scheme [7] is a very robust controller including compensations in order to cope with the boost variable operation point condition. But it has also current loop and circuit complexity. So a simplified controller with only voltage mode control and meet the IEEE Standard for THD of the output voltage can be a simplified option to control the boost inverter.

This paper proposes two voltage mode control strategy. First one is based on traditional design of feedback loops which is based on frequency domain analysis after linearization. A linearized model of the boost inverter is determined and control to output transfer function is obtained from it. Second method is based on FF-PWM [8] which has inherent capability of reducing input source disturbances, improved steady-state and dynamic responses. A simplified outer voltage feedback loop is found to be sufficient to get good regulation of output

voltage under different load and source variations. The main advantage of this method over current mode control is its simplicity.

II. Modelling of the Boost Inverter and Conventional Control Approach

The boost inverter is modeled in [9] using PWM switch model described in [2]. Here we used a simplified method to model it. The main step is to replace converter phase legs with voltage and current sources [4]. The waveforms of voltage and current sources are identical to the switch waveforms of the converter. Then the converter waveforms are averaged over one switching period to remove switching harmonics. Any nonlinear element in the phase leg's average model is perturbed and linearized leading to small-signal ac model [10].

The boost inverter is usually classified as current bidirectional converters because they share the same switching cells that are current bidirectional. The switching network averaging is performed on a phase-leg basis [4]. After the phase-leg averaging, the average model of any current-bidirectional converter can be easily obtained by connecting the averaged phase legs.

In the current-bidirectional switch based converters, a generic switching unit, called a phase leg, can be identified, as shown in Fig. 3.

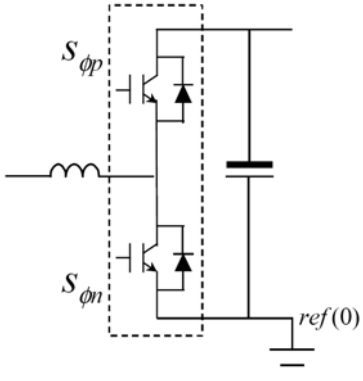


Fig. 3 Generic phase leg in current-bidirectional converters.

The phase leg is composed of two switching cells, and has a voltage source (or a capacitor) on one side and a current source (or an inductor) on the other. These features make the phase leg a generic switching unit. The phase leg can be represented by a single-pole, double-throw switch, as shown in Fig. 4.

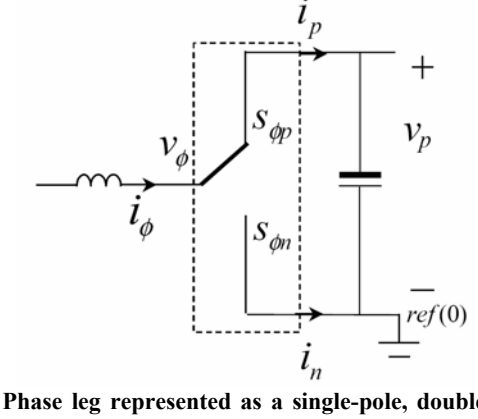


Fig. 4 Phase leg represented as a single-pole, double-throw switch.

A. DC model of PWM switch using Phase-Leg Average Model

Let us first assume that the duty ratio is fixed at D_ϕ and it is the duty cycle of the top switch $S_{\phi p}$. The PWM of the phase leg in DC is shown in Fig. 5, where T is the switching period. The corresponding voltage and current waveforms are also shown in Fig. 5.

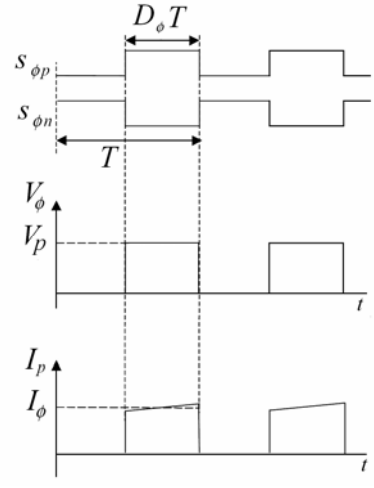


Fig. 5 PWM of Phase leg in DC and corresponding current and voltage waveforms

Based on the waveforms, the voltage and current relationships in average, assuming the current I_ϕ and the voltage V_p are continuous with small ripples:

$$V_\phi = D_\phi V_p \quad (1)$$

$$I_p = D_\phi I_\phi \quad (2)$$

The average DC model based on (1) and (2) of the phase leg is depicted in Fig. 6.

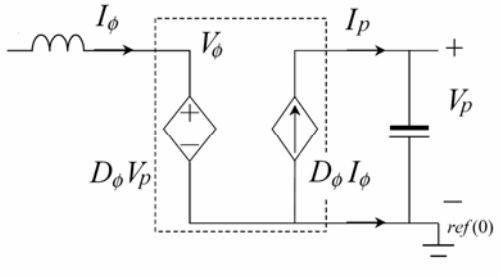


Fig. 6 Phase leg's average DC model.

B. Small signal model of PWM switch using Phase-Leg Average model

To construct a small-signal ac model at a quiescent operating point we assume that, duty cycle is perturbed around a quiescent value D_ϕ such that.

$$\langle d_\phi(t) \rangle = D_\phi + d_\phi(t) \quad (3)$$

where $d_\phi(t)$ is small ac variation. Using similar arguments we can write

$$\langle v_\phi(t) \rangle_{T_S} = V_\phi + v_\phi(t) \text{ etc.} \quad (4)$$

With the assumption that the ac variations are small in magnitude compared to the dc quiescent values, such as

$v_\phi(t) << V_\phi$. We will employ the basic approximation of removing the high-frequency switching ripple by averaging over one switching period

$$\langle x(t)\rangle_{T_S} = \frac{1}{T_S}\int_{t}^{t+T_S} x(\tau)d\tau \tag{5}$$

By averaging all signals over one switching period using equation (1)

$$\langle v_\phi(t)\rangle_{T_S} = \langle d_\phi(t)\rangle\langle v_p(t)\rangle_{T_S} \tag{6}$$

$$\Rightarrow V_\phi + v_\phi(t) = \underbrace{D_\phi V_p}_{DC\ terms} + \underbrace{d_\phi(t)V_p + D_\phi v_p(t)}_{1st\ order\ ac\ terms(linear)} + \underbrace{d_\phi(t)v_p(t)}_{2nd\ order\ ac\ terms(nonlinear)} \tag{7}$$

We can neglect the non-linear ac terms provided that the small signal assumption is satisfied, then each of the second order non-linear terms is much smaller in magnitude than on or more of the linear first-order terms [10]. The linearization step can be explained more analytically by taking the Taylor expansion of a nonlinear relation and retaining only constant and linear terms. The dc terms on the right hand side of the equation are equal to the dc terms on the left hand side. Then (7) becomes

$$v_\phi(t) = D_\phi v_p(t) + d_\phi(t)V_p \tag{8}$$

Using similar arguments

$$i_p(t) = D_\phi i_\phi(t) + d_\phi(t)I_\phi \tag{9}$$

For convenience we use $d_\phi(t) \Rightarrow d_\phi$, $v_\phi(t) \Rightarrow v_\phi$ etc.

Then equations (10) and (11) becomes

$$v_\phi = D_\phi v_p + d_\phi V_p \tag{10}$$

$$i_p = D_\phi i_\phi + d_\phi I_\phi \tag{11}$$

Equations (10) and (11) are small signal linearized equation that describes the phase-leg's average model in ac. The small signal model using equations (10) and (11) of the phase leg is depicted in Fig. 7.

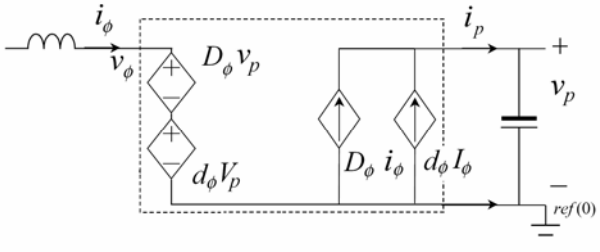


Fig. 7 Phase leg's average small signal ac model.

C. DC Analysis

Phase leg's average DC model in the boost inverter of Fig. 6 is substituted into the dc-ac boost inverter circuit of Fig. 1. The resulting circuit is shown in Fig. 8. In this circuit all inductances are shorted and capacitances are opened which is required for DC.

After analyzing the circuit of Fig. 8 we find the following relations:

$$V_1 = \frac{V_{in}}{(1-D)};\ V_2 = \frac{V_{in}}{D};\ V_o = \frac{V_{in}(2D-1)}{D(1-D)}$$

$$I_{L1} = \frac{V_{in}(2D-1)}{2RD(1-D)^2};\ I_{L2} = \frac{-V_{in}(2D-1)}{2RD^2(1-D)}$$

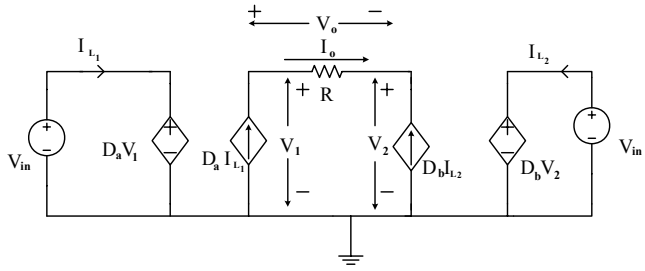


Fig. 8 DC model of the boost inverter.

D. Control to Output Transfer Function

In this case we are only considering the perturbations in the duty ratio. Here the input voltage source V_{in} is shorted to ground and the Phase leg's average small signal ac model of Fig. 7 is inserted in Fig. 1 as shown in Fig. 9.

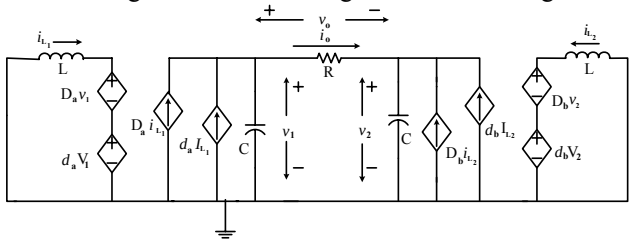


Fig. 9 Small signal ac model of the boost inverter

Here $d_b = d$, $d_a = -d$, $D_b = D$, $D_a = 1-D$

After analyzing the circuit of Fig. 9 we obtain the following control to output transfer function

$$\frac{v_o(s)}{d(s)} = -R\frac{Q_1s^3 + Q_2s^2 + Q_3s + Q_4}{T_1s^4 + T_2s^3 + T_3s^2 + T_4s + T_5} \tag{12}$$

Where, v_o = output voltage; d = variation of duty cycle around 0.5; R =load resistance

$Q_1 = (I_{L_1} + I_{L_2})CL^2$; $Q_2 = [(D-1)V_1 - DV_2]CL$;

$Q_3 = [I_{L_1}D^2 + I_{L_2}(D-1)^2]L$;

$Q_4 = [DV_1 - (D-1)V_2](D-1)D$;

$T_1 = RC^2L^2$; $T_2 = 2CL^2$; $T_3 = RCL(1-2D+2D^2)$;

$T_4 = L(1-2D+2D^2)$; $T_5 = R(D-1)^2D^2$

E. Conventional Control Using Small Signal Model

A conventional voltage mode controller is designed using the transfer function. We designed it taking the following parameters as shown in Fig. 10. $H = 1/36$, $V_M = 10$, $RC = 2.2736\times10^{-4}$, low pass filter cut off $f_{LP} = 700$, $k_p = 0.2$, $k_i = 50$. The capacitance, inductance and source voltage values of [1] are chosen. Capacitors 40uF each; Inductors 800uH each; Vin=100V. When the phase margin of the loop gain T is positive, then the feedback system is stable. Moreover increasing the phase margin causes the system transient response to be better behaved, with less overshoot and ringing [10]. The bode plot of T reveals that there is only one crossover frequency and phase margin of T is 113 degree

which is positive. So the feedback system is stable with good transient response.

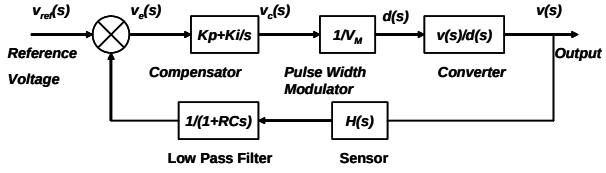


Fig. 10 Small signal block diagram of conventional voltage mode control.

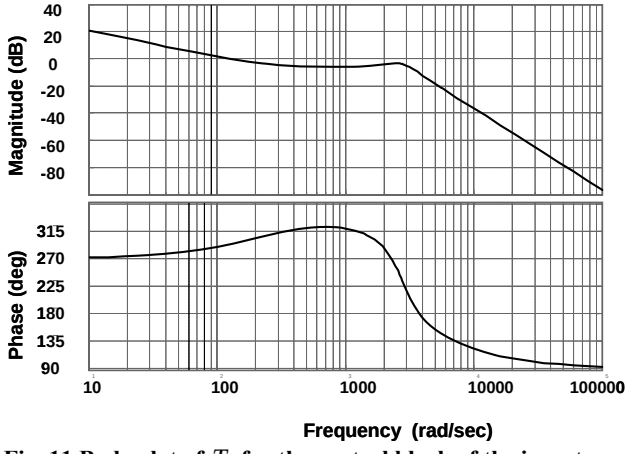


Fig. 11 Bode plot of T for the control block of the inverter.

F. Simulation Results

The inverter was simulated in ORCAD using the following parameters along with those mentioned in last section: S1,S2,S3,S4=> IRGBC40F; Diodes D1-D4=> MR2406F; Lock out time=2us; Switching frequency fc=5k; Nominal output voltage, Vomax=182.2 and frequency=50Hz and Nominal Load=30Ω. For load 30 Ω, THD=4.76% and Fourier component of fundamental voltage=182.2V. For load 60 Ω, THD=3.91% and fourier component of fundamental voltage=185.6V.

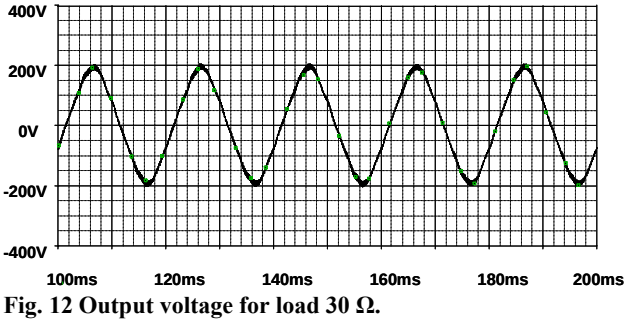


Fig. 12 Output voltage for load 30 Ω.

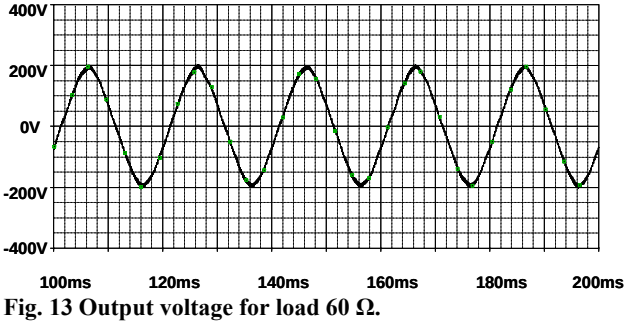


Fig. 13 Output voltage for load 60 Ω.

The controller works moderately well with resistive load keeping THD within IEEE Std 519 voltage harmonic limits. But it fails to regulate the output for inductive loads. In next section feedforward compensation technique is used to regulate the output voltage against source and inductive and resistive load variations.

III. Control of the Boost Inverter by FF-PWM with Simple Voltage Feedback Loop

For switching converters, feedforward compensation [8] is effective in reducing effects of source disturbances on converter outputs and improving steady-state and dynamic responses. The FF-PWM [8] yields good wide-range open loop line regulation and simplified design of an outer voltage feedback loop. A converter with FF-PWM behaves at low frequencies as a linear power amplifier with constant gain independent of operating conditions such that $V_o = Av_m$ where v_m is the modulating input, and A is a constant, independent of operating conditions. The saw-tooth carrier waveform of conventional PWM is periodic as in Fig. 14, it is sufficient to define the modulator during one switching period, say from 0 to T_s.

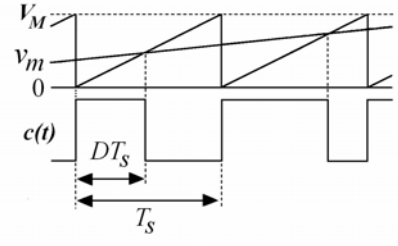


Fig. 14 Typical waveforms in the conventional PWM.

In this interval, the output logic-level function $c(t)$ is determined by

$$c(t) = 1, \quad \text{if} \quad g(t/T_s, v_m) > 0 \text{ and}$$
$$c(t) = 0, \quad \text{if} \quad g(t/T_s, v_m) \le 0, \quad 0 \le t \le T_s \qquad (13)$$
$$g(t/T_s, v_m) = v_m - V_M(t/T_s) \qquad (14)$$

Where, $g(t/T_s, v_m)$ is the modulator function. The output duty ratio D of the pulsating waveform $c(t)$ solves the equation

$$g(D, v_m) = 0 \qquad (15)$$

where $D \longrightarrow t/T_s$. In duty-ratio controlled switch-mode power converters operating in the continuous conduction mode, the output voltage V_o is a function of the input voltage V_g and the duty ratio D,

$$\frac{V_o}{V_g} = M(D) = \frac{P(D)}{Q(D)} \qquad (16)$$

The basics of FF-PWM are briefly described below for clarity. Detail analysis is presented in [8]. The first step in the synthesis of FF-PWM is to remove the assumption about the up-going saw-tooth carrier waveform and to allow feedforward inputs for the disturbance (independent source) that is desired to compensate. The converter is supplied from a single voltage source V_g. A modulator function, $g(t/T_s, V_g, v_m)$ is constructed so that the converter steady-state output V_o is independent of V_g

and directly proportional to the modulating control input v_m i.e.

$$V_o = Av_m \tag{17}$$

where A is a constant. If we combine the objective (17) and the conversion ratio (16), the duty ratio at the output of the FF-PWM must solve

$$\frac{V_o}{V_g} = \frac{P(D)}{Q(D)} \Rightarrow v_m Q(D) - \frac{1}{A} V_g P(D) = 0 \tag{18}$$

Equations (18) and (15) are similar equation with same variables. So equation (18) also gives a modulator function for the FF-PWM, if we let $D \longrightarrow t/T_s$.

So, a FF-PWM for any converter with specified $M(D)$ is defined by the modified modulator function

$$g_{TE}(t/T_s, V_g, v_m) = v_m Q(t/T_s) - \frac{1}{A} V_g P(t/T_s) \tag{19}$$

The subscript TE indicates the usual trailing-edge modulator. An equivalent leading-edge FF-PWM can be obtained simply by taking $D' = (1-D) \longrightarrow t/T_s$ in (18), which yields

$$g_{LE}(t/T_s, V_g, v_m) = v_m Q(1-t/T_s) - \frac{1}{A} V_g P(1-t/T_s) \tag{20}$$

In the modulator functions given by (19) or (20), there are terms of the form, $u(t)(t/T_s)^k$, where $u(t)$ is a linear combination of V_g and v_m . These terms can be implemented based on the use of integrators with reset.

For $k = 1$ we have,

$$u(t)\left(\frac{t}{T_s}\right) \approx \frac{1}{T_s}\int_0^t u(\tau)d\tau \tag{21}$$

The required building blocks include a voltage comparator, an integrator with reset, and a monostable pulse circuit (one-shot) with a short pulse output.

A voltage comparator outputs a logic high (‘‘1’’) if $v^+ > v^-$, and a logic low (‘‘0’’) if $v^+ \le v^-$, where v^+ and v^- are the voltages at the + and - the input of the comparator. An integrator with reset produces zero output $v_o = 0$ if the logic-level reset input R is high ($R = 1$), or the integral of the input v_i if $R = 0$. Assuming that R goes from 1 to 0 at $t = 0$, the output is given by

$$v_o = \frac{1}{T_i}\int_0^t v_i(\tau)d\tau \tag{22}$$

A constant-frequency clock generator CLK is required which outputs short pulses with period $T_s = T_i$ where T_i is the time-constant of the integrator. For the boost converter $P(D) = 1$ and $Q(D) = 1 - D$.

A leading-edge modulator function for the boost converter is

$$g_{LE}(t/T_s, V_g, v_m) = v_m \frac{t}{T_s} - \frac{1}{A} V_g \tag{23}$$

The implementation is shown in Fig. 15, with typical waveforms shown in Fig. 16.

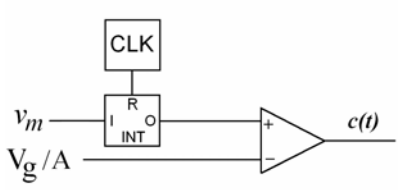


Fig. 15 Leading-edge FF-PWM for the boost converter.

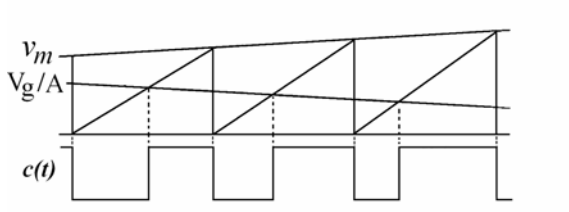


Fig. 16 Typical waveforms in the FF-PWM for the boost converter with leading-edge modulator.

The leading edge modulator is simpler than trailing edge modulator and in the leading-edge FF-PWM, the integrator time constant T_i does not have to match the switching period T_s [8]

A. Simplification of the Feedback Loop using FF-PWM

The FF-PWM results in a constant control-to-output gain. To regulate the output voltage, an outer voltage feedback loop is closed with fewer difficulties as shown in Fig. 17 Here we used two separate controllers for each of the individual dc-dc boost converter.

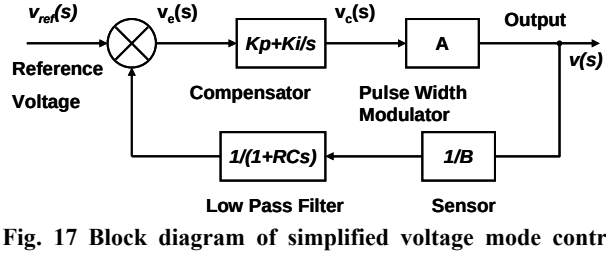


Fig. 17 Block diagram of simplified voltage mode control using FF-PWM.

We selected the same circuit component values [1] with $B = A$; $k_p = 0.3$, $k_i = 50$. The bode plot of T is shows is only one crossover frequency and phase margin of T is 107 degree which is positive. So the feedback system is stable with good transient response.

B. Simulation Results:

The inverter was simulated in ORCAD using the same parameters mentioned in last section II along with the following ones: Gain of feedforward pulse width modulator A=58.75; Sensor gain 1/B=1/58.75; Clock pulse =1us; Lock out time=2us; Switching frequency fc=5k; Nominal output voltage Vomax=185.8V frequency=50Hz and Nominal Load=30Ω. For load 30 Ω, THD=2.43%% and Fourier component of fundamental voltage=185.8V. For load 60 Ω, THD=1.07% and Fourier component of fundamental voltage=188V. For load, Load R=30 ohm and L=50mH, THD=3.29% and Fourier component of fundamental voltage=181V.

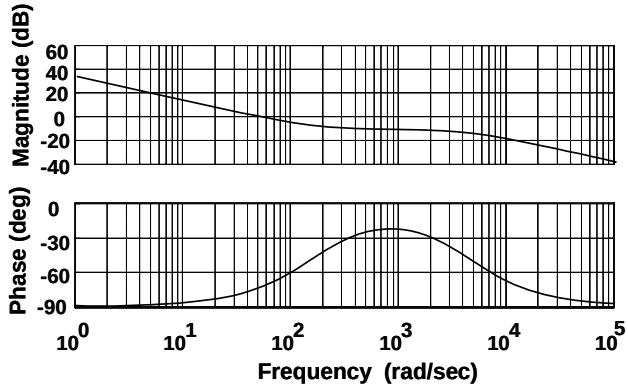


Fig. 18 Bode plot of T for the control block using FF-PWM

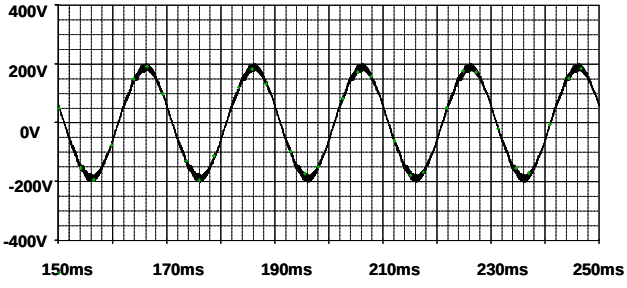


Fig. 19 Output voltage for load 30 Ω

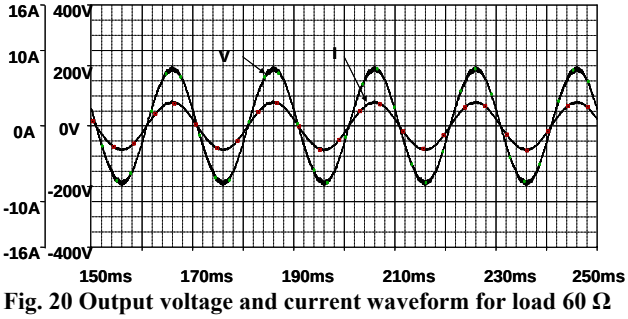


Fig. 20 Output voltage and current waveform for load 60 Ω

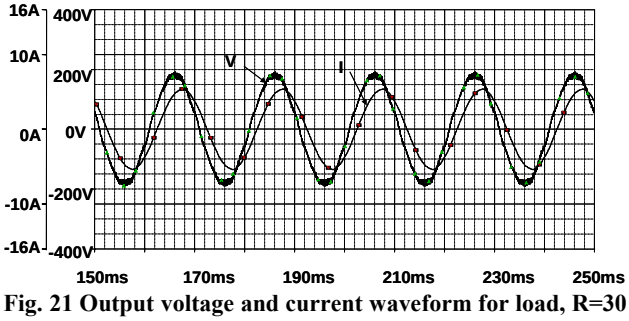


Fig. 21 Output voltage and current waveform for load, R=30 ohm and L=50mH.

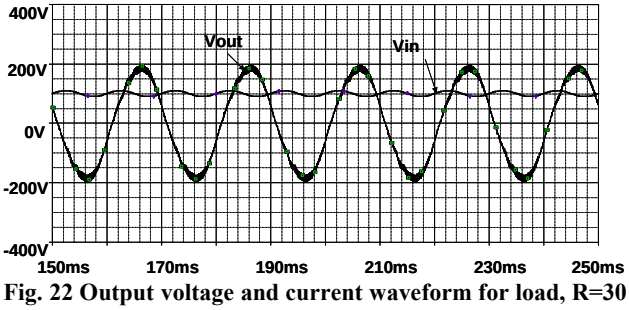


Fig. 22 Output voltage and current waveform for load, R=30 ohm and source voltage, Vin=100+10sin2ωt

For load 30 Ω, and Input voltage, Vg = Vin = 100 + 10 sin(2ωt), THD = 3.80% and Fourier component of fundamental voltage=186.2V. The controller works extremely well to regulate output voltage with source variation, inductive and resistive load variations as well as keeping THD within IEEE Std 519 i.e. THD<5%.

IV. Conclusion

This paper presents a simplified control strategy for the boost inverter with load and source variations. The method is based on feedforward compensation of source variation and voltage feedback technique using PI (proportional plus integral) compensator. It is shown that the resultant controller achieves the objective quite well. The results are tested by simulations. Future research can be done considering capacitive load and reduction of THD using second method. There is also scope to modify the conventional voltage mode controller considering inductive loads.

References

[1] Ram´on O. C´aceres and Ivo Barbi, "A Boost DC–AC Converter: Analysis, Design, and Experimentation," IEEE Transactions on Power Electronics, vol. 14, no. 1, pp. 134-141, January 1999.

[2] Vatche Vorperian, "Simplified Analysis of PWM Converters Using Model of PWM Switch Part I: Continuous Conduction Mode," IEEE Transactions on Aerospace and Electronic Systems, vol. 24, no. 3, pp. 490-496, May 1990.

[3] Khai D. T. Ngo, "Alternate Forms of the PWM Switch Models," IEEE Transactions on Aerospace and Electronic Systems, vol. 35, no. 4, pp. 1283-1292, October 1999.

[4] http://scholar.lib.vt.edu/theses/available/etd-11062000-08510043/unrestricted/Chapter2.pdf. From (DLA) Digital Library and Archives, Virginia Tech. DLA, 1989.

[5] Carolina Albea and Francisco Gordillo, "Control of the Boost DC-AC Converter with RL Load by Energy Shaping," Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, USA, pp. 2417-2422, Dec. 12-14, 2007.

[6] Carolina Albea, Carlos Canudas-de-Wit and Francisco Gordillo, "Adaptive Control of the Boost DC-AC Converter," 16th IEEE International Conference on Control Applications Part of IEEE Multi-conference on Systems and Control Singapore, pp. 611-616, 1-3 October 2007.

[7] C Pablo Sanchis, Alfredo Ursæa, Eugenio Gubía and Luis Marroyo, "Boost DC–AC Inverter: A New Control Strategy," IEEE Transactions on Power Electronics, vol. 20, no. 2, pp. 343-353, March 2005.

[8] Barry Arbetter and Dragan Maksimovic, "Feedforward Pulse Width Modulators for Switching Power Converters," IEEE Transactions on Power Electronics, vol. 12, no. 2, pp. 361-368, March 1997.

[9] Ram´on O. C´aceres and Ivo Barbi, "A Boost DC–AC Converter: Operation, Analysis, and Experimentation," Proc. IEEE IECON'95 Conf., Orlando, FL, pp. 546-551, 5-11 November 1995.

[10] http://www.engr.colostate.edu/ECE562/lectures.html

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Development of Control Strategy for Load Sharing in Grid-Connected PV Power System

Muhammad Quamruzzaman
Department of Electrical and Electronic Engineering
Chittagong University of Engineering & Technology
Chittagong-4349
Email: qzaman359@yahoo.com

Kazi Mujibur Rahman
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering & Technology
Dhaka-1000
Email: kmr@eee.buet.ac.bd

Abstract – In a power system of conventional parallel connected generators, share of real power and reactive var of an incoming generator are controlled by adjusting shaft power input and field excitation. The scenario of load sharing by a grid connected PV system is however different since no prime mover or excitation source are present. Due to serious power crisis, there are needs for transfer of PV power to grid systems. However, this needs intensive analysis on the load sharing phenomena. In this paper, aspects of load sharing of a grid connected PV system are analyzed, and a strategy is proposed where the load sharing task can be undertaken controlling both modulation index and phase angle of the inverter. It is seen that both real power and reactive var are affected upon change in index of modulation and phase angle of the PV inverter. Analysis and simulation results are presented to demonstrate effectiveness of the proposed control technique.

I. Introduction

In today's world, electricity is a vital ingredient for both economic and social development. Adequate, reliable and reasonably priced supply of electricity is an essential pre-requisite for national development. Due to the growing energy consumption around the world and the eminent exhaustion of fossil-fuel reserves, a great interest on alternative energy sources can be noticed nowadays. The threat of electrical energy rationing, blackouts, and overtaxes, in addition to the great environmental awareness, increases the requirement of research on alternative renewable energy systems. Among the clean and green power sources, the photovoltaic (PV) solar energy comes up as being a very interesting alternative to supplement the generation of electricity.

Photovoltaic (PV) arrays as an alternative energy source has been becoming feasible due to enormous researches and development work being conducted over a wide area. Interconnecting a PV system with utility is the current design trend. The advancement of power electronics and semiconductor technologies, the declining cost of solar panels, and the favorable incentives in a number of countries had profound impact on the commercial acceptance of grid-connected PV systems—which have been used in peak shaving, demand reduction, and supply of remote loads [1], [2]. Where utility power is available, consumers can use a grid-connected PV system to supply a portion of the power they need while using utility generated power at night and on very cloudy days. Grid-connected PV systems can provide most of a consumer's need. If the generation is greater than consumer's demand, the excess electricity can be fed through a meter back into the utility grid.

The dc power obtained from PV array is converted to ac through inverter and fed to the load. In case of conventional parallel connected generators, share of real power and reactive var of an incoming generator are controlled by adjusting shaft power input and field excitation. In grid connected PV system, no prime mover or excitation source are present. Therefore, its scenario of load sharing is different and needs intensive analysis on the load sharing phenomena.

Droop control method and average power control method are the load sharing techniques developed in stand alone ac power system based on the power flow theory of an ac system [3],[4]. To guarantee proper performance of load-sharing under the wire impedance mismatches, voltage/current measurement error mismatches, and interconnection tie-line impedance effect, combined droop control and average power control method is proposed [5]. Also to ensure sharing of harmonic contents of the load currents, a harmonic droop sharing technique is proposed [5]. To determine the power that the inverter can handle, a criterion is proposed [6] to find reactive power which can avoid sophisticated detections of phase and magnitude of the fundamental component of a nonlinear load current.

In this research work, analysis is performed on the aspects of load sharing of a grid connected PV system and a strategy is proposed where the load sharing task can be undertaken controlling both modulation index and phase angle of the inverter. It is found that both real power and reactive var are affected upon change in modulation index and phase angle of the PV inverter. Analysis and simulation results are presented to demonstrate effectiveness of the proposed control technique.

II. Fundamental Building Block Diagram

A PV power inverter system in parallel with a load and utility grid is shown in Fig. 1. X_g represents sum of synchronous reactance of generator and reactance of transmission line.

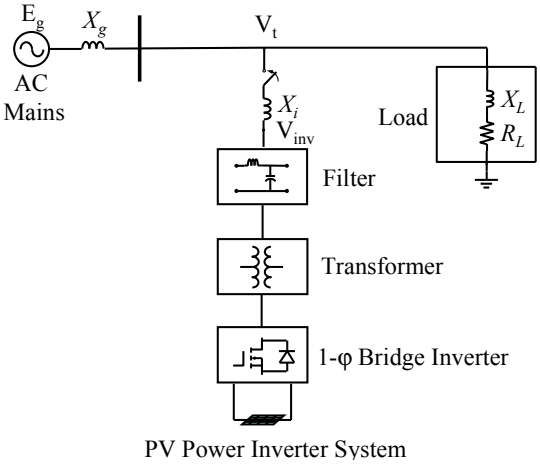


Fig. 1 Fundamental Building Block diagram of the grid-connected PV system

The dc voltage obtained from PV array is converted to ac through the inverter. The voltage is stepped up by transformer. An LC filter is connected to remove unwanted harmonics. The power is then fed to the load.

III. Principle of Load Sharing

The load sharing principle in case of parallel connected generators and grid connected PV system are described below separately.

A. In Parallel Connected Generators

Two parallel generators A and B shown in Fig. 2 supply a load of a certain power and power factor. Real and reactive power delivered from generators can be controlled by controlling the mechanical power input and excitation respectively.

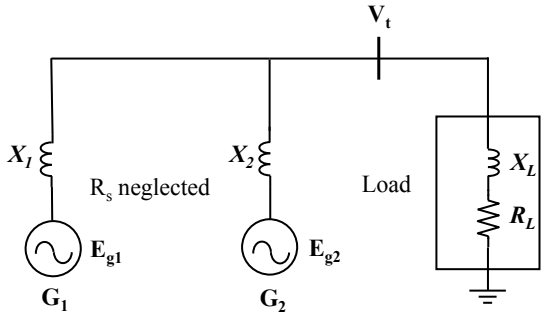


Fig. 2 Two parallel generators supplying a load

Control by changing mechanical power input

The input mechanical power or shaft power to the generator can be changed by changing the opening of the valves through which steam (or water) enters a turbine. Real power delivered by generator is given by

$$P = \frac{E_g . V_t}{X} \sin \delta \qquad (1)$$

If excitation I_f is kept constant, E_g will remain constant. Increasing the input shaft power will result the rotor speed start increasing. But rotor speed cannot exceed bus frequency. So δ will start to increase if V_t and X remain constant. The generator will start delivering more real power. Again the real power delivered is

$$P = V_t . I_a . \cos\theta \qquad (2)$$

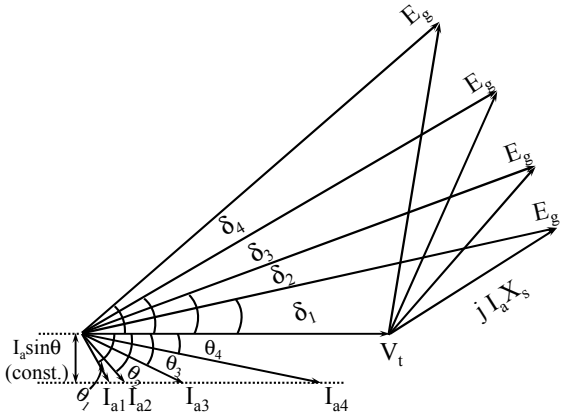


Fig. 3 Vector diagram of voltages and currents for change in shaft power input only in case of parallel generators

From the vector diagram shown in Fig. 3, we see that

$$E_g = V_t + jI_a X_s \qquad (3)$$

If δ is increased keeping E_g constant, $I_a X_s$ will increase. Since X_s is constant, I_a will increase. Again θ will decrease with the increase of δ. So $I_a \cos\theta$ will increase. Therefore, delivered real power $P = V_t . I_a . \cos\theta$ will increase. It is seen in the diagram that $I_a \sin\theta$ remains same with the increase of δ. So the reactive power $Q = V_t I_a \sin\theta$ will remain constant.

Control by changing excitation

Let us reduce excitation I_f keeping input shaft power constant. Since input shaft power is constant, delivered real power will be constant. It implies that $E_g \sin\delta$ will remain constant. A reduction in I_f must reduce E_g. So $\sin\delta$ increases i.e. δ increases. Since V_t is considered constant, $I_a \cos\theta$ must remain constant for constant output real power. From the vector diagram shown in Fig. 4 it is seen that $I_a \sin\theta$ varies with the change of excitation. So the reactive power $Q = V_t I_a \sin\theta$ varies. Therefore it can be inferred that changing excitation only varies reactive power if input shaft power is kept constant.

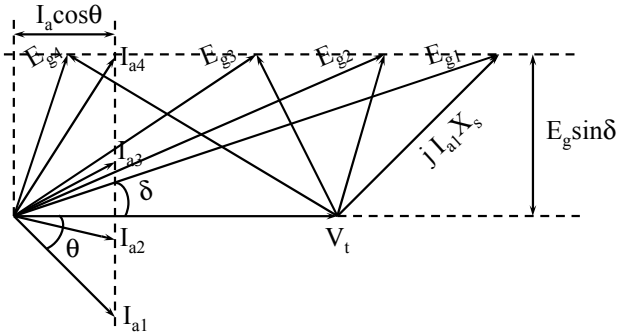


Fig. 4 Vector diagram of voltages and currents for change in excitation only in case of parallel generators

B. In grid-connected PV system

In the present grid-connected PV system scheme as shown in Fig. 5, power is fed to the load from the PV array. If the generation of power from PV array is not sufficient to serve the load demand, the deficit can be compensated from the grid. If the generated power is greater than the load demand, the excess power can be fed back to the grid.

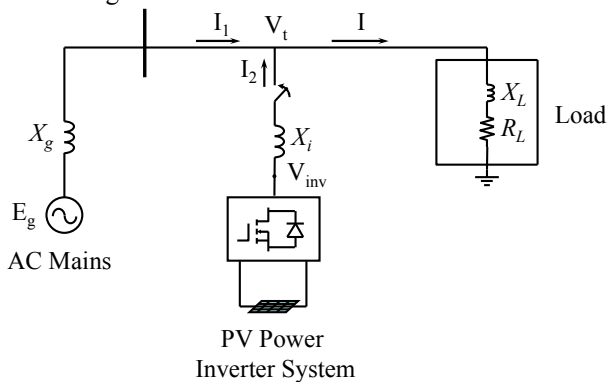


Fig. 5 Block diagram of the grid-connected PV system showing the sharing of load.

The vector diagram comprising the voltage and current vectors for the grid-connected PV system is shown in Fig. 6.

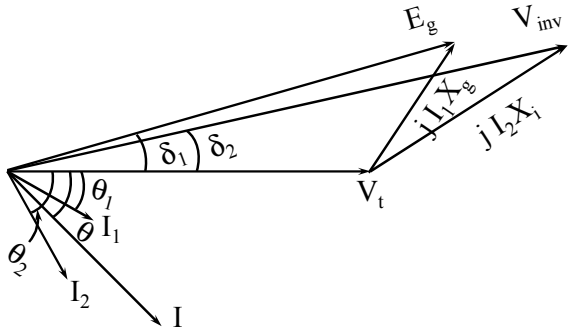


Fig. 6 Vector diagram for voltages and currents in grid-connected PV system

The active power supplied from the inverter side is

$$P = \frac{V_{inv} V_t}{X_i} \sin \delta \qquad (4)$$

And the reactive var supplied is

$$Q = \frac{|V_t|}{X_i} \left(|V_{inv}| \cos \delta - |V_t| \right) \qquad (5)$$

The control parameters to vary the real power and reactive var to be supplied from inverter side are modulation index and phase angle. The vector diagrams comprising voltages and currents may not be exactly same as those of parallel connected generators. This should be examined through simulation.

The maximum power that can be delivered from the inverter side is limited by the maximum power rating of the PV array. The losses in the switching devices and transformer should also be taken into account.

IV. Inverter branch: Equivalent circuit and analysis for voltages and currents

The PWM voltage is fed to the transformer and its output is then filtered. The filtered voltage is fed to the grid through an inductance X_i. The grid voltage is V_t whose variation must be within a specified range.

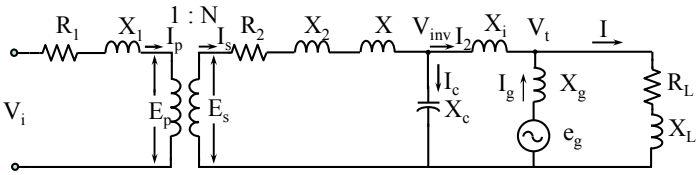


Fig. 7 Equivalent circuit of the grid-connected PV system

The transformer shown in Fig. 7 is expressed by its equivalent circuit. If all the parameters on the primary side are referred to the secondary then the above circuit takes the form shown in Fig. 8.

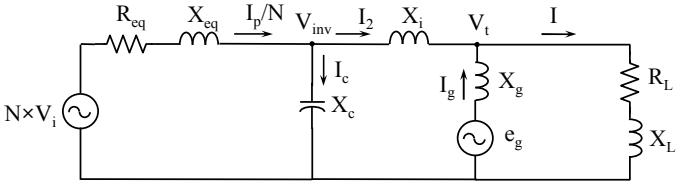


Fig. 8 Primary parameters of inverter branch of Fig. 8 referred to secondary

Here, $R_{eq} = R_1 N^2 + R_2$; $X_{eq} = X_1 N^2 + X_2 + X$

A. Formation of equation for voltages and currents to draw their vector diagrams and waveshapes

Applying superposition theorem in the circuit shown in Fig. 8, the equation for I_2, I_g and I can be developed.

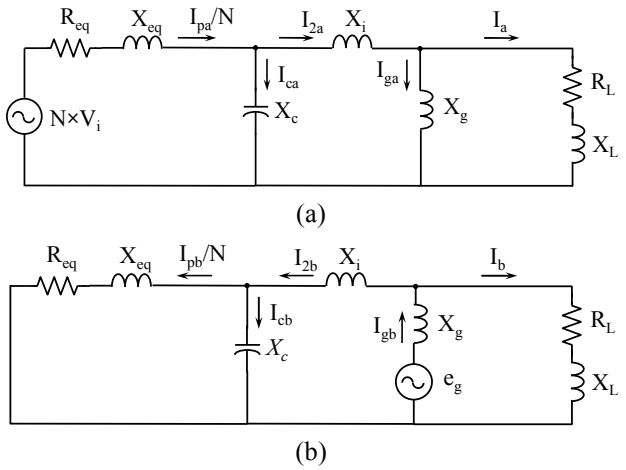


Fig. 9 Applying superposition theorem in the circuit of Fig.8, (a). When $N*V_i$ present only (b). When e_g present only

From Fig. 9(a),

$$\frac{I_{pa}}{N} = \frac{N \times V_i}{Z_{eq} + (-jX_c) \| (jX_i + Z_a)}$$

where, $Z_a = (jX_g) \| (R_L + jX_L)$

$$I_{2a} = \frac{I_{pa}}{N} \times \frac{-jX_c}{-jX_c + jX_i + Z_a}$$

$$I_{ga} = I_{2a} \times \frac{R_L + jX_L}{R_L + jX_L + jX_g} \; ; \; I_a = I_{2a} - I_{ga}$$

From Fig.9(b),

$$I_{gb} = \frac{e_g}{(R_L + jX_L) \| (Z_{aa} + jX_i) + jX_g}$$

where, $Z_{aa} = (-jX_c) \| (R_{eq} + jX_{eq})$

$$I_{2b} = I_{gb} \times \frac{R_L + jX_L}{R_L + jX_L + Z_{aa} + jX_i} ; \; I_b = I_{gb} - I_{2b}$$

$$I_2 = I_{2a} - I_{2b} \; ; \; I_g = I_{gb} - I_{ga} \; ; \; I = I_a + I_b$$

V. Simulation Results

The proposed scheme is simulated using Matlab and PSpice for various modulation indices and phase angles. Sine PWM (SPWM) technique is used for the generation of PWM patterns for switching signals of inverter. SPWM patterns are formed comparing a modulating sinusoidal wave with a high frequency triangular carrier wave.

The maximum power rating of the PV array used in the scheme is 1500 watt. The rated power, rated voltage and power factor of the load are 1000 watt, 240 V and 0.8 (lagging) respectively.

Control of real power

The modulation index and phase angles were adjusted in such a way that real power delivered from the inverter side varied while reactive var remained constant. The vector diagram of voltages and currents are shown in Fig. 10.

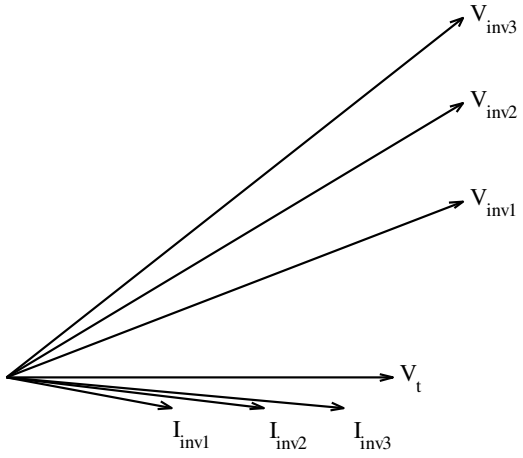


Fig. 10 Vector diagram showing voltages and currents when active power is controlled keeping reactive var constant (M_1 = 0.5, $Phase_1$ = 43.56^0; M_2 = 0.6, $Phase_2$ =54.55^0; M_3= 0.7, $Phase_3$ =61.4^0)

The data obtained from the simulation is given in Table 1

Table 1 Data for real power control

M	*Phase (deg)*	V_{inv} *(volt)*	δ_{inv} *(deg)*	I_{inv} *(amp)*	P_{inv} *(watt)*	Q_{inv} *(var)*
0.4	15	283.47	2.34	0.71	44.10	-165.4
0.5	43.56	325.88	29.64	2.66	615.43	-165.38
0.6	54.55	378.3	41.50	4.05	957.26	-165.66
0.7	61.4	433.93	49.27	5.28	1255.57	-165.06

From Table 1 it is seen that by increasing both M and Phase angle simultaneously, the real power delivered can be increased. The adjustments of M and Phase angle are done in such a way that real power delivered is varied but reactive power remained constant. From the vector diagram shown in Fig. 10 it is seen that the inverter output voltage is not constant. This differs with the parallel connected generators where generated voltage of generator remains constant while varying real power keeping reactive var constant.

Control of reactive power

The modulation index and phase angles were adjusted in such a way that reactive var delivered from inverter side varied while real power remained constant. The vector diagram of voltages and currents are shown in Fig. 11.

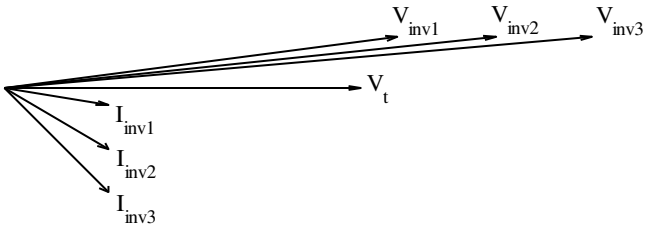


Fig. 11 Vector diagram showing voltages and currents when reactive var is controlled keeping active power constant (M_1 = 0.4, $Phase_1$ = 30.01^0; M_2 = 0.5, $Phase_2$ =24.93^0; M_3= 0.6, $Phase_3$ =21.63^0)

The data obtained from the simulation is given in Table 2.

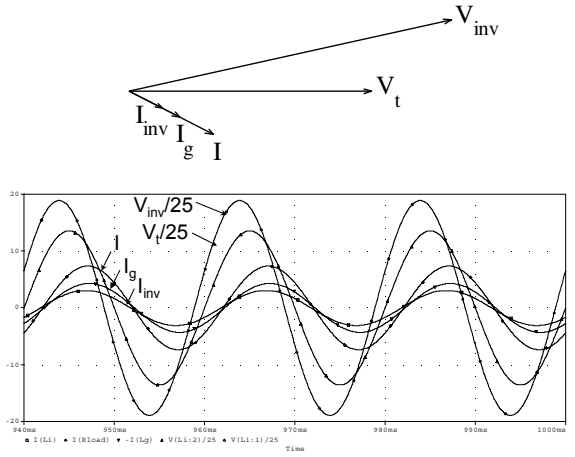
Table 2 Data for reactive power control

M	*Phase (deg)*	V_{inv} *(volt)*	δ_{inv} *(deg)*	I_{inv} *(amp)*	P_{inv} *(watt)*	Q_{inv} *(var)*
0.4	30.01	274.53	15.49	1.23	280.00	-94.07
0.5	24.93	339.06	12.49	1.86	280.14	-347.85
0.6	21.63	402.58	10.5	2.74	280.25	-595.30
0.7	19.3	465.57	9.06	3.69	280.17	-839.41
0.8	17.57	528.24	7.98	4.65	280.18	-1081.47
0.9	16.23	590.7	7.13	5.63	280.11	-1322.25

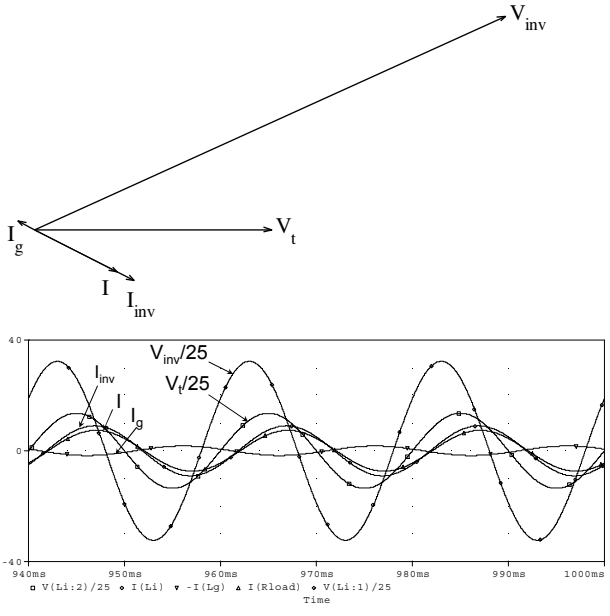
From Table 2 it is seen that by increasing M and decreasing phase angle simultaneously, the reactive var delivered can be increased. The adjustments of M and phase angle are done in such a way that reactive var delivered is varied but real power is kept constant. The vector diagram comprising voltages and currents shown in Fig. 11 is same as that of parallel connected generators shown in Fig. 4.

Control of real and reactive power maintaining currents in phase or 180^0 out of phase

Again the modulation indices and phase angles are adjusted in such a way that both real power and reactive var are controlled while the currents remain in phase or 180^0 out of phase. Fig. 12 shows vector diagrams and waveshapes for different M and phase angle.



(a)



(b)

Fig. 12 Vector diagrams of voltages and currents when currents are in phase or 180^0 out of phase (a). M = 0.5, Phase = 31.1^0 ; (b). M = 0.9, Phase = 43.9^0

The data obtained from the simulation is given in Table 3:

Table 3 Data for power control with currents in phase or 180^0 out of phase

M	*Phase (deg)*	V_{inv} *(volt)*	δ_{inv} *(deg)*	I_{inv} *(amp)*	P_{inv} *(watt)*	Q_{inv} *(var)*
0.4	23.4	278.72	9.65	0.93	178.46	-133.09
0.5	31.1	334.99	18.11	2.07	397.55	-299.62
0.6	36.0	392.87	23.81	3.16	605.77	-456.25
0.7	39.5	451.71	28.01	4.22	810.31	-606.53
0.8	42.0	511.30	31.11	5.25	1008.99	-755.47
0.9	43.9	571.38	33.50	6.27	1204.72	-903.23

From Fig. 12 and Table 3 it is seen that by increasing M and phase angle simultaneously both the real power and reactive var from the inverter side can be increased keeping the currents in phase or 180^0 out of phase. The vector diagram and waveshapes are in consistent in respect of both magnitude and phase angle.

In Table 1, Table 2 and Table 3 we can see that the active power delivered in all three cases are within the specified limit considering the maximum power rating of the PV array and the losses.

V. Conclusion

A control strategy for proper load sharing in the grid-connected PV system is designed in this research paper. It is different from that of conventional parallel connected generators since no prime mover or excitation source are present. The load sharing task can be performed controlling both modulation index and phase angle of the inverter. Sine PWM technique is adopted for generating switching signals for the inverter.

It is shown that by adjusting modulation index and phase angle, the real power delivered by the inverter branch can be adjusted keeping reactive var constant. The reactive var is also shown to be adjusted keeping real power constant. Again both the real power and reactive var are shown to be adjusted keeping the currents in phase or 180^0 out of phase. The PV power inverter can supply the total load or share a portion of it according to the demand. The excess power developed by the PV array can be fed back to the utility. The power delivered from the PV branch is shown to be within the specified limit defined by the maximum power rating of the PV array considering the losses.

References

[1]. C. J. Hatiziadoniu, F. Chalkiadakis, and V. Feiste, "A power conditioner for a grid-connected photovoltaic generator based on the 3-level inverter," *IEEE Trans. Energy Conv.*, vol. 14, no. 4, pp. 1605–1610, Dec. 1999.

[2] K. Y. Khouzam, "Technical and economic assessment of utility interactive PV systems for domestic applications in South East Queensland," *IEEE Trans. Energy Conv.*, vol. 14, no. 4, pp. 1544–1550, Dec. 1999.

[3] Chandorkar M.C., Divan M.D., Adapa R., "Control of parallel connected inverters in standalone ac supply systems" *IEEE Transactions on Industry Applications*, Vol. 29 No 1, 1993, pp. 136 –143

[4] Anil Tuladhar, Hua Jin, Tom Unger, and Konrad Mauch, "Control of parallel inverters in distributed ac power systems with consideration of line impedance effect," *IEEE Trans. Ind. Applicat.*, vol. 36, pp. 131–138, Jan./Feb. 2000.

[5] Marwali, M.N. Jin-Woo Jung Keyhani, "A. Control of distributed generation systems - Part II: Load sharing control," *IEEE Tran.,. Power Electron.,* Vol. 19, Issue. 6, pp. 1551-1561. Nov. 2004

[6] Wu, T.-F., Shen, C.-L., Chang, C.-H. Chang, and Chiu, J.-Y., 2003a, "A 1φ3W Grid-Connection PV Power Inverter with Partial Active Power Filter," *IEEE Transactions on Aerospace and Electronic Systems*,Vol. 39, No. 2, pp. 635-646, 2003

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Sensor-less Adaptive Rotor Parameter Estimation Method for Three Phase Induction Motor

Rajib Mikail, Kazi Mujibur Rahman

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET)
Dhaka-1000, Bangladesh
E-mail: rajib_mikail@eee.buet.ac.bd

Abstract – Accurate online parameter estimation is very important for inverter driven induction motor drives. Among the motor parameters the rotor is more prone to high percentage of parameter change. An adaptive rotor parameter estimation method is proposed including a detailed motor model. When the motor reaches steady state under constant load the stator current is monitored. The stator current of the motor is compared with the stator current of the model and the error is used recurrently to reestablish the model. The adaptation process stops when a threshold of the error value is reached. The estimated rotor resistance is compared with the original value. The response of the model with different change of parameters is also analyzed.

I. Introduction

It is expected that computational power will lead to realization of more powerful control techniques in future. It will be possible to identify the required parameters, decide on the control strategy and self-commission the drive. In case of direct field orientation control [1-2] the motor flux is calculated using measured voltages and currents with some known parameters of the machine. This scheme leads to poor accuracy because of the machine parameter dependency and the use of pure integration for the flux calculation and hence is seldom employed in industrial drives. The feed-forward slip control method [3] used in Rotor Flux Oriented Control also depends significantly on the rotor time constant *Tr* of the machine, which varies with changes in temperature and the load of the machine. The practical temperature excursion of the rotor is approximately 120°C above ambient. This increases the rotor resistances by 50 percent over its ambient or nominal value. Hence the calculated slip frequency is incorrect and the flux angle is no longer appropriate for field orientation. Spectral analysis techniques are based on the measured response to a deliberately injected test signal or an existing characteristic harmonic in the voltage/current spectrum. Observer based techniques, such as Extended Kalman filter [4-5], are computational intensive. Model reference adaptive control techniques [6-7] are usually operational in steady-states only and are disabled during transients. Application of observer based [8-10], artificial neural networks and fuzzy logic [11] for the on-line rotor time constant / rotor resistance adaptation is also a hot topic. But still the accuracy is an issue for researchers to work in this field. In this paper a sensor-less adaptive method is proposed having a dynamic fast response to parameter change both in steady state and transient state.

II. Induction motor model

An induction motor is composed of three stator windings and three rotor windings. The rotor windings move with respect to the stator windings like a transformer with a moving secondary. The machine model is described by differential equations with time varying mutual inductances. A three phase symmetrical induction machine can be represented as an equivalent two phase machine. The three phase stationary reference frame variables is converted to two phase stationary reference frame variables using the following equation

$$\begin{bmatrix} f_{qs}^{s} \\ f_{ds}^{s} \\ f_{0s}^{s} \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \cos\theta & \cos(\theta-120^{o}) & \cos(\theta+120^{o}) \\ \sin\theta & \sin(\theta-120^{o}) & \sin(\theta+120^{o}) \\ 0.5 & 0.5 & 0.5 \end{bmatrix}\begin{bmatrix} f_{as} \\ f_{bs} \\ f_{cs} \end{bmatrix} \quad ..(1)$$

Where f_{0s}^{s} is the zero sequence component and voltage, current and flux linkage is placed as variables. For the two phase equivalent machine if we consider the $d^{e}-q^{e}$ reference frame is rotating at synchronous speed then the model equations [12] become as follows

$$v_{ds} = R_s i_{ds} + \frac{d}{dt}\psi_{ds} - \omega_e \psi_{qs} \quad .....(2)$$

$$v_{qs} = R_s i_{qs} + \frac{d}{dt}\psi_{qs} + \omega_e \psi_{ds} \quad .....(3)$$

$$v_{dr} = R_r i_{dr} + \frac{d}{dt}\psi_{dr} - (\omega_e - \omega_r)\psi_{qr} \quad .....(4)$$

$$v_{qr} = R_r i_{qr} + \frac{d}{dt}\psi_{qr} + (\omega_e - \omega_r)\psi_{dr} \quad .....(5)$$

Flux linkage expressions are

$$\psi_{qs} = L_{ls} i_{qs} + L_m (i_{qs} + i_{qr}) \quad .....(6)$$

$$\psi_{qr} = L_{lr} i_{qr} + L_m (i_{qs} + i_{qr}) \quad .....(7)$$

$$\psi_{ds} = L_{ls} i_{ds} + L_m (i_{ds} + i_{dr}) \quad \text{.....(8)}$$

$$\psi_{dr} = L_{lr} i_{dr} + L_m (i_{ds} + i_{dr}) \quad \text{.....(9)}$$

Where L_s, L_m, ω_e, ω_r denote self-inductance, mutual inductance, synchronous speed, rotor electrical speed, respectively. The subscripts s and r stand for stator and rotor. The dynamic model equivalent circuits for synchronously rotating reference frame that satisfy the equations (2)-(5) is shown in figure 1.

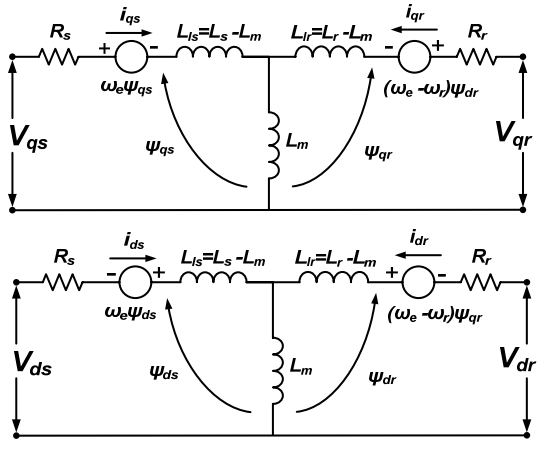


Fig. 1 The q^e-axis and d^e-axis equivalent circuits.

Based on the two-phase equivalent machine representation with two stator windings and two rotor windings, different choices of vector variables and reference frame (static or rotating) may be used. In this paper, stator currents, rotor currents, rotor speed and position are chosen as the states, and a fixed stator reference frame is used as the reference frame. As a result, the following model is obtained

$$v_{ds} = R_s i_{ds} + (L_{sl} + L_m)\frac{di_{ds}}{dt} + L_m \frac{di_{dr}}{dt} \quad \text{.....(10)}$$

$$v_{qs} = R_s i_{qs} + (L_{sl} + L_m)\frac{di_{qs}}{dt} + L_m \frac{di_{qr}}{dt} \quad \text{.....(11)}$$

$$v_{dr} = 0 = R_r i_{dr} + L_m \frac{di_{ds}}{dt} + (L_{rl} + L_m)\frac{di_{dr}}{dt} + \omega_m L_m i_{qs} + \omega_m (L_{rl} + L_m) i_{qr} \quad \text{.....(12)}$$

$$v_{qr} = 0 = R_r i_{qr} + L_m \frac{di_{qs}}{dt} + (L_{rl} + L_m)\frac{di_{qr}}{dt} - \omega_m L_m i_{ds} - \omega_m (L_{rl} + L_m) i_{dr} \quad \text{.....(13)}$$

$$T_e = \frac{3}{2}\left(\frac{p}{2}\right) L_m (i_{qs} i_{dr} - i_{ds} i_{qr}) \quad \text{.....(14)}$$

$$\frac{d\omega_m}{dt} = (T_e - T_L)/J - B_r \omega_m / J \quad \text{.....(15)}$$

$$\frac{d\theta}{dt} = \omega_m \quad \text{.....(16)}$$

Where ω_m, θ, T_L, T_e, J, and B_r denote rotor mechanical speed, rotor mechanical angle, the load torque, generated torque, the moment of inertia of the rotor and frictional damping, respectively.

III. Adaptation Process

To illustrate the adaptation process we need to rewrite the fixed stator reference frame model equations of (10)-(13) in matrix formulations:

$$V = RI + WLI + L\frac{dI}{dt} \quad \text{.....(17)}$$

Where,

$$V = \begin{bmatrix} v_{ds} \\ v_{qs} \\ 0 \\ 0 \end{bmatrix}, \; R = \begin{bmatrix} R_s & 0 & 0 & 0 \\ 0 & R_s & 0 & 0 \\ 0 & 0 & R_r & 0 \\ 0 & 0 & 0 & R_r \end{bmatrix}, \; I = \begin{bmatrix} i_{ds} \\ i_{qs} \\ i_{dr} \\ i_{qr} \end{bmatrix},$$

$$L = \begin{bmatrix} L_s & 0 & L_m & 0 \\ 0 & L_s & 0 & L_m \\ L_m & 0 & L_r & 0 \\ 0 & L_m & 0 & L_r \end{bmatrix}, \; \begin{matrix} L_s = L_{sl} + L_m \\ L_r = L_{rl} + L_m \end{matrix}$$

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_m \\ 0 & 0 & -\omega_m & 0 \end{bmatrix}$$

The flux linkage for fixed stator reference frame is calculated from the equation (18).

$$F = LI \quad \text{.....(18)}$$

Where, $F = \begin{bmatrix} \psi_{ds} \\ \psi_{qs} \\ \psi_{dr} \\ \psi_{qr} \end{bmatrix}$

The detail adaptation process is explained in fig.2. Here it is explained step by step.

Step 1

The three phase voltages and currents are sampled to the computer and a detailed adaptive model for the motor is established, where the nameplate parameters and the no load and blocked rotor test data are used. For the simulation the same model (14)-(17) is used to get the real stator current magnitude for a constant load.

Step 2

The stator currents for changing the rotor resistance to different percentage of increment and decrement are calculated by the motor model and stored to estimate the change of rotor resistance by the adaptation model later. These currents act as if the current is found from the actual motor.

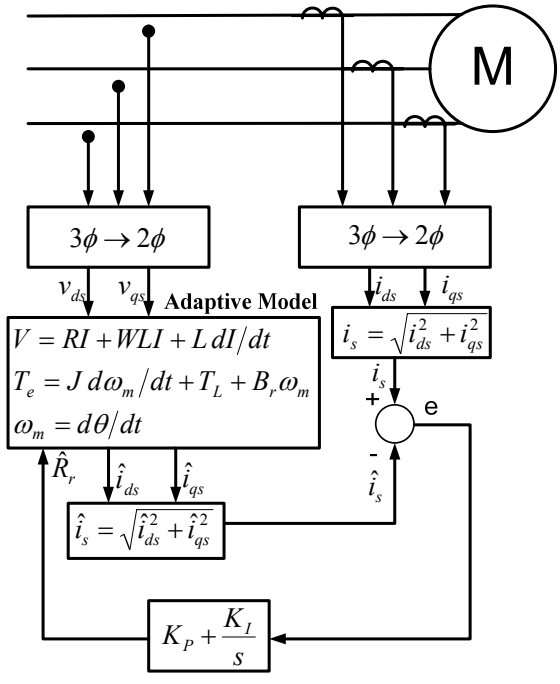


Fig. 2 Rotor resistance estimation by the adaptive scheme.

Step 3

The state model is solved numerically with runge-kutta method for a short period of time. The simulation starts from transient state with zero initial values for the variables. From the simulation the direct axis and quadrature axis stator current, flux, torque and speed is found as a result.

Step 4

When the model output goes to steady state then the simulated output current found in step 2 for changing rotor resistance of different percentage is applied to calculate the error in the adaptation process. Both the adaptive model and the real motor model gives the stator current in d-q reference frame from which the magnitude of the stator current is calculated as

$$i_s = \sqrt{i_{ds}^2 + i_{qs}^2} \quad \text{.....(19)}$$

The reference frame can be selected stator fixed, synchronously rotating or fixed at rotor. This output stator current from the adaptive model is compared with the real motor model and an error signal is generated.

Step 5

The three phase input voltages and the error are fed recurrently to the adaptive model to re-establish the motor parameters for the model.

$$\hat{R}_r = K_P(i_s - \hat{i}_s) + K_I \int (i_s - \hat{i}_s)dt \quad \text{.....(20)}$$

For re-establishing the model the error sign and magnitude plays a role and the rotor resistance is changed in a manner so that the error tends to a predefined threshold value.

Step 6

When the adaptive model gives almost the same output of the original motor the adaptation process ends.

The threshold value for the error is not fixed. For fast adaptation it can be changed to an optimal value which depends on the motor size and application. The PI block parameters are also selected according to the requirement. According to the required accuracy the parameters vary to a certain limit. If the accuracy is very concerned then the time for adaptation increases and the error threshold value and PI parameters are chosen accordingly.

Though in practical situation the change of rotor resistance will happen slowly we change the original value abruptly with a greater percentage and observe the response time and accuracy of the algorithm.

IV. Simulation Results

For the simulation the parameter of a three phase induction motor of 1hp is used. The parameters for the motor are given in table 1.

Table 1 Motor parameters

Description	Parameter	Value	Units
Stator inductance	L_s	$2.5\text{x}10^{-3}$	H
Rotor inductance	L_r	$2.5\text{x}10^{-3}$	H
Mutual inductance	L_m	$90\text{x}10^{-3}$	H
Stator Resistance	R_s	0.6	Ω
Rotor Resistance	R_r	0.4	Ω
Moment of inertia	J	0.15	Kgm^2
Frictional coefficient	B_r	0.0	Nms
Torque	T_L	10	Nm
Stator voltage	v	220	V
Frequency	f	50	Hz
No. of poles	p	4	-

At first the model is simulated for a constant load from transient to steady state. Different reference frames are used to validate the model output. In figure 3 the generated torque is plotted against speed. This same result is found for each reference frame.

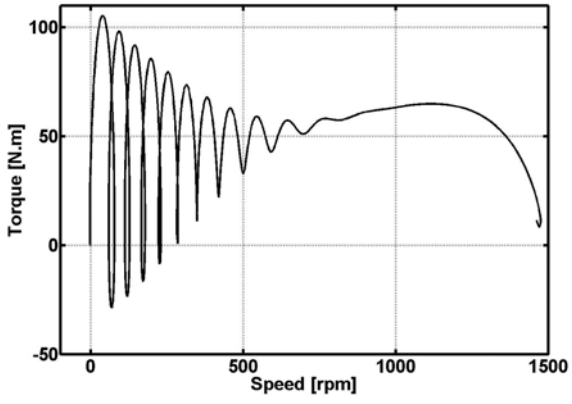


Fig. 3 The generated torque for a constant load torque for the proposed model from transient to steady state.

In figure 4 the torque is plotted against time for the same situation. Here also all the reference frame give the same output.

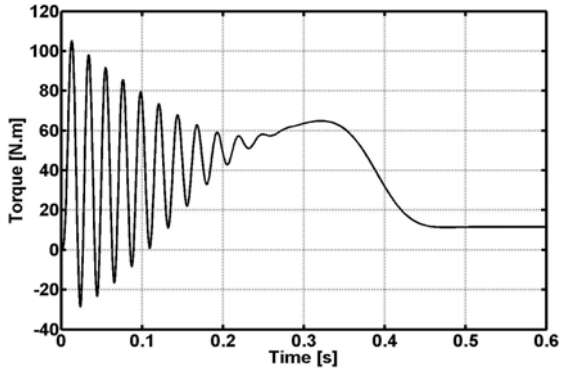


Fig. 4 The generated torque vs. time for a constant load.

In figure 5 the direct axis and quadrature axis current for synchronously rotating reference frame.

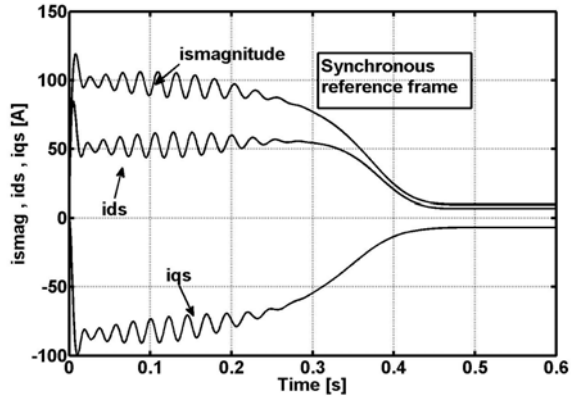


Fig. 5 The q^e-axix and d^e-axis stator currents with the magnitude of that.

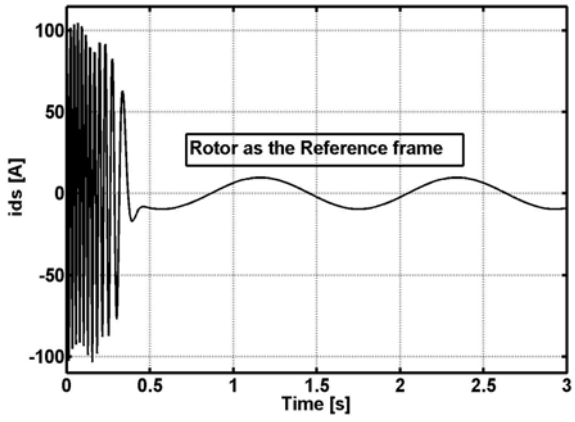


Fig. 6 The rotor reference frame direct axis current in the transient state to steady state.

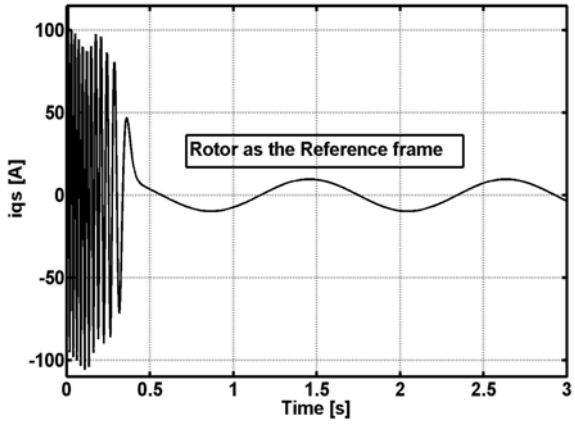


Fig. 7 The rotor reference frame quadrature axis current in the transient state to steady state.

In figure 6 and 7 direct axis and quadrature axis stator currents are shown for rotor reference frame. Figure 8 and 9 shows the stator reference frame d-q stator currents. The magnitude of stator current calculated from any reference frame is the same and shown in figure 5.

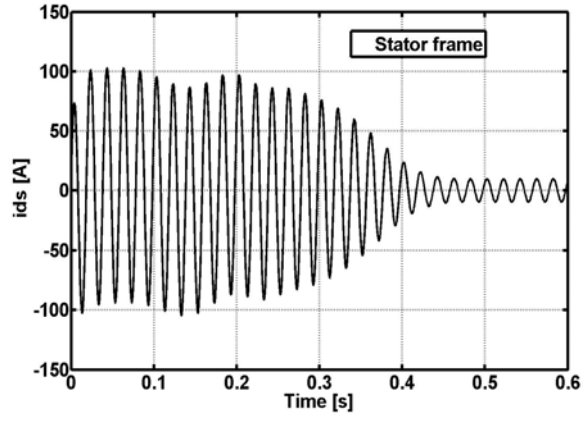


Fig. 8 The stator reference frame direct axis current in the transient state to steady state.

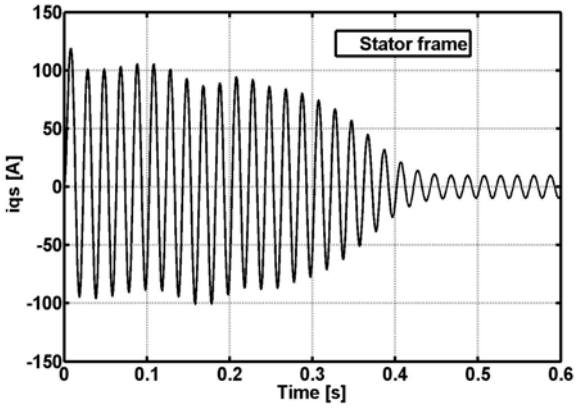


Fig. 9 The stator reference frame quadrature axis current in the transient state to steady state.

The flux is also calculated from (18) and shown in figure 10. Here the direct axis and quadrature axis stator flux linkages are shown for synchronously rotating reference frame.

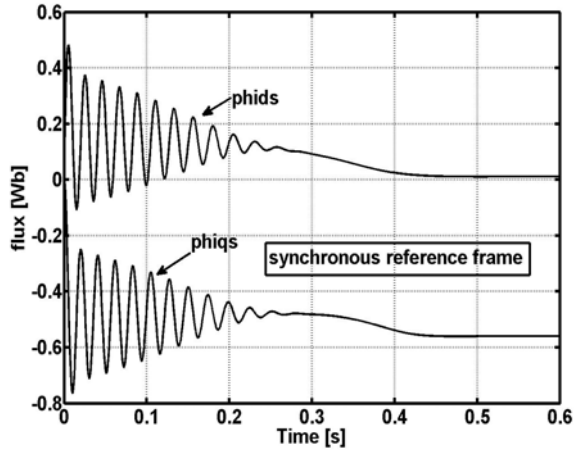


Fig. 10 The direct axis and quadrature axis flux linkages for synchronously rotating reference frame.

To observe the change of speed due to rotor resistance change, the model is simulated for 50% increase of rotor resistance. The result is shown in figure 11. From the result it is found that the speed of the motor running at constant load changes from 1475 rpm to 1462 rpm. It is obvious from the result that the calculated speed could be different if the change in rotor resistance is not estimated properly. For the inverter control circuit it is needed to estimate the speed for proper control of flux and generated torque. So for applications where speed is not sensed by any physical sensor it must be estimated by proper modelling and estimation. The adaptive model developed here can calculate the speed along with other variables.

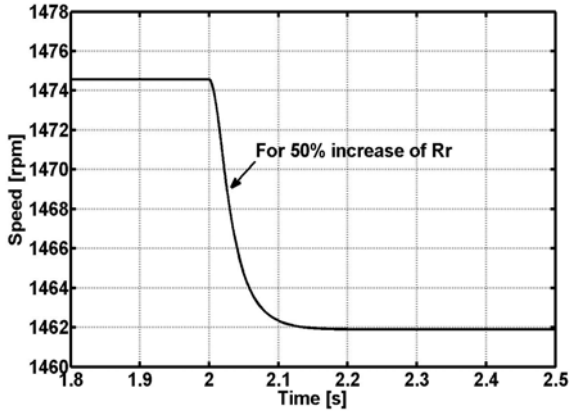


Fig. 11 Change of speed due to change of rotor resistance.

The adaptive algorithm was applied for 50% increase of rotor resistance and 20% decrease of rotor resistance from initial value. The estimated rotor resistance along with the original resistance is shown in figure 12. It is found that the estimator can track the change of resistance with good precision.

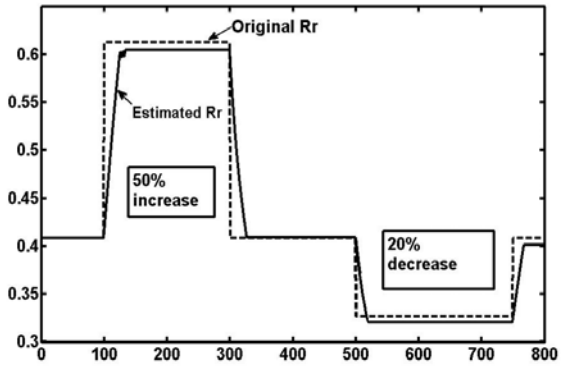


Fig. 12 Estimated rotor resistance is plotted with the original rotor resistance for 50% increase and 20% decrease.

The percentage of error is shown in figure 13. It is found that the error percentage is below 2%. The error percentage for increasing the resistance is lower than that of decreasing. As the resistance usually increase due to skin effect and rise of temperature, it is a good sign for the estimator. The time to reach the original value is also a concern for the controller circuit. If the estimator can not track with the rate of change of the resistance then it may lose the track of flux and the motor may go beyond control. So the time to reach the original value is shown in figure 14. Here the time is calculated for sudden 50% and 20% change. Though in practical situation the rotor resistance will not change so quickly, it is found from the result that the estimator can track the change within acceptable time.

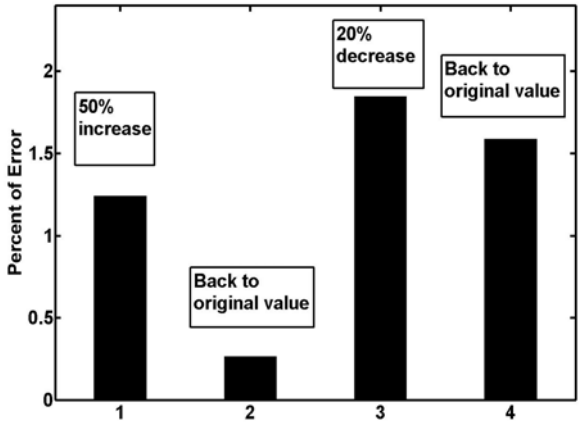


Fig. 13 The percentage of error of the estimated value is shown for different change of original rotor resistance.

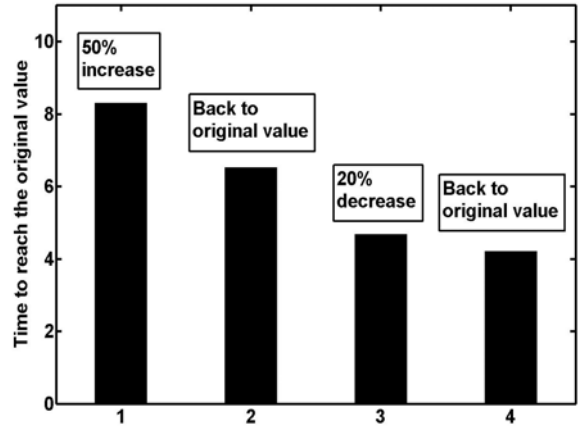


Fig. 14 Time to reach the original value for the proposed model.

V. Conclusion

As the rotor circuit faces different slip frequency at different load and speed conditions, the rotor resistance changes due to skin effect which is dependant on operation frequency. The rise of temperature of the rotor also causes the resistance to change. Due to all these reasons rotor resistance change by significant percentage over its nominal value. If the parameter is not estimated correctly the flux level cannot be correctly maintained, resulting torque will be out of track, the motor will take more current than necessary. Tracking and adapting the rotor parameter within the controller in real time can solve these problems. The proposed estimator is simple and easy to implement. The results from zero speed to steady state are analyzed in different reference frame. The tracking performance of the proposed estimator is analyzed and the result from figure 12, 13 and 14 vindicate that the estimator can dynamically track the change with greater accuracy and smaller response time.

References

[1] F. Blaschke, “The principle of field orientation as applied to the new transvector closed loop system for rotating field machines,” Siemens Review, vol. 34, pp. 217–220, May 1972.

[2] H. Kubota, K. Matsuse, "Speed Sensorless Field-Oriented Control of Induction Motor with Rotor Resistance Adaptation", IEEE Trans. on Industry Applications, Vol. 30, No. 5, Sept/Oct 1994, pp. 1219-1224, Sept/Oct 1994.

[3] T. Matsuo, T. A. Lipo, “A Rotor Parameter Identification Scheme for Vector-Controlled Induction Motor Drives”, IEEE Proceedings on Industry Applications, Vol. IA-21, No. 4, May/June, 1985, pp. 624-632.

[4] L. Guo, "Estimating Time-varying Parameters by Kalman Filter Based Algorithms: Stability and Convergence", IEEE Trans. on Automatic Control, 1990, Vol. 35, pp. 141-147.

[5] L. Loron and G. Laliberte, "Application of the Extended Kalman Filter to Parameter Estimation of Induction Motors", Conf. Proceedings EPE'93, Brighton, U.K., 1993, pp. 85-90.

[6] C. Attaianese, G. Tomasso, et al, "A Novel Approach to Speed and Parameters Estimation in Induction Motor Drives", IEEE Trans. on Energy Conversion, Vol. 14, No. 4, Dec. 1999, pp. 939-945.

[7] K. Ohnishi, Y. Ueda, and K. Miyachi, “Model Reference Adaptive System Against Rotor Resistance Variation in Induction Motor Drive”, IEEE Trans. Ind. Electron., vol. IE- 33, pp. 217-223, June 1986.

[8] Nilsen R, Kazmierkowski MP. “Reduced-order observer with parameter adaptation for fast rotor flux estimation in induction machines”, IEEE Proceedings 1989; 136 (Pt. D) (1): 35-43.

[9] Verghese GC, Sanders SR. “Observers for flux estimation in induction machines”, IEEE Trans. Ind. Electron. 1988; 35(1): 85-94.

[10] T. Orlowska-Kowalska, "Application of Extended Luenberger Observer for Flux and Rotor Time-constant Estimation in Induction Motor Drives", IEE Proceedings on Control Theory and Applications, Vol. 136, No. 6, Nov 1989, pp. 324-330.

[11] B. Karanayil, M.F. Rahman and C. Grantham, “Stator and Rotor Resistance Observers for Induction Motor Drive Using Fuzzy Logic and Artificial Neural Networks”, IEEE Trans. Energy Conversion, 25 October 2004.

[12] C. M. Ong, Dynamic Simulation of Electric Machinery, McGraw-Hill, New York, 1986.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Measured Impedance by Distance Relay Elements in a Single Phase to Ground Fault

H. Shateri and S. Jamali

Centre of Excellence for Power Systems Automation and Operation
Department of Electrical Engineering, Iran University of Science and Technology (IUST), Narmak 16846, Tehran, Iran
shateri@iust.ac.ir and sjamali@iust.ac.ir

Abstract - Distance relays are one of the mostly used protective devices in the transmission systems. These relays are used as the main and backup protective device. Fault resistance is a source of error for distance relays in the form of the measured impedance deviation from its actual value. This paper discusses the measured impedance by six elements of distance relay first zone in the case of a single phase to ground (AG) fault. This is done by presenting the measured impedance at the relaying point and the distance relay ideal tripping characteristic for each element. The variation of the ideal tripping characteristic due to the changes in the power system parameters is investigated.

I. Introduction

Frequency variation, ground fault resistance in the single phase to ground faults, and power swing are some of the phenomena adversely affect the distance protection performance. Among these, frequency variation and power swing relate to dynamic states of power systems. The ground fault resistance depends on the state of the fault arc creation between the ground and the faulted phase of the transmission line and the ground path [1].

Many efforts have studied the measured impedance at the relaying point, the measured impedance by the element corresponding to the faulted phase to ground, for single phase to ground faults [2]-[5]. But, no attention has been paid to the other elements of the distance relay.

Ground fault resistance is an unknown phenomenon causing distance relay mal-operation. Reference [2] presents an adaptive distance protection method to overcome this problem. Due to unknown magnitude of the ground fault resistance, this method defines an operational characteristic, which is robust against the variations of the ground fault resistance and does not mal-operate.

References [3]-[5] discuss different aspects of adaptive distance protection. For example, [4] suggests a fixed quadrilateral characteristic for operational characteristic of distance relays, regarding to the mentioned characteristic in [2]. Reference [5] utilizes an artificial intelligence technique to determine the boundaries of the characteristic presented in [2].

The fault resistance not only is an unknown quantity from protective system point of view, but this resistance could be measured as an inductive or capacitive impedance, depending on the far end infeed and the line pre-fault loading. Variation of the imposed impedance due to the presence of the ground fault resistance from its resistive feature depends on both the structural and operational conditions of the system.

This paper discusses the measured impedance by the six elements of a distance relay in the case of a single phase to ground (AG) fault. This is done by presenting the measured impedance and the ideal tripping characteristic for each element. In addition, the variation of the measured impedance due to changes in power system conditions is investigated.

II. Measured Impedance at Relaying Point

Distance relays operate based on the measured impedance at the relaying point. In the case of zero fault resistance, the measured impedance at the relaying point by element AG in the case of phase A to ground fault only depends on the length of the line section between the fault and the relaying points. According to Fig. 1 this impedance is equal to pZ_{1L} where p is the per-unit length of the line section between the relaying and the fault points, and Z_{1L} is the line positive sequence impedance in ohms.

In the case of a non-zero fault resistance, the measured impedance at the relaying point is not equal to the impedance of the line section between the fault and relaying points. In this case, the structural and operational conditions of the power system affect the measured impedance. The structural conditions are evaluated by short circuit levels at the line ends, S_{SA} and S_{SB}. The operational conditions prior to the fault instance can be represented by the load angle of the line, δ, and the ratio of the magnitude of the line end voltages, h, or $E_B / E_A = he^{-j\delta}$. With respect to Fig. 1 and Fig. 2, the measured impedance by each element and the measured impedance by distance relay before the fault occurrence can be expressed by the following equation. More detailed calculations for the measured impedance by element AG in the case of AG fault can be found in [2].

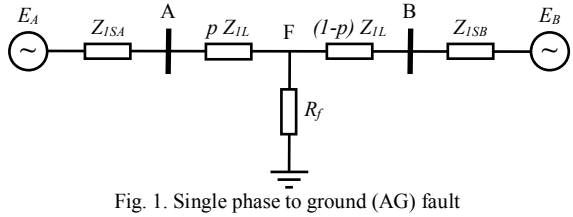


Fig. 1. Single phase to ground (AG) fault

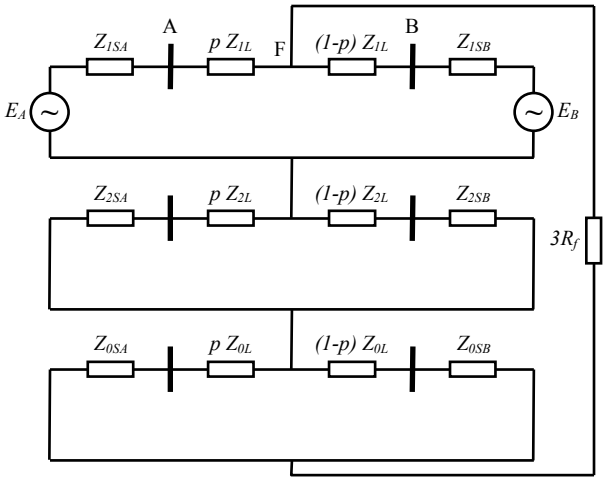


Fig. 2. Equivalent circuit of phase A to ground fault

The measured impedance before the fault occurrence by the all six elements of the distance relay is:

$$Z_{NF} = \frac{Z_{1SA}\, he^{-j\delta} + Z_{1L} + Z_{1SB}}{1 - he^{-j\delta}} \tag{1}$$

On the other hand, after fault occurrence the measured impedance by the distance elements can be expressed by the following equations:

$$a = 1\angle 120° \tag{2}$$

$$Z_{1A} = Z_{1SA} + p\,Z_{1L} \tag{3}$$

$$Z_{1B} = Z_{1SB} + (1-p)Z_{1L} \tag{4}$$

$$Z_1 = \frac{Z_{1A}Z_{1B}}{Z_{1A} + Z_{1B}} \tag{5}$$

$$Z_{0A} = Z_{0SA} + p\,Z_{0L} \tag{6}$$

$$Z_{0B} = Z_{0SB} + (1-p)Z_{0L} \tag{7}$$

$$Z_0 = \frac{Z_{0A}Z_{0B}}{Z_{0A} + Z_{0B}} \tag{8}$$

$$Z_\Sigma = 2Z_1 + Z_0 \tag{9}$$

$$C_1 = \frac{Z_{1B}}{Z_{1A} + Z_{1B}} \tag{10}$$

$$C_0 = \frac{Z_{0B}}{Z_{0A} + Z_{0B}} \tag{11}$$

$$K_{0L} = \frac{Z_{0L} - Z_{1L}}{3Z_{1L}} \tag{12}$$

$$K_{ld} = \frac{1 - he^{-j\delta}}{Z_{1A}he^{-j\delta} + Z_{1B}} \tag{13}$$

$$C_{ld} = (Z_\Sigma + 3R_f)K_{ld} \tag{14}$$

The measured impedance by element AG is:

$$Z_{\underset{AG}{A}} = p\,Z_{1L} + \frac{3R_f}{C_{ld} + 2C_1 + C_0(1+3K_{0L})} \tag{15}$$

The measured impedance by element BG is:

$$Z_{\underset{BG}{A}} = p\,Z_{1L} + \frac{[(2+a)Z_1 + (2+a^2)Z_0] + 3R_f}{C_{ld} + a[C_0(1+3K_{0L}) - C_1]} \tag{16}$$

The measured impedance by element CG is:

$$Z_{\underset{CG}{A}} = p\,Z_{1L} + \frac{[(2+a^2)Z_1 + (2+a)Z_0] + 3R_f}{C_{ld} + a^2[C_0(1+3K_{0L}) - C_1]} \tag{17}$$

The measured impedance by element AB is:

$$Z_{\underset{AB}{A}} = p\,Z_{1L} + \frac{[-a^2Z_1 - Z_0] + 3R_f}{C_{ld} + 3C_1/(2+a)} \tag{18}$$

The measured impedance by element BC is:

$$Z_{\underset{BC}{A}} = p\,Z_{1L} + \frac{[2Z_1 + Z_0] + 3R_f}{C_{ld}} \tag{19}$$

which is also equal to:

$$Z_{\underset{BC}{A}} = \frac{Z_{1SA}\, he^{-j\delta} + Z_{1L} + Z_{1SB}}{1 - he^{-j\delta}} \tag{20}$$

The measured impedance by element CA is:

$$Z_{\underset{CA}{A}} = p\,Z_{1L} + \frac{[-aZ_1 - Z_0] + 3R_f}{C_{ld} + 3C_1/(2+a^2)} \tag{21}$$

It can be seen that the measured impedance in the case of element BC is unchanged, in the case of element AG is equal to the impedance of the line section between the fault and the relaying point for zero fault resistance; otherwise deviated from its actual value depending on the fault resistance and the power system conditions. The other four elements (BG, CG, AB, and CA) are affected by fault occurrence, but not as seriously as element AG. In the case of these elements, in addition to the fault resistance, there is another deviating term.

III. Ideal Tripping Characteristic

Knowing the structural and operational conditions, distance relay ideal tripping characteristic can be defined, in the case of each distance relay element. This characteristic has four boundaries. First boundary is the measured impedance for zero fault resistance; fault location varies from near end up to the far end of the line. In the second boundary, the fault location is at the far end; fault resistance varies between 0 and 200 ohms. Third boundary is the result of the fault point variation along the line for the fault resistance of 200 ohms. Forth is achieved by variation of the fault resistance between 0 and 200 ohms for the faults on the near end of the line. Setting distance relay operational characteristic to this characteristic, the relay would not mal-operate.

The tripping characteristic of a distance relay is presented for a practical system. A 400 kV Iranian transmission line with the length of 300 km has been used for this study. By utilizing the Electro-Magnetic Transient Program (EMTP) [6] various sequence impedances of the line are evaluated according to its physical dimensions. The calculated impedances are:

R_{1L} = 0.01133 Ω/km
X_{1L} = 0.3037 Ω/km
R_{0L} = 0.1535 Ω/km
X_{0L} = 1.1478 Ω/km

Fig. 3 shows the ideal tripping characteristics for six elements of the distance relay, when the short circuit levels at Buses A and B are 10 and 20 GVA, the voltage ratio is 0.96, and the load angle is equal to 16°.

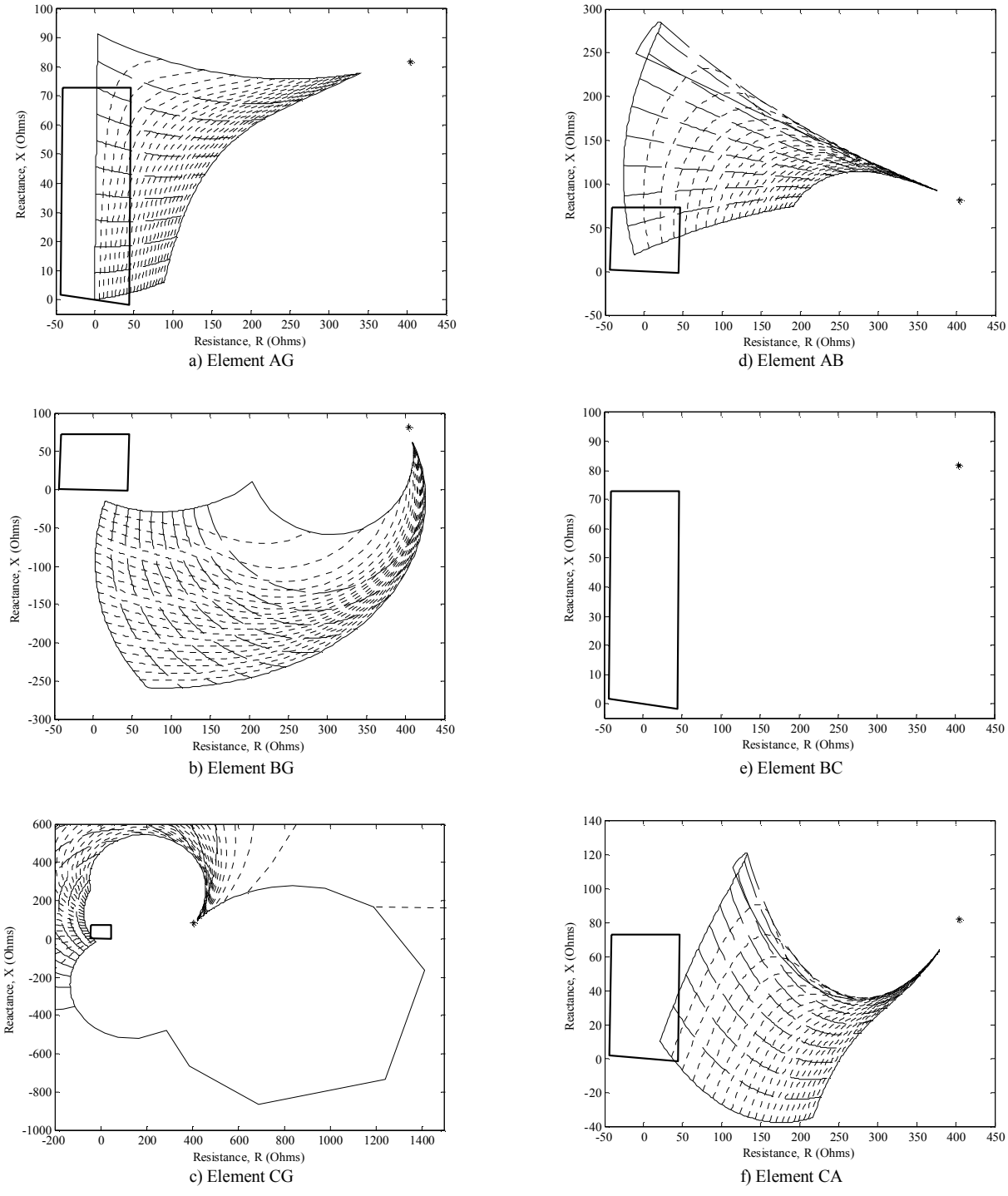


Fig. 3. Distance relay ideal tripping characteristics

In Fig. 3, in addition to the boundaries of the tripping characteristic which are shown by full lines, the measured impedance when the fault point varies from 0.1 to 0.9 of the line length in steps of 0.1 as the fault resistance varies from 0 to 200 ohms is shown with dashed lines; furthermore, the measured impedance when the fault resistance varies from 10 to 190 ohms in steps of 10 ohms as the fault position varies from near end up to the far end of the line is shown with dotted lines. The star at the upper right side is the measured impedance before fault occurrence, which is the same for the all six elements. The distance relay quadrilateral characteristic set to 80% of the line length is also plotted in Fig. 3.

In Fig. 3a, element AG, it can be seen that in the case of zero fault resistance the measured impedance is equal to the impedance of the line section between the fault and the relaying points and for non-zero fault resistances, the measured impedance deviates from the mentioned value. As it is expected, this element would operate for the reasonable magnitudes of the fault resistance.

In Fig. 3b, element BG, it can be seen that there is no overlapping region between the tripping and quadrilateral characteristics; therefore, this element would not operate.

In Fig. 3c, element CG, the tripping characteristic is placed outside the boundaries, unlike the other cases. Here, there is a small overlapping region between the

tripping and quadrilateral characteristics for the faults close to the near end of the line and the low magnitudes of the fault resistance. Therefore, the probability of this element operation is low.

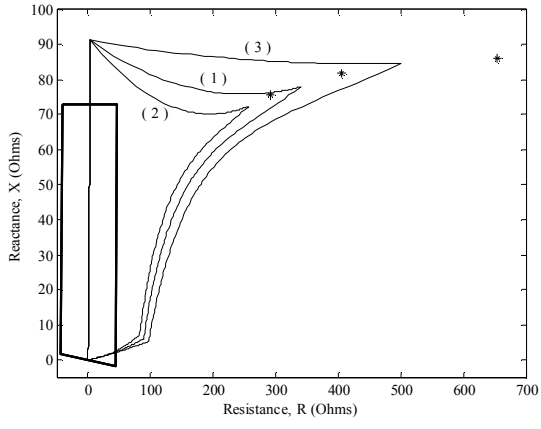
In Fig. 3d, element AB, there is a considerable overlapping region between the tripping and quadrilateral characteristics for the faults close to the near end of the line and the low and medium magnitudes of the fault resistance. Therefore, the probability of this element operation is not high, but not low either.

In Fig. 3e, element BC, as it is expected the measured impedance is not affected by the fault occurrence; and the measured impedances before and after the fault occurrence are the same. Therefore, this element would not operate.

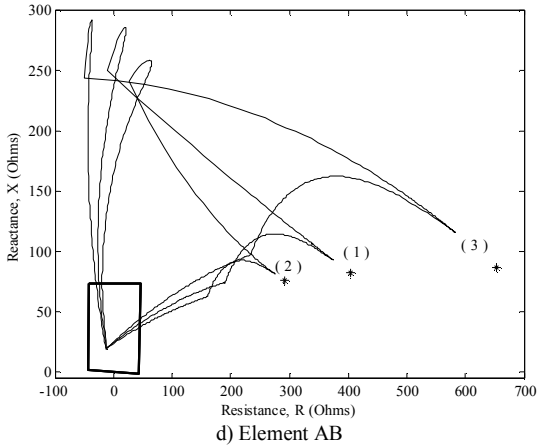
In Fig. 3f, element CA, there is a small overlapping region between the tripping and quadrilateral characteristics for the faults close to the near end of the line and the low magnitudes of the fault resistance. Therefore, the probability of this element operation is low.

IV. Ideal Tripping Characteristic Variation due to Changes in Power System Conditions

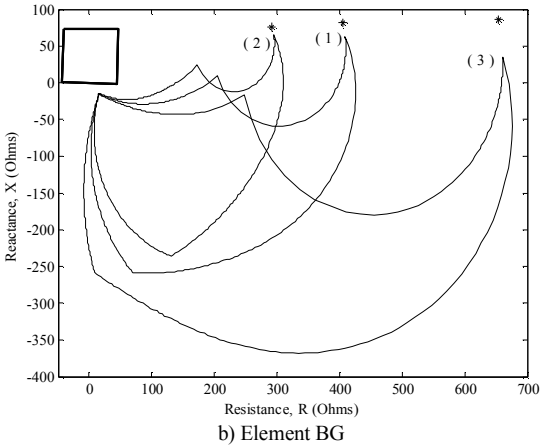
Once the operational conditions of a power system has changed, for instance, due to load level changes, the load angle and the voltage ratio would vary. Otherwise, when a switching takes place in the network that changes the topology of the network, short circuit levels at the line



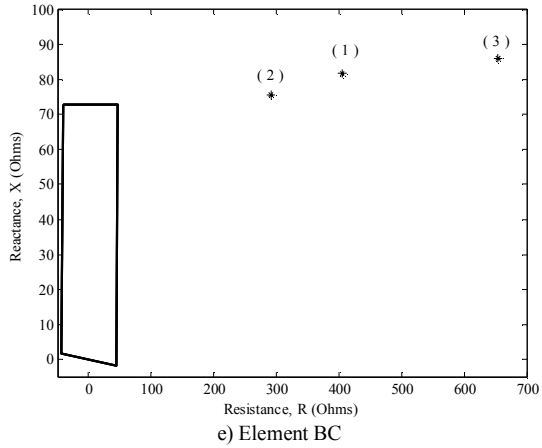
a) Element AG



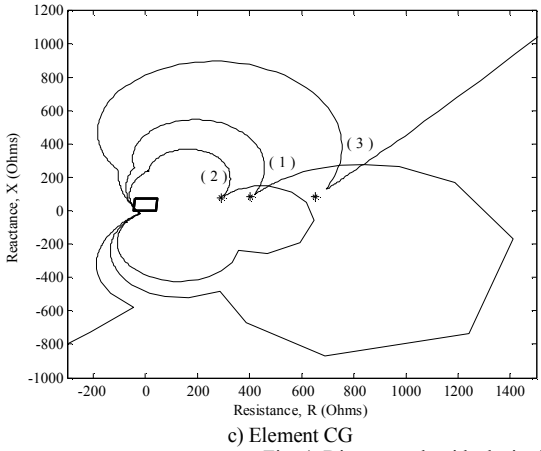
d) Element AB



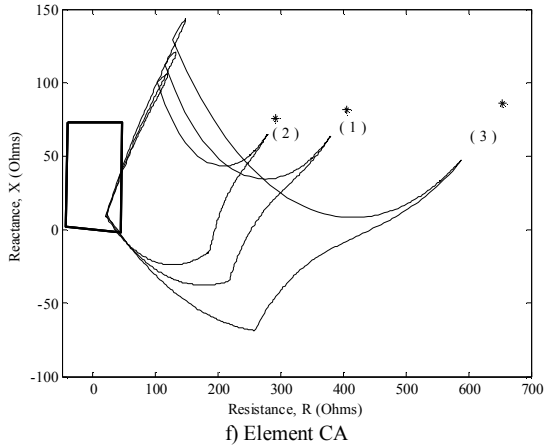
b) Element BG



e) Element BC



c) Element CG



f) Element CA

Fig. 4. Distance relay ideal tripping characteristics, operational conditions variation

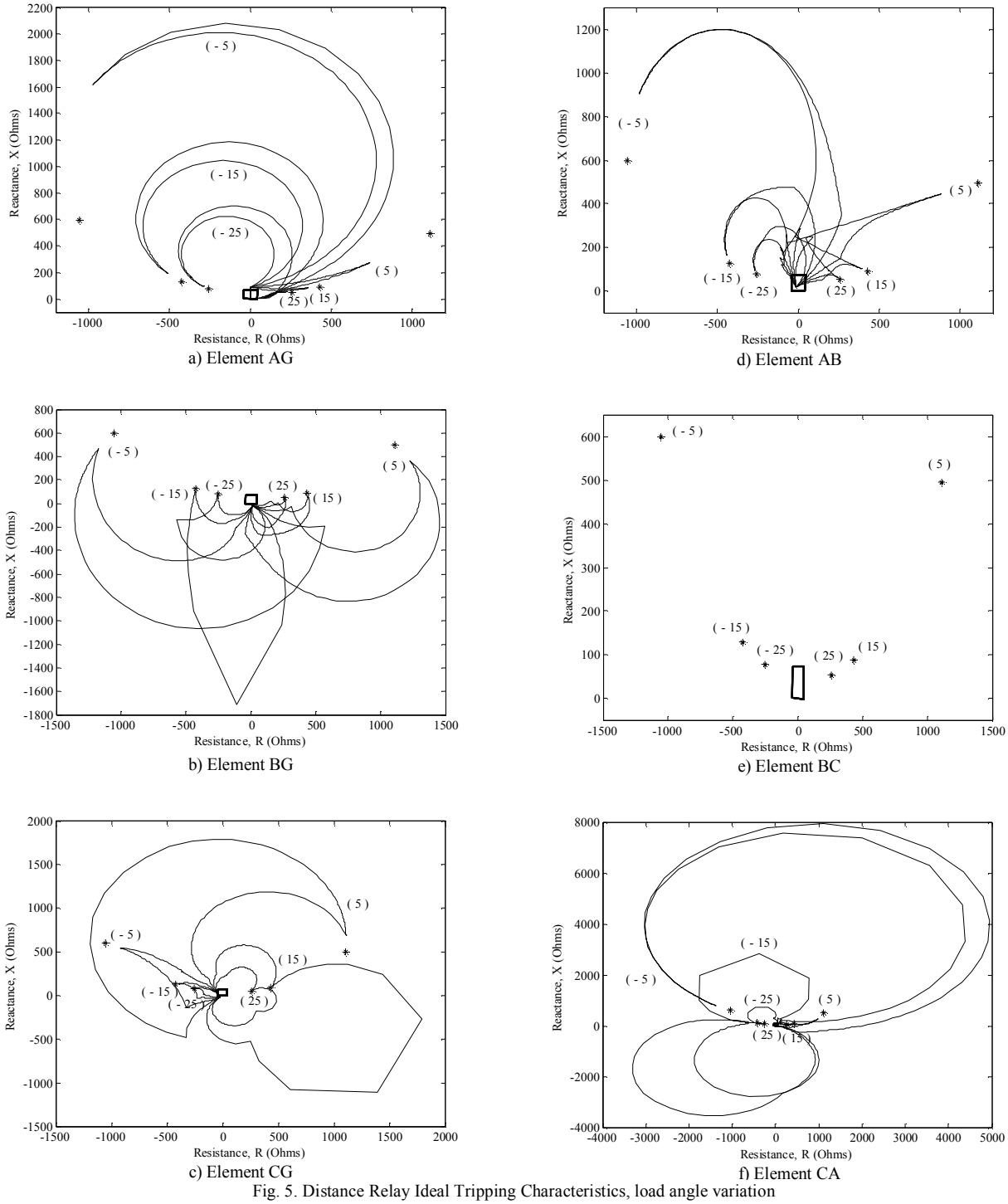
ends would also vary. As operational or structural conditions of the network vary, the measured impedance and consequently the ideal tripping characteristic would change.

Fig. 4 shows the distance relay tripping characteristic in the case of the six elements for three different power system operational conditions, with the same structural conditions as Fig. 3. Curve (1) in each sub-figure is the tripping characteristic of Fig. 3. When the load angle increases from 16° to 22°; and the voltage ratio decreases from 0.96 to 0.94, Curve (2) would be the result. Otherwise, when the load level decreases, Curve (3) is resulted for $\delta = 10°$ and $h = 0.98$. For better observation, the dashed and dotted lines are omitted.

It can be seen that as the power system operational conditions change, the tripping characteristics for all of the distance relay elements, with exception of element BC, varies considerably.

Usually load angle and voltage ratio have a close relation with each other, knowing power factor is usually constant. Here, the effect of the load angle variation on the measured impedance at the relaying point is discussed individually.

Fig. 5 shows the effect of the load angle variation on the tripping characteristics of the distance relay elements. Here, the load angle takes the magnitudes 25°, 15°, 5°, – 5°, – 15°, and – 25°. The other parameters are just the same as Fig. 3.



a) Element AG

d) Element AB

b) Element BG

e) Element BC

c) Element CG

f) Element CA

Fig. 5. Distance Relay Ideal Tripping Characteristics, load angle variation

It can be seen that as the load angle varies, the tripping characteristics for all of the distance relay elements, with exception of element BC, changes considerably. In the case of the negative load angles, the tripping characteristic expands considerably for the most of the elements.

V. Conclusion

This paper presents the measured impedance at the relaying point by the six elements of a distance relay in the case of the single phase to ground (AG) fault. The ideal tripping characteristic is presented for the six elements. In addition, the variation of the measured impedance due to changes in power system conditions is investigated.

In the case of element AG, in the case of zero fault resistance the measured impedance is equal to the impedance of the line section between the fault and the relaying points; and for non-zero fault resistances, the measured impedance deviates from the mentioned value.

In the case of element BG, there is no overlapping region between the tripping and the quadrilateral characteristics.

In the case of element CG, the tripping characteristic is placed outside the boundaries, unlike the other cases and there is a small overlapping region between the tripping and quadrilateral characteristics for the faults close to the near end of the line and the low magnitudes of the fault resistance.

In the case of element AB, there is a considerable overlapping region between the tripping and quadrilateral characteristics for the faults close to the near end of the line and the low and medium magnitudes of the fault resistance.

In the case of element BC, the measured impedance is not affected by the fault occurrence; and is equal to the measured impedance before the fault occurrence.

In the case of element CA, there is a small overlapping region between the tripping and the quadrilateral characteristics for the faults close to the near end of the line and the low magnitudes of the fault resistance.

Therefore, it can be concluded that in the case of a single phase to ground (AG) fault, in addition to the ground element corresponding to the faulted phase (AG), the other elements might operate. The probability of the multiple elements operation is high in the case of the faults close to the near end of the line and in the case of zero or low fault resistances. Therefore, the operation of up to four elements in a single phase to ground fault is possible.

For all elements, with exception of element BC, as the fault point approaches toward the far end of the line and the fault resistance increases (around 200 ohms and higher), the measured impedance approaches toward the measured impedance before the fault occurrence. As it can be seen the measured impedance for the fault at the far end of the line and faults resistance of 200 Ohms is very close to the pre-fault measured impedance. In the case of element BC, the measured impedance always is equal to the pre-fault measured impedance.

References

[1] Zhang Zhizhe and C. Deshu, "An adaptive approach in digital distance protection", *IEEE Trans. Power Delivery*, vol. 6, no. 1, pp. 135–142, Jan. 1991.

[2] Y. Q. Xia, K. K. Li, and A. K. David, "Adaptive relay setting for stand-alone digital distance protection", *IEEE Trans. Power Delivery*, vol. 9, no. 1, pp. 480–491, Jan. 1994.

[3] S. Jamali, "A fast adaptive digital distance protection", in *Proc. 2001 IEE 7th International Conference on Developments in Power System Protection, DPSP2001*, pp. 149–152.

[4] Chang-Ho Jung, Dong-Joon Shin, and Jin-O Kim, "Adaptive setting of digital relay for transmission line protection", in *Proc. 2000 IEEE International Conference on Power System Technology, PowerCon2000*, vol. 3, pp. 1465–1468.

[5] K. K. Li, L. L. Lai, and A. K. David, "Stand alone intelligent digital distance relay", *IEEE Trans. Power Systems*, vol. 15, no. 1, pp. 137–142, Feb. 2000.

[6] H. W. Dommel, "EMPT reference manual", Microtran Power System Analysis Corporation, Vancouver, British Columbia, Canada, August 1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Saving Of Natural Gas Through Optimal Operation Of Bangladesh Power System

Mohammad Tawhidul Alam[1] and Q. Ahsan[2]

[1]Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
Email: tawhidul_bclt@ yahoo.com

[2]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
Email: qahsan@eee.buet.ac.bd

Abstract - The depletion of gas with its use is a great concern of Bangladesh. Electricity sector consumes 47% of total annual use of gas. Because of the peculiar input energy requirement characteristic of thermal units with their output the generating units are operated such that the total input energy requirement is minimum in a power system. As load management through load shedding, is a common phenomenon of Bangladesh Power System (BPS), it is generally considered that the scope of optimal loading of units in BPS does not exist. This paper investigates the scope of optimal loading of units of BPS. It applies dynamic economic unit commitment scheduling technique for BPS to evaluate the production cost and the corresponding gas requirement for the past and future period. It also studies the possibility of reducing gas requirement in BPS through its optimal operation. The investigation results deserve the attention of natural resource planners and utilities.

I. Introduction

In BPS electricity is generated from fossil fuels and hydro resource. The statistics of last one decade shows that the average generation of hydro power is only 3% of total energy generation [1]. It also shows that about 87% of electricity is generated by natural gas.

The gas reserve of this country is significant but not abundant. The estimated gas reserve is presented in Table 1[2]. In Table 1, P_1 is estimated with reasonable certainty with recovery probability higher than P_2 and P_3. The recovery probability of P_2 is higher than P_3.

Table 1 Estimated gas reserve

Category of Estimate	Range of Reserve (TCF)
P_1+P_2	12.04 – 15.55
P_3	4.14 – 11.84
$P_1+P_2+P_3$	16.18 – 27.39

*P_1= Proved reserve, P_2= Probable reserve, P_3= Possible reserve

With the increase of population and industrial growth the use of natural gas in different sectors is increasing over the years. The average use of gas in different sectors during 1986-87 to 1999-2000 is shown in Figure 1 [2]. It is clearly observed from the figure that the use of gas in electricity sector is 47% of the total gas consumption of the country.

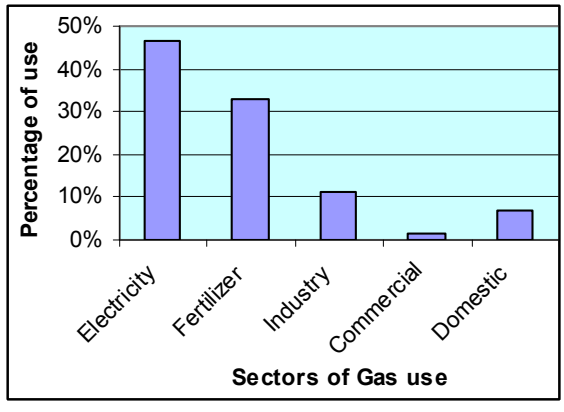


Figure 1 Average use of gas in different sectors

In an analysis, presented in [2] [3], the gas required to produce electricity in coming 17 years (2009 – 2025) is 9.72 TCF. The gas deficit/surplus in coming 17 years is presented in Table 2.

Table 2 Gas deficit or surplus in coming 17 years

Total gas requirement of the country in coming 17 years (Considering previous average percentage of use of all the sectors constant)		20.83 TCF
Gas deficit(-)/ surplus(+) (TCF)	Considering P_1+P_2	(-)5.28 to (-) 8.79)
	Considering $P_1+P_2+P_3$	(+)6.56 to (-)4.56

Table 2 shows that if P_1+P_2 is considered as the natural gas reserve of the country it will not be able to meet its own demand; rather it will have a deficit. The analysis clearly presents that the existing gas reserve of the

country will be completely exhausted within the next 15 years [2] if additional amount is not ensured through further exploration. Therefore, the reduction of gas use in electricity sector should duly be taken care.

II. Scope of Economic Unit Commitment for Reducing Gas Consumption Keeping the Generation Same

The present load profile of BPS reveals that the base load varies from 1800MW to 2900MW and the maximum load varies from 3200MW to 4200MW depending on the seasons. The average duration of peak hour is about 4 hours and it occurs at the evening period, usually from 18.00 to 22.00. The daily load factor varies from 70% to 80%. The available generation capacity of BPS is not always sufficient to meet the peak demand, but it is sufficient to meet the base load and mid level load.

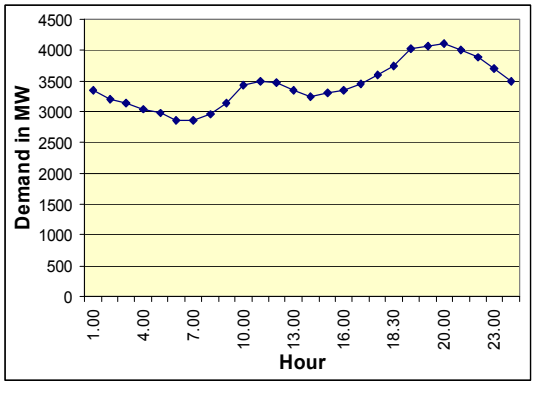


Figure 2 Daily load of a typical summer day of

Figure 2 shows the load curve of a typical summer day of 2007. It shows that the minimum load occurs during the morning period from 6 .00 am to 8.00 am and peak load occurs from 19.00 pm to 22.00 pm. The curve also reveals that the minimum load of the system is about 2900 MW and maximum load is 4100 MW. A difference of 1200MW occurs between the maximum and minimum load in the summer.

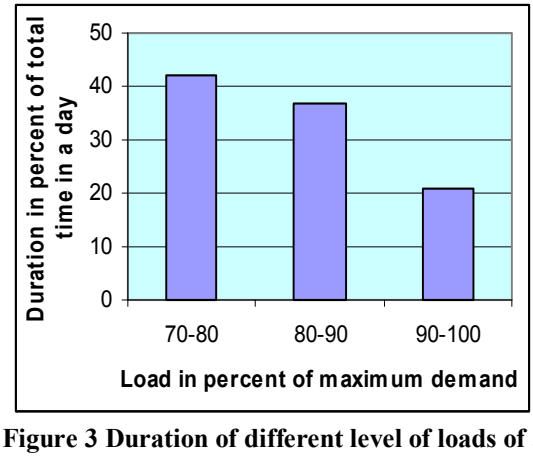


Figure 3 Duration of different level of loads of a typical summer day

The duration in percent of total time of different range of load levels of a typical summer day is depicted in figure 3. It shows that 42% time of a day the demand is less than 80% percent of the peak demand and only 20% time the demand exceeds 90% of the peak demand.

The hourly load of a typical winter day is depicted in figure 4. It is observed from this figure that the minimum load occurs during 3.00am to 5.00am and it is about 2100 MW.

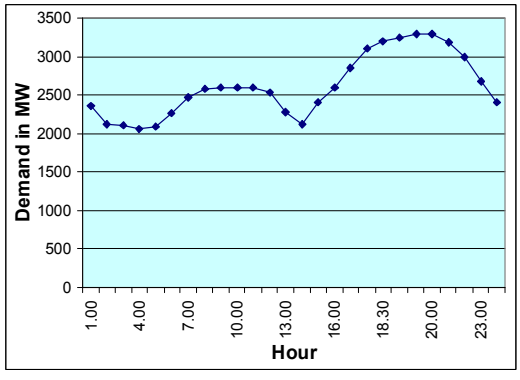


Figure 4 Daily load curve of a typical winter day

The peak load occurs during 18.00 pm to 21.00 pm and it is about 3300 MW. The difference between maximum and minimum load is about 1200 MW in the winter season, which is similar to that of the summer season.

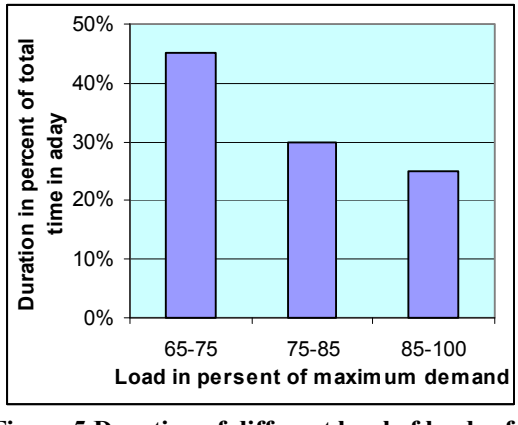


Figure 5 Duration of different level of loads of a typical winter day

The duration, in percent of total time, of different range of load levels of a typical winter day is depicted in figure 5. It shows that 45% time of a day the demand is less than 75% percent of the peak demand and only 25% time the demand exceeds 85% of the peak demand.

It is clear from the figures 2 and 4 that the minimum demand is 70% and 64 % of the maximum demand for the summer and winter season, respectively. Therefore, during the peak load hours of summer season all the units may be operated at their rated capacity but during the off peak hours of this season and all day long of winter season there is definitely a scope of economic unit commitment.

III. Technique of Economic Unit Commitment

The economic unit commitment is divided into two sub problems; one is economic allocation of system load on the generating units i.e. economic loading of the units. The other one is to choose the sets of generating units considering startup and shutdown cost of units and economic loading of units at the different load levels such that the global production is minimum. The set represents a group of economically loaded units for a given level.

A. Economic Loading

The input output characteristic of any unit 'i' can be expressed in the quadratic form as,

$$f_i = \frac{a_i}{2} P_{gi}^{\ 2} + b_i P_{gi} + c_i \qquad (1)$$

Where f_i is the fuel cost Tk/h and a_i, b_i and c_i are the coefficient and P_{gi} is the output of unit i.

Using the well established theory of loading units at equal incremental cost, the system incremental cost λ_s for economic distribution of load neglecting transmission loss may be written as [4-6],

$$\lambda_S = \left(\sum_{i=1}^{n} \frac{1}{a_i}\right)^{-1} \left(\sum_{i=1}^{n} P_{gi}\right) + \left(\sum_{i=1}^{n} \frac{1}{a_i}\right)^{-1} \left(\sum_{i=1}^{n} \frac{b_i}{a_i}\right) \qquad (2)$$

Where n is the total number of thermal units in the system.

B. Economic Scheduling

For the power system optimization, economic scheduling of the generating units is an important function. When the load level is changed for the successive hours, the optimal set of the generating units is also changed due to startup and shutdown of the units.

Among the unit commitment techniques the dynamic programming is one of the best technique [5],[6]. The mathematical technique for the optimization of multistage decision problem is called dynamic programming.

A recursive formula for the forward dynamic programming unit commitment is [7].

$$F_{i^*}(k) = \min_{\{x_j(k-1)\}} \{P_{i^*}(k) + T_{i^*j}(k) + F_j(k-1)\} \qquad (3)$$

Here, F is the fuel cost, k represents the load level stage number, i and j represent the unit combination number of adjacent stages, and T indicates the transition (startup and shutdown)cost from one combination of units to the other combination.

IV. Bangladesh Power System (BPS)

The economic unit commitment analysis requires the input output characteristic and transition cost of each generating unit in the system. In BPS these data are not readily available. The input output characteristic data of some units are directly collected from the corresponding power station and the characteristics of rest of the units are determined from their generation and fuel consumption data of different days obtained from load dispatch center (LDC) of BPS. The average incremental cost of different units of BPS are presented in Table 3 [7].

Table 3 Average incremental cost of the units of BPS

Types of generator	Types of fuel	Average incremental cost (Tk/kWh)
Gas turbine	Natural gas	0.6-1.2
	Furnace oil	3.2
	High speed diesel	10-11.8
	Super speed kerosene oil	7.5
Steam turbine	Natural gas	0.7-0.9
	Coal	1.53
	Furnace oil	3.2-4.3

The table shows that the generation cost of the units with the fuels FO, HSD, SKO is very high compared to that of gas or coal fired units.

BPS has 102 generating units of different capacities ranging from 2 MW to 235 MW. Among these 5 units are hydro and all are located in the same place at Kaptai. The total capacity of these hydro units is 230MW. The present installed capacity of BPS is 5245MW. Many of the units are operated at de-rated capacity because of their old age and maximum available capacity is 4631 MW. The generation system of BPS describing the types of fuel, average incremental cost, and rated capacity along with other characteristics is presented in Appendix.

In the study, the considered load model is the hourly actual generation of 2007 obtained from the daily log of LDC. The peak, base and total energy generation of 2007 are 3900MW, 1800MW, and 24673GWh, respectively, and the annual load factor of this year is 78%. The model also considers the forecasted hourly load of coming five years, 2009-2013 [1] for the investigation. The addition and retirement of the generating units during 2009-2013 are considered according to power system master plan (PSMP) [7].

V. Simulated Result

To investigate the impacts of optimal operation of the generating units of BPS on the consumption of natural gas the production cast and gas consumption are evaluated using the economic unit commitment (EUC) technique for all the twelve months of 2007. In the evaluation, the actual hourly load and the corresponding available generating units are considered. The monthly production cost and gas consumption are compared with the actual ones in Table 4. The table also presents monthly production cost savings and increased amount of gas requirement due to EUC along with the annual ones.

Table 4 Comparison of production cost and gas consumption for EUC with those of actual ones

Month	Cost (MillionTaka)			Gas used (Billion CFT)		
	Actual	EUC	Savings	Actual	EUC	Increased use
Jan	1782	1142	640	14.58	15.44	0.86
Feb	1771	1092	678	14.22	14.79	0.57
Mar	2120	1446	673	18.00	19.50	1.50
Apr	2151	1552	599	18.96	20.77	1.81
May	2103	1643	460	19.94	21.92	1.98
Jun	1952	1490	462	19.10	20.15	1.05
Jul	2083	1568	514	19.23	21.22	1.99
Aug	2132	1655	477	19.53	21.86	2.33
Sep	2151	1644	506	19.45	21.63	2.18
Oct	1996	1454	542	17.12	19.50	2.38
Nov	1697	1201	495	15.20	16.14	0.94
Dec	1659	1111	547	14.73	14.95	0.22
Total	23604	17004	6599	210.07	227.87	17.80

Table 4 clearly shows that 27.45% production cost could be saved during 2007 if the units were optimally loaded. However, he simulation results of 2007 also shows that the use of gas would be increased by 8.73%. The increased use of gas indicates that the low cost gas fired units offload the high cost liquid fuel units.

With an objective to evaluate the amount of reduced gas consumption through optimal loading of units, the amount of energy generated by hydro, coal fired, liquid fuel units is kept constant for 2007. That is, these units are simulated as must run units. Then the EUC technique is applied to load the gas fired units to supply the remaining demand energy for all 365 days of 2007. The required gas for each month of 2007 for the optimal loading of gas fired units is compared in Table 5 with the actual consumption resulting from conventional loading. The saving in gas requirement, total energy generation and the energy generated by the gas fired units only, are also presented in the table.

Table 5 Gas savings due to Economic unit commitment during 2007

Month	Generated Energy (GWh)		Gas use (Million CFT)		Gas Savings (Million CFT)
	Total	By gas fired units only	actual	For optimal operation of gas fired units	
Jan	1697.27	1507.71	14577.59	13836.91	740.68
Feb	1611.25	1445.66	14216.61	13308.96	907.65
Mar	2036.74	1867.02	17992.83	17753.18	239.65
Apr	2127.11	1912.55	18956.84	18381.33	575.5
May	2211.25	2038.74	19939.39	19631.97	307.41
Jun	2095.18	1939.88	19089.97	18649.32	440.65
Jul	2253.91	2064.45	20372.75	19922.3	450.44
Aug	2311.17	1991.25	19520.09	19152.13	367.9
Sep	2328.55	1968.91	19446.76	19002.92	443.84
Oct	2355.25	1778.74	17119.33	16889.25	230.08
Nov	1853.03	1490.38	15195.64	13886.12	1309.5
Dec	1792.58	1525.34	14729.54	14231.68	497.85
Total	**24673.34**	**21530.67**	**211157.4**	**204646.1**	**6511.29**

It is clearly observed from the Table 5 that if the gas fired units are optimally loaded then it would be possible to save 6511.29 Million CFT of gas, which is 3.08% of the actual consumption for the generation of 21530.67GWh energy of 2007.

Table 6 Gas savings in coming five years for optimal loading

Year	Forecasted energy generation (GWh)		Gas requirement (TCF)		Savings (TCF)
	Total	By gas fired units only	Forecasted	For Optimal operation of units	
2009	30718	29203	0.3345	0.242475	0.092025
2010	32756	31241	0.3585	0.25104	0.10746
2011	34889	33374	0.3837	0.261266	0.122434
2012	37117	35602	0.4105	0.264611	0.145889
2013	39459	37944	0.43755	0.277127	0.160423
Total	174939	167364	1.92475	1.296519	0.628231

Again considering the same strategy for hydro, coal fired and liquid fuel units as before, that is operating them as must run units, the gas fired units are optimally loaded to evaluate the gas requirement for the coming 5 years, 2009-2013. In the simulation, the forecasted hourly loads and the units of PSMP [1],[7] are considered. The annual gas requirement obtained from the simulated result of EUC is compared with the forecasted ones [2] obtained through conventional use in table 6. This table also presents annual gas savings, total energy generation and energy generated by the gas fired units only. Note that 1515 GWh energy is generated by the units other than gas fired units in each year. Table 6 shows that the optimal operation of the gas fired units will save a total amount of 0.62831 TCF gas in coming 5 years, which is 32.6% of the forecasted value.

VI. Conclusion

Because of higher demand for electricity compared to generation capability and also for forced load reduction, load shedding in all most all the days of the year, it is generally considered that there is no scope of EUC for BPS. However, a detailed analysis of loads and available generation capacity reveals that the scope of EUC exist for the system.

The results shows that the production cost could be saved up to 27.95% in last one year, 2007, if the units were optimally loaded. People of the country are concerned about the gas reserve compared to its enormous demand in different sectors and the investigation clearly reveals that the gas requirement in electricity sector can be reduced significantly if the generating units are optimally operated, in future.

VII. References

[1] Moin, U. 'Development of a Methodology for Long Term Hourly Electrical Load Forecasting and Assessment of Gas Requirement', M. Sc. Engineering Thesis, Department of EEE, BUET, Dhaka, 2003.

[2] Moin Uddin and Q.Ahsan "Expected Status of Natural Gas in Bangladesh in Meeting the Electricity Demand in the Next Twenty Five Years.", ICECE 2004,28-30 December, Dhaka, Bangladesh.

[3] Ahsan, Q. " Electric Power Crisis in Bangladesh- long term solution", Presented in the seminar of Mist, April 26, 2007, pp 1-19.

[4] Stevenson, W. D. Jr. "Elements of Power System Analysis" McGraw-Hill Book Company, New York, 1996.

[5] Wood, A.J. and Wollenbery, B.F. "Power Generation Operation and Control" New York ; wiley,1996.

[6] Stevenson, W. D. Jr., Grainger, J. J. "Power System Analysis" McGraw-Hill, New York, 1994.

[7] Mohammad Tawhidul Alam, "Evaluation of Fuel Cost Savings for Optimal Loading of the generating units of Bangladesh Power Syatem" M.Sc. Engg. Thesis (EEE), Bangladesh University of Engineering and Technology, Dhaka, 2008.

Appendix

The BPS has 102 generating units at different locations of the country with an installed capacity of 5245MW. The different characteristic features of the units are presented in Table A.

Table A: Basic information of generating units of BPS

Name of Power Stations/ Units **Under BPDB**	Forced outage rate(FOR)	Capacity (MW)		Types of fuel	Average incremental cost (Tk/MWh)
		Rated	Available		
Kaptai Hydro (2*40, 3*50)	0.0000014	230	230	Hydro	0
Siddhirganj Steam-(1* 210)	0.16	210	210	Gas	885
Siddhirganj Steam- (1*50)	0.113	50	32	Gas	897
Tongi gas turbine-(1*109)	0.07	109	109	Gas	980
Ghorasal steam-1&2 (2*55)	0.185	110	80	Gas	995
Ghorasal steam-3 (1*210)	0.095	210	190	Gas	845
Ghorasal steam-4 (1*210)	0.019	210	200	Gas	763
Ghorasal steam-5 (1*210)	0.08	210	200	Gas	825
Ghorasal steam-6 (1*210)	0.08	210	200	Gas	828
Ashuganj steam- 1&2 (2*64)	0.116	128	128	Gas	848
Ashuganj steam-3 (1*150)	0.013	150	150	Gas	815
Ashuganj steam-4(1*150)	0.014	150	150	Gas	794
Ashuganj steam-5(1*150)	0.014	150	0	Gas	810
Ashuganj GT-1,2 (2*56)	0.321	112	80	Gas	1011
Ashuganj CCPP Steam(1*30)	0.15	30	20	Gas	0
Haripur GT (3*33.33)	0.30	100	32	Gas	1188
Raozan steam –1 (1*210)	0.197	210	180	Gas	761

Name of Power Stations/ Units	Forced outage rate(FOR)	Capacity (MW)		Types of fuel	Average incremental cost (Tk/MWh)
Under BPDB		Rated	Available		
Raozan steam-2 (1*210)	0.197	210	210	Gas	837
shikalbaha steam (1*60)	0.117	60	50	Gas	888
Shikalbaha BMPP (2*28)	0.6	56	10	Gas	8870
Shahjibajar GT (1-7)(3*12, 4*15)	0.15	96	34	Gas	2416
Shahjibajar GT (8-9)(2*35)	0.1	70	60	Gas	954
Sylhet GT(1*21)	0.122	21	21	Gas	1478
Fenchuganj Comb. Cycle(3*30)	0.04	90	90	Gas	732
Baghabari GT-1 (1*77)	0.101	77	77	Gas	856
Baghabari GT-2 (1*100)	0.04	100	100	Gas	871
Barapukuria Steam-1,2 (2*125)	0.1	250	250	Coal	1525
Khulna Steam (1*110)	0.301	110	90	FO	3277
Khulna Steam (1*60)	0.402	60	50	FO	4368
Khulna BMPP (2*28)	0.5	56	32	SKO	7536
Sayedpur GT (1*20)	0.045	20	20	HSD	11847
Barisal GT(2*20)	0.2	40	32	HSD	11648
Rangpur GT(1*20)	0.119	20	20	HSD	11119
Bheramara GT(3*20)	0.5	60	56	HSD	10002
Combined Diesel (3+3+2)	0.3	8	8	Diesel	-
Total for BPDB		**3983**	**3369**		
Under Private Sector					
CDC Meghnaghat GT 1,2 (2*150) Steam (1*150)	0.07	450	450	Gas	647
CDC Haripur GT(1*235), Steam (1*125)	0.07	360	360	Gas	618
Haripur Barge(1-8)	0.11	110	110	Gas	765
RPCL Mymensing GT (4*35.57)	0.07	142	142	Gas	1190
WMPL Baghabari-1,2(2*45)	0.07	90	90	Gas	678
KPCL (Tiger) Khulna(1-19)	0.07	110	110	FO	3284
Total for Private Sector		**1262**	**1262**		
Grand total for BPS		**5245**	**4631**		

* Ashugang steam-5 is under Maintenance.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Lightning Surge Impedance Measurement on Control Building Using Electromagnetic Transient Program

Md. Mostafizur Rahman1, M.O. Goni2, Kazutaka Mitobe1 and Masafumi. Suzuki1

Faculty of Engineering and Resource Science, Akita University, Japan1 and Dept. of ECE, KUET2
Suzuki Lab, 1-1 Tegata Gakuen-cho, Akita-shi-010-8502, Japan and Khulna-920300, Bangladesh
E-mail: mostafiz963@yahoo.com

Abstract - This paper demonstrates the electromagnetic behavior of transient response of lightning surge strike on control building and power system. Surge responses of control building are analyzed using Electromagnetic Transient Program (EMTP). A direct lightning stroke to a building may cause damages or interference in the operation of electronic equipment inside the structure and even be dangerous for the people in the more critical case. In order to decide on an efficient lightning protection system, it is very important to determine the lightning current distribution in the building. The accuracy of this method is shown to be satisfactory with the simulation result based on the Numerical Electromagnetic Code (NEC-2) that were carried out on control building as reduced scale model. Lightning over-voltages must be taken into account when designing a control building. In the world, especially in Japan various studied, therefore, have been done to measure the characteristics of the line components and to develop their equivalent circuit models for the Electromagnetic Transient Program (EMTP) simulations. Surge impedance is calculated by both the vertical and horizontal stroke analyses method.

I. Introduction

The Control Building (CB) is called the group of conductors, makes a rectangular form situated above the ground with other conductor. And if the lightning surge hits inside or on the control building then there will be some effect of voltage and the current properties, so the surge impedance should proper for designing a control building. To protect the electronic components inside the control building and overhead transmission lines, lightning is very important cause of unscheduled interruption. In the calculation of the lightning induced over-voltages the accurate representation of the control building is a difficult problem and has been the subject of much discussion. Prediction of lightning surges is very important for the design of electric power systems and control building. Therefore, a number of experimental and theoretical studies on tower surge impedance and control building have been carried out [1]-[17].

Now a days the numerical analysis of lightning protection system is carried out in the two arrangements of the current lead wire of a building: i) horizontal and vertical conductors placed outside the structure, ii) conductive elements in reinforced concretes or steel constructions. During a lightning stroke to a protection systems the surge currents of high amplitude [2] and short rise time which flows to the ground may cause: i) damages in lightning protection system, ii) secondary spark (fire or explosive hazards), iii) unequal high voltage distribution (it may be danger for the people inside the building) iv) damages or functional disturbances on electronic communication, control and measuring components inside.

So the purpose of this paper is to predict surge current distribution and to measure the surge impedance taking into account results from simulation and theoretical calculations. The first theoretical formulation of surge impedance was proposed by Jordan [7]. It was assumed that the current distribution inside the tower was uniform from the tower bottom to the top of the tower. However, the effect of return stroke current was neglected. The tower was approximated as a vertical cylinder having a height equal to that of the tower, and a radius equal to the mean equivalent radius of the tower. Propagation velocity inside the tower was assumed to be the velocity of light.

In lightning surge simulations the control building models used can range from simple lumped inductances or resistances to complicated non-uniform transmission line circuits. Representation of the control building, as a lumped element is only valid if surge current rise time is long compared to surge travels time in the control building.

For simulation the surge impedance of control building, two methods has been used, one is direct method and the other is the refraction or reflection method. Simulation methods to investigate the surge characteristics of reduced scale model of control building. EMTP is used world wide for switching and lightning surge analysis. It is a universal program system for digital simulation of transient phenomena of electromagnetic as well as electromechanical nature.

In the simulation by the direct method, the step current is injected into the top of the control building, and the voltage between the top of the control building and a voltage measuring wire or the voltage across an insulator string is measured by a voltage divider. The numerical analysis is carried out in the two arrangements of the current lead wire: i) vertical and ii) horizontal and in

extension of a voltage measuring wire. Thus the surge impedance with horizontal and vertical current [1], [12-13] is considered again for the theoretical values of surge impedance given in the equation i and ii.

$$Z=60\ \{\ln(2\sqrt{2}h/r_0)-1.983\}\ \Omega \qquad (i)$$

$$Z=60\ \{\ln(2\sqrt{2}h/r_0)-1.540\}\ \Omega \qquad (ii)$$

A theoretical work was reported by Ishii and Baba [14]. Recently, the theoretical work was reported by Takahashi [3], [15, 16]. He derived the theoretical formula of surge impedance of a vertical conductor including the effect of ground plane and without ground plane. Thus theoretical value of surge impedance agrees satisfactory with the experimental and simulation results [3], [14]-[16].

II. Modelling Guidelines and Structure

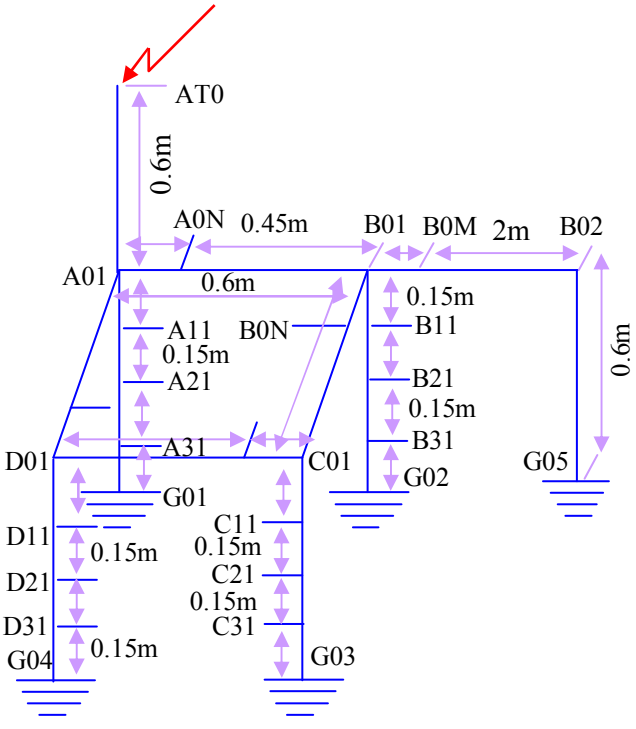


Fig. 1 Control Building-1 (Vertical Injection)

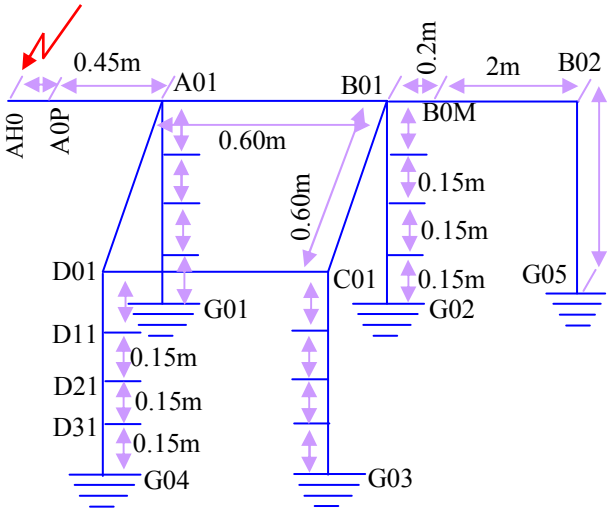


Fig. 2 Control Building-2 (Horizontal Injection)

In this paper the authors work with different models of control building each arm of building is 0.6 m in length (A01-B01, B01-C01, C01-D01 and D01-A01) and each building is 0.6 m in height from the ground. In these models simulation voltage measuring wire is taken 2.2 m in length and current lead wire 0.6 m in length. The radius of each arm is taken 0.08 cm. The legs and the arms of models of control building are divided into two to four segments for simulation work such as, A01-A11, A11-A21, A21-A31, A31-G01 and A01-A0N, A0N-B01, B01-B02, B02-G05 etc.

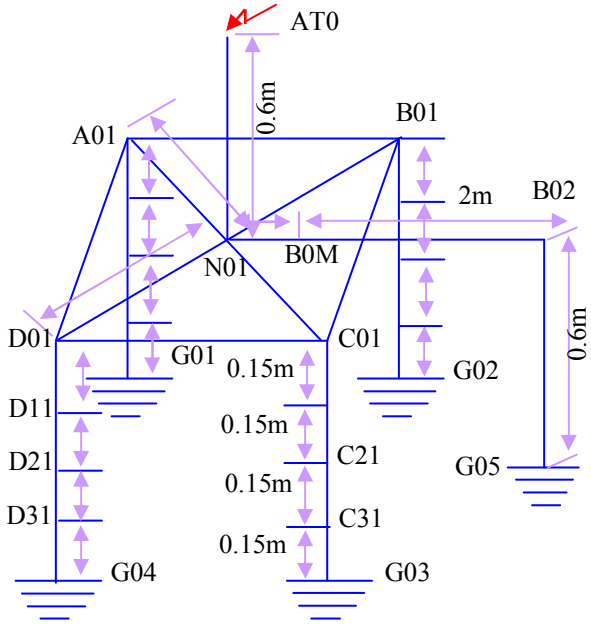


Fig. 3 Control Building-3 (Vertical Injection)

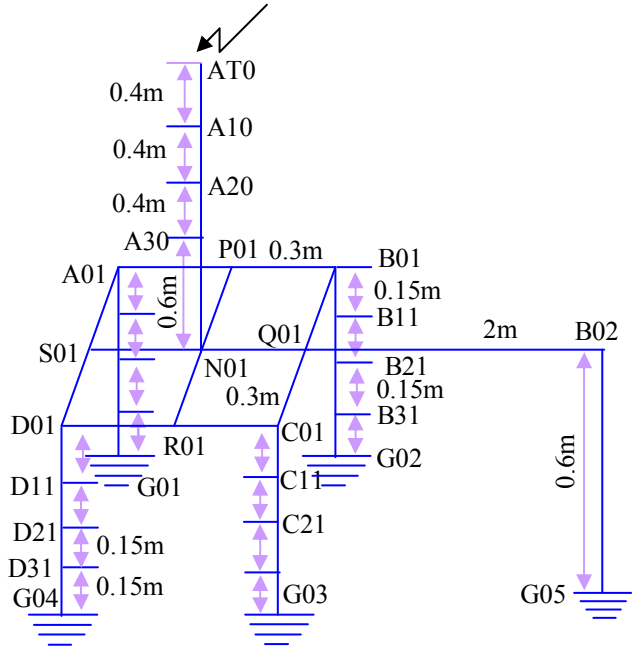


Fig. 4 Control Building-4 (Vertical Injection)

The simulation procedure is carried out for different methods of current injection using Electromagnetic

Transient Program (EMTP), where 5V voltage or pulse current generator is applied at the top of the current lead wire or lower segment of the current lead wire instead of surge over voltage. The setup for the numerical analysis is illustrated in Fig. 1, Fig. 2, Fig. 3 and Fig. 4

It is observed in the Fig. 1 that surge voltage is injected vertically at the top of the current lead wire and the current lead wire is situated at the corner of the control building AT0-A01, in the Fig. 2, the voltage is injected horizontally at the edge of the current lead wire and in the Fig. 3 current channel is diagonally situated at the Centre of the control building and voltage is injected vertically at the top of the current channel and finally Fig. 4 shows the current channel is just at the centre of the control building and the voltage is injected vertically at the top of the channel and the channel is divided into different segments. By varying current lead wire and voltage measuring wire the voltage waveform of the model of Fig. 1, Fig. 2 , Fig. 3 and Fig. 4 are shown later.

A. Voltage Analysis of Control Building by Vertical and Horizontal Injection.

The simulated voltage of control building (CB) is shown in the Fig. 5. If the surge current propagates along the control building at the speed of light, the reflected wave from the ground should return to the top after twice of the CB travel time which is 4 ns for the structure of the CB.

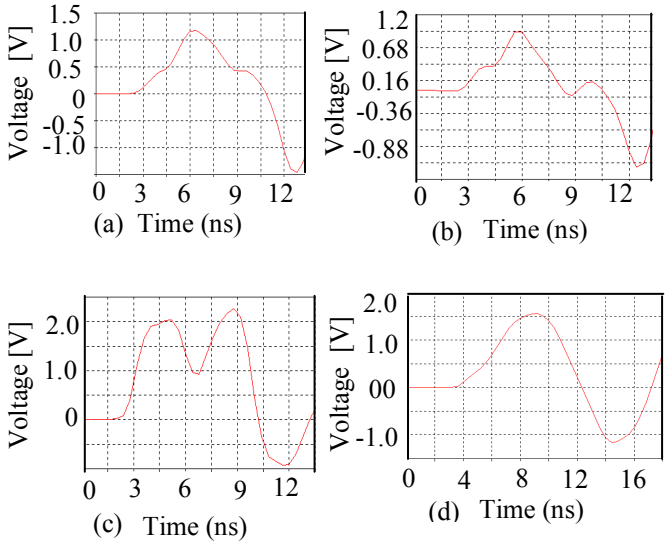


Fig. 5 Simulation voltage of different Control Buildings.

But in the simulated result in Fig. 5a the voltage characteristics of the control building of vertically applied voltage and each conductor leg is divided into four segments and overhead conductor line divided into two segments of the control Building. 40 ns is taken for one complete pulse-width and 5V is applied for surge representation. There consists small delay time in the voltage waveform before the beginning of rising point. The delay time is 2.6 ns and the pulse of the voltage wave-shape is approximately 10 ns.

So that from the above voltage waveform of measuring voltage curve of Fig. 5a is shown that voltage rises at 2.6 ns and start falling at 6.3 ns. The rising time of this model shown in the Fig. 5a is 3.7 ns. Now if the surge current propagates along the CB at the speed of light, the reflected wave from the ground should return to the tower top after twice of the tower travel time which is 4 ns for the structure of the model of Fig. 1. So that for the structure of the model of Fig. 1 indicates that a negative voltage wave arrives at the CB top before the arrival of the reflected wave from the ground.

Now the voltage waveform Fig. 5b is shows that voltage rises at 2.5 ns and start falling at 6 ns. The rising time of this model shown in the Fig. 5b is 3.5 ns which is almost to the same of Fig. 5a that is structure of the model of Fig. 2 indicates that a negative voltage wave arrives at the CB top before the arrival of the reflected wave from the ground.

And Fig. 5c and 5d shows the voltage waveform rises at 2 ns and 3.5 ns and start falling at 8.8 ns and 9.2 ns. The rising time of this model shown in the Fig. 5c and 5d is 6.8 ns and 5.7 ns which is almost equal to the calculated value.

The delay time of control building of Fig. 1, 2 and 3 is almost 2.7 ns and Fig. 4 is for 3.8 ns which is shown in the voltage waveform of Fig. 5. Delay time of control building 4 is greater than other three building as because current channel of control building 4 which is shown in Fig. 4 is greater than other three control buildings.

B. Current Analysis of Control Building

There are current wave shape of the different control building are shown in the Fig. 6. Also the current wave shape had shown the rising time, delay time and reflection time.

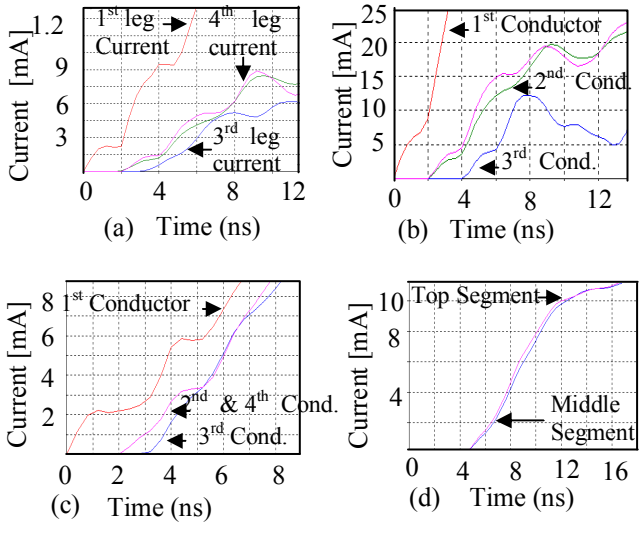


Fig. 6 Upper segment current of the Control Buildings

From the current wave-shape of the above Figs. 6a, 6b, 6c and 6d shows the characteristics of input current or upper segment current of control building. Segment current and each input current have some delay time and reflection time for some short length. In Fig. 6a, 6b and 6c shows the 1st curve of the input current starts from origin and no

delay time, because node A01 in the model is the first point and surge current didn't propagate any distance, on the other hand input current of 2nd and 4th curve starts with same delay time 2 ns as the calculated value, because they propagate same distance and current 3rd curve starts with more delay time 4 ns which is exact to the calculated value 4 ns.

Now it is shown another current wave shape in Fig. 6d that the input current of the Fig. 4 A01-A11, B01-B11, C01-C11 and D01-D11 of the conductor leg of the control building have also delay time, reflection time and rise time. The input current or upper segment of the current starts with same delay time 5 ns for propagating the same distance of the control building and it rises at the peak value after 10 ns.
The figure shows these currents starts from same point that is the delay time is almost same for surge currents propagate the same distance. The reflection of all these currents is almost 10 ns.

C. Surge Impedance Measurement of Different Control Building

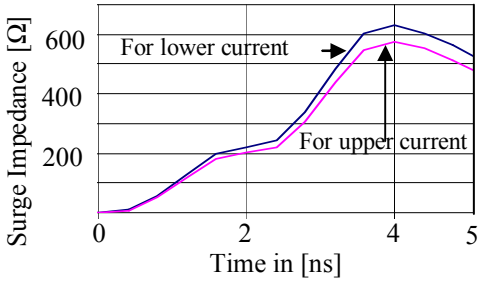


Fig. 7 Surge Impedance of Control Building–1

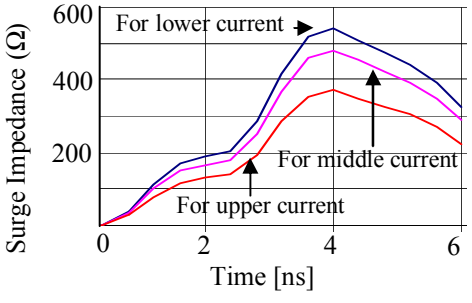


Fig. 8 Surge Impedance of Control Building–2

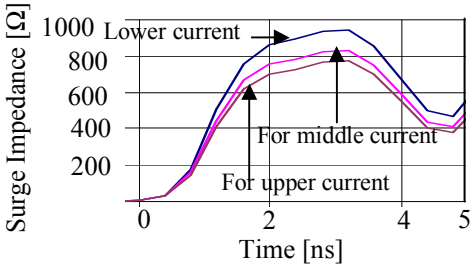


Fig. 9 Surge impedance of Control Building-3

Figure 7, 8, 9 and 10 shows the surge impedance curve of the control building 1, 2, 3 and 4 which is summarized in the table 1.

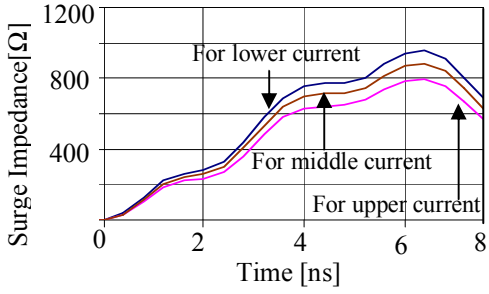


Fig. 10 Surge Impedance of Control Building-4.

Table: 1 Comparative Statement of Surge Impedance.

	CB-1	CB-2	CB-3	CB-4
Upper	631Ω	540Ω	947Ω	956Ω
Middle	------	480Ω	833Ω	883Ω
Lower	574Ω	370Ω	731Ω	797Ω
Theoretical	720	648	720	720

CB-1: Vertically applied Voltage
CB-2: Horizontally applied Voltage
CB-3, CB-4: Vertically (Centrally) applied Voltage

III. Conclusions

From this work it is seen that electromagnetic transient behaviors of lighting surge on control building are analyzed using EMTP and the result is compared with theoretical values. Voltage, Current and Surge response are analyzed in the different models of the control building. Different Control buildings of reduced scale model are investigated using different procedure, i.e. using the voltage applied on the top of the control building of vertically applied voltage and horizontally applied voltage. It can easily compare from the comparison table that surge impedance for vertically applied is greater than that of horizontally applied voltage for control building.
Again Centrally applied voltage of control building is also investigated and the surge impedance is greater effect for this reason. Surge impedance is very high for centrally applied voltage than applied voltage at the edge or corner of the control building.
In future any one can simulate and measure the surge impedance by considering different components inside the control building.

References

[1] Md. Mostafizur Rahman, Md. Foisal Hossain, A B M Aowlad Hossain, M. O. Goni" Lightning Surge Analysis on Vertical Tower Using Electromagnetic Transient Program." *4th International Conference, Proceedings of ICECE-2006, 19-21 December, 2006*, (ISBN: 98432-3614-1)

[2] Md. Osman Goni, Md Faruque Hossain, Md. Salah Uddin Yusuf, Md. Mostafizur Rahman, E. Kaneko and H. Takahashi, "Simulation and experimental analysis of transient behavior of lightning surge on vertical conductors", *IEEE Trans. Power Delivery*, Volume 21, October 2006

[3] M.O.Goni, P.T.Cheng and H.Takahashi, "Theoretical and experimental investigation of the surge response of a vertical

conductor", *in Proc.of IEEE Power engineering society conf*., vol.2, pp.699-704, 2002.

[4] M.O.Goni and H.Takahashi,"Theoretical and experimental investigation of the surge response of a vertical conductor", *The ACES Journal*, vol.18, no.1, pp.41-47, Mar.2003.

[5] M.O.Goni, E.Kaneko and H.Takahashi,"Thin wire representation of the vertical conductor surge simulation", *The ACES Journal*, vol.18, no.1, pp.41-47, Mar.2004.

[6] S. Cristina, M. D Amore and A Orlandi, " Lightning stroke to a structure protection system. Part I: Current distribution analysis", presented on the 6th *International Symposium on High voltage Engineering. New Orleans, USA*, 1989.

[7] C. A. Jordan, "Lightning Computation For Transmission Line with Ground Wires," *General Electric Review*, vol. 34, pp. 180-185, 1934.

[8] A. Sargent and M. Darveniza, "Tower surge impedance," *IEEE Trans. PAS*, vol. 88, pp.680-687, 1969.

[9] K. Okumura and A. Kijima, " A method for computing surge impedance of transmission line tower by electromagnetic field theory," *IEE of Japan Trans. B*, vol. 105, pp. 733-740,1985.

[10] M. Kawai, "Studies of the surge response on a transmission line tower, " *IEEE Trans. PAS*, vol. 83, pp. 30-34, 1964.

[11] W. A. Chisholm, Y. L. Chow, and K. D. Srivastava, "Lightning surge response of transmission towers," *IEEE Trans.,* vol. PAS-102, pp. 3232-3242, 1983.

[12] M. A. A. Wahab, I. Matsubara, and H. Kinoshita, "An experimental evaluation of some factors affecting tower surge impedance," *Trans. IEE of Japan*, vol. 107-E, no.9/10, pp.171-178, 1987.

[13] T. Yamada, A. Mochizuki, J. Sawada, E. Zaima, T. Kawamura, A. Ametani, M. Ishii, and S. Kato, "Experimental evaluation of a UHV tower model for lightning surge analysis," *IEEE Trans., PWRD*, vol. 10, no.l, pp. 393-402, Jan. 1995.

[14] M. Ishii and Y. Baba, "Numerical electromagnetic field analysis of tower surge response, " *IEEE Trans., PWRD*, vol. 12, pp. 483-488,1997.

[15] H. Takahashi,"A consideration on the vertical conductor problem", *Proc. of ICEE,* pp.635-638, 2001.

[16] H.Takahashi and M.Kodama,"Theoretical derivation of surge impedance about a vertical conductor", *in Proc.of IEEE Power engineering society Conf.*, vol.2, pp.688-693, 2002.

[17] A. Sowa, "Lightning Over voltages in wires within the building", presented at *IEEE 1991 International Symposium on EMC*, New Jersey.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Voltage Fluctuations in a Remote Wind-Diesel Hybrid Power System

Sheikh Mominul Islam, M. Tariq Iqbal, John E. Quaicoe

Faculty of Engineering and Applied Science
Memorial University of Newfoundland
St. John's, NL, Canada A1B 3X5
Email: mominul511@gmail.com

Abstract: Generation of electricity using diesel is costly for a small remote isolated system. At a remote location electricity generation from renewable energy such as wind can help to reduce the overall operating costs by reducing the fuel costs. Siting of wind turbines connected to the diesel in a remote location effects wind penetration and power quality. In this paper we study how the location of a wind turbine effects the voltage in a remote wind-diesel system. We selected an isolated grid system in Cartwright, Labrador, Canada. Four possible sites for a wind turbine are selected and for all four cases we study the voltage variations in the system due to addition of a wind turbine. Hybrid power system modeling and simulation results are presented. Results will be used to finalize siting of the wind turbine.

I. Introduction

Cartwright is an isolated community in Southern Labrador, Canada. It generates electricity using four diesel generators which consume more than 1.2 million liters of fuel per year. As fuel price is increasing and so the cost of electricity generation in Cartwright. By adding electricity generation from a wind turbine will reduce the diesel consumption in Cartwright. The utilization of wind energy is not a new technology but draws on the long tradition of wind power technology. Generation of electricity from wind is a fastest growing energy technology in the world. Figure 1 shows the topographical map of Cartwright, Labrador.

Figure 1. Topographical location of Cartwright.

Cartwright is situated on southern cost of Labrador. Community's main source of income is fishery. Diesel plant is located in the community. Wind turbine can be installed outside the community. To extract maximum power from wind, turbine can be placed on top of any of the hill in the area. Small circles in figure 1 represent the top of the hills. We have considered different lengths of transmission line for the wind turbine. Figure 2 and figure 3 show the satellite image of two possible wind turbine locations and distance from the diesel plant.

Figure 2. Wind turbine is placed 6.5km away from grid

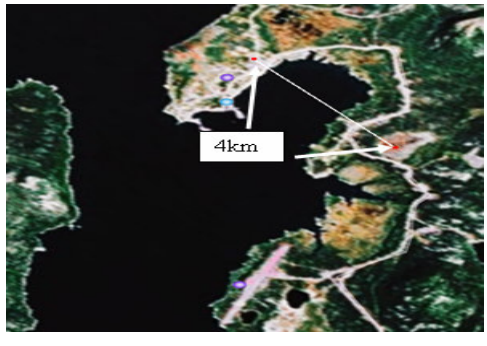


Figure 3. Wind turbine is placed 4km away from grid

To study voltage fluctuations for each possible wind turbine site we had to model diesel generator, transmission line and the whole wind energy conversion system, which includes wind turbine, gear box and induction generator. Generally there are two types of utility scale wind turbines, fixed speed and variable speed. Fixed speed wind turbines operate at a near constant rotor speed at all times which are directly connected to the grid. The fundamental frequency of the grid determines the rotor speed. Variable speed wind turbines can operate at a wider range of speeds depending on the wind speed. A variable rotor speed can be achieved by varying the blade pitch angle and generator speed. In

this case power converters are needed to interface the wind turbine to the power system. For this study we consider a 100kW fixed speed wind turbine. The MATLAB/SIMULINK modeling of a fixed speed wind turbine, diesel generator and transmission line are discussed in the following sections.

.

II. Wind turbine model

With the help of elementary momentum theory the typical output characteristic of a wind turbine is given by:

$$P = \frac{1}{2} C_p(\lambda, \alpha) \rho v^3 A \tag{1}$$

Here A (m^2) is the rotor swept area and C_p is known as the performance co-efficient which can be expressed as a function of tip speed ratio (λ). The tip speed (λ) is obtained from the quotient of the peripheral velocity V_u (rad/sec) to the undisturbed wind velocity V_1 (m/s)

$$\lambda = \frac{V_u}{V_1} \tag{2}$$

For a fixed blade pitch turbine C_p can be approximately expressed as [1]:

$$\text{Cp} = \frac{1}{2} \times \left(\frac{116}{\lambda_p} - 0.4 \times (\alpha - 5) \right) \exp^{\frac{-16.5}{\lambda_p}} \tag{3}$$

Where

$$\lambda_p = \frac{1}{\frac{1}{(\lambda + 0.089)} - \frac{0.035}{\alpha^3 + 1}} \tag{4}$$

And α is the blade pitch angle. A mid sized wind turbine with a rated power of 100kW at rated rotor speed of 32 rpm and wind speed of 13m/s is used for the modeling. The blade pitch angle is assumed to be constant.

III. Gear Box Model

The drive power of a wind turbine engenders torque in its mechanical drive train or generator that is subject to fluctuation as a result of both periodic and aperiodic process, such as

a) change in wind speed
b) tower-shadow or tower-occasioned upwind overpressure
c) blade asymmetry
d) blade bending and skewing and
e) tower oscillation

In addition, load moments in generator and converter due to static, dynamic and electromechanical disturbances also act on the wind turbine via the drive train [2]. The equivalent model of a wind turbine drive train is presented in figure 4. It includes turbine low speed shaft, gear box and generator high speed shaft. Here aerodynamic torque (T_T) is counteracted by the electromechanical torque generated by the generator (T_g) through the gear box.

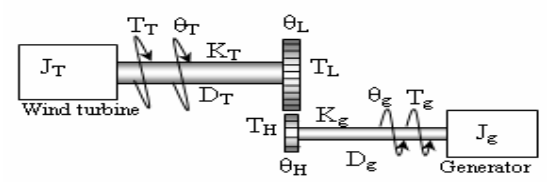


Figure 4. Two mass Model of the wind turbine drive train

The model can be reduced to a single mass model by considering high stiffness of wind turbine rotor and generator and also neglecting the moment of inertia, damping co-efficient and stiffness for the gear box as these are very small compared with the wind turbine rotor and generator[4]. The torque equation for this simplified model is given by-

$$T_g - T_T = J_{equ} \times \frac{d^2\theta_G}{dt^2} \tag{5}$$

$$J_{equ} = J_G + \frac{J_T}{n_g^2} \tag{6}$$

Here n_g is the gear ratio of the gear box. Torque Equation is represented for induction machine as motor operation. For generator operation the only change is instead of negative sign there will be a positive sign [5] in torque equation.

IV. Induction Machine Model

The induction machines commonly used on fixed – speed wind turbines are very similar to conventional industrial induction motors. The d-q or dynamic equivalent circuit of induction machine is shown in figure 5. There are numerous ways of formulating the equations of an induction machine for the purposes of computer simulation. One of the most popular induction motor models derived from this equivalent circuit is Krause's model [7]. The current equations which were used in this simulation are given below:

$$i_{qs} = \frac{1}{X_{ls}} \left(\varphi_{qs} - \varphi_{mq} \right) \tag{7}$$

$$i_{ds} = \frac{1}{X_{ls}} \left(\varphi_{ds} - \varphi_{md} \right) \tag{8}$$

$$i_{qr} = \frac{1}{X_{lr}} \left(\varphi_{qr} - \varphi_{mq} \right) \tag{9}$$

$$i_{dr} = \frac{1}{X_{lr}} \left(\varphi_{dr} - \varphi_{md} \right) \tag{10}$$

here

$$\varphi_{mq} = X_m \left(i_{qs} + i'_{qr} \right)$$
$$\varphi_{md} = X_m \left(i_{ds} + i'_{dr} \right)$$

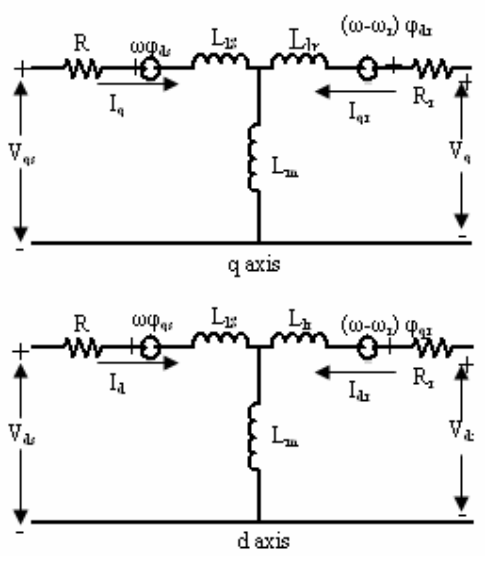


Figure 5. d-q axis equivalent circuit of a 3-phase induction machine.

And the flux linkage in terms of voltage quantities can be written as:

$$\varphi_{qs} = \omega_b \int \left[v_{qs} - \frac{\omega}{\omega_b}\varphi_{ds} + \frac{R_s}{X_{ls}}\left(\varphi_{mq} - \varphi_{qs}\right)\right] dt \qquad (11)$$

$$\varphi_{ds} = \omega_b \int \left[v_{ds} + \frac{\omega}{\omega_b}\varphi_{qs} + \frac{R_s}{X_{ls}}\left(\varphi_{md} - \varphi_{ds}\right)\right] dt \qquad (12)$$

$$\varphi'_{qr} = \omega_b \int \left[v'_{qr} - \frac{\omega-\omega_r}{\omega_b}\varphi'_{dr} + \frac{R'_r}{X'_{lr}}\left(\varphi_{mq} - \varphi'_{qr}\right)\right] dt \qquad (13)$$

$$\varphi'_{dr} = \omega_b \int \left[v'_{dr} + \frac{\omega-\omega_r}{\omega_b}\varphi'_{qr} + \frac{R'_r}{X'_{lr}}\left(\varphi_{md} - \varphi'_{dr}\right)\right] dt \qquad (14)$$

Since a squirrel cage induction machine is used, V'_{dr} and V'_{qr} in the above equations can be set to zero. So the magnetizing flux equation can be written as:

$$\varphi_{mq} = X_t\left(\frac{\varphi_{qs}}{X_{ls}} + \frac{\varphi'_{qr}}{X'_{lr}}\right) \qquad (15)$$

$$\varphi_{md} = X_t\left(\frac{\varphi_{ds}}{X_{ls}} + \frac{\varphi'_{dr}}{X'_{lr}}\right) \qquad (16)$$

where

$$X_t = \left(\frac{1}{X_m} + \frac{1}{X_{ls}} + \frac{1}{X'_{lr}}\right)^{-1} \qquad (17)$$

The electromagnetic torque equation is given by

$$T_g = \frac{3P}{4\omega_b}\left(\varphi_{ds} i_{qs} - \varphi_{qs} i_{ds}\right) \qquad (18)$$

V. Diesel Generator

Diesel generator is equipped with diesel engine and a synchronous generator.

Diesel Engine: For modeling, diesel engine statistical data curves were used which are provided by the manufacturer is shown in figure 6.

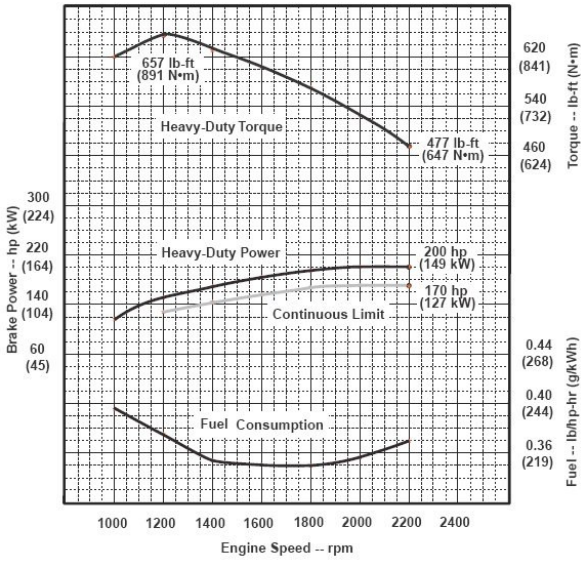


Figure 6. Diesel engine characteristic curves

By using the best fitting curve option in MATLAB a sixth order polynomial curve is derived for both engine speed vs. power and engine speed vs. fuel consumption curves.

Output power is given by:

$$\begin{aligned} P = &\left(5.1554\times10^{-8}\,\omega^6\right) - \left(7.0552\times10^{-5}\,\omega^5\right) \\ &+ \left(0.037776\omega^4\right) - \left(10.241\omega^3\right) + \left(1485.8\omega^2\right) \\ &- \left(1.0823\times10^5\,\omega\right) + 3.1539\times10^6 \end{aligned} \qquad (19)$$

Fuel consumption is given by:

$$\begin{aligned} F_c = &\ 5.667\times10^{-11}\,\omega^6 - 6.6665\times10^{-8}\,\omega^5 \\ &+ 3.2056\times10^{-5}\,\omega^4 - 0.0080567\,\omega^3 \\ &+ 1.1168\,\omega^2 - 81.071\,\omega + 2618.6 \end{aligned} \qquad (20)$$

Here ω is the engine speed (rad/sec), P is the engine power (kW) and F_c is the engine's fuel consumption (g/kWh).

Synchronous generator: The d-q or dynamic equivalent circuit of a synchronous machine is shown in figure 7. The current equations expressed in rotor reference frame can be given by [7]:

$$i_{qs}^{r} = -\frac{1}{X_{ls}}\left(\varphi_{qs}^{r} - \varphi_{mq}^{r}\right) \quad (21)$$

$$i_{ds}^{r} = -\frac{1}{X_{ls}}\left(\varphi_{ds}^{r} - \varphi_{md}^{r}\right) \quad (22)$$

$$i_{kq1}^{/r} = -\frac{1}{X_{kq1}^{/}}\left(\varphi_{kq1}^{/r} - \varphi_{mq}^{r}\right) \quad (23)$$

$$i_{kq2}^{/r} = -\frac{1}{X_{kq2}^{/}}\left(\varphi_{kq2}^{/r} - \varphi_{mq}^{r}\right) \quad (24)$$

$$i_{fd}^{/r} = -\frac{1}{X_{lfd}^{/}}\left(\varphi_{fd}^{/r} - \varphi_{md}^{r}\right) \quad (25)$$

$$i_{kd}^{/r} = -\frac{1}{X_{lkd}^{/}}\left(\varphi_{kd}^{/r} - \varphi_{md}^{r}\right) \quad (26)$$

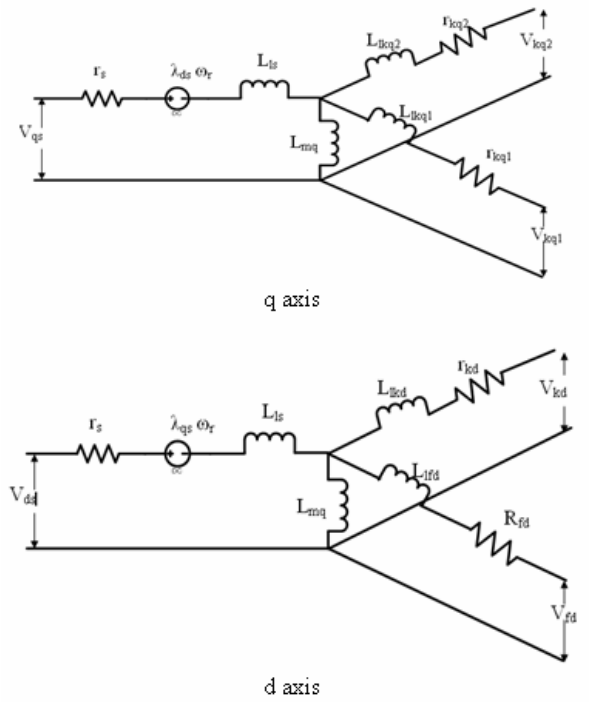


Figure 7. d-q axis equivalent circuit of a 3-phase synchronous machine.

Flux linkage equations which were used in this simulation are given as follows:

$$\varphi_{qs}^{r} = \omega_b \int \left[v_{qs}^{r} - \frac{\omega_r}{\omega_b}\varphi_{ds}^{r} + \frac{r_s}{X_{ls}}\left(\varphi_{mq}^{r} - \varphi_{qs}^{r}\right)\right] dt \quad (27)$$

$$\varphi_{ds}^{/r} = \omega_b \int \left[v_{ds}^{r} + \frac{\omega_r}{\omega_b}\varphi_{qs}^{r} + \frac{r_s}{X_{ls}}\left(\varphi_{md}^{r} - \varphi_{ds}^{r}\right)\right] dt \quad (28)$$

$$\varphi_{kq1}^{/r} = \omega_b \int \left[v_{kq1}^{/r} + \frac{r_{kq1}^{/}}{X_{lkq1}^{/}}\left(\varphi_{mq}^{r} - \varphi_{kq1}^{/r}\right)\right] dt \quad (29)$$

$$\varphi_{kq2}^{/r} = \omega_b \int \left[v_{kq2}^{/r} + \frac{r_{kq2}^{/}}{X_{lkq2}^{/}}\left(\varphi_{mq}^{r} - \varphi_{kq2}^{/r}\right)\right] dt \quad (30)$$

$$\varphi_{fd}^{/r} = \omega_b \int \left[\frac{r_{fd}^{/}}{X_{md}} e_{xfd}^{/r} + \frac{r_{fd}^{/}}{X_{lfd}^{/}}\left(\varphi_{md}^{r} - \varphi_{fd}^{/r}\right)\right] dt \quad (31)$$

$$\varphi_{kd}^{/r} = \omega_b \int \left[v_{kd}^{/r} + \frac{r_{kd}^{/}}{X_{lkd}^{/}}\left(\varphi_{md}^{r} - \varphi_{kd}^{/r}\right)\right] dt \quad (32)$$

Magnetizing flux equation can be given by:

$$\varphi_{mq}^{r} = X_{aq}\left(\frac{\varphi_{qs}^{r}}{X_{ls}} + \frac{\varphi_{kq1}^{/r}}{X_{lkq1}^{/}} + \frac{\varphi_{kq2}^{/r}}{X_{lkq2}^{/}}\right) \quad (33)$$

$$\varphi_{md}^{r} = X_{ad}\left(\frac{\varphi_{ds}^{r}}{X_{ls}} + \frac{\varphi_{fd}^{/r}}{X_{lfd}^{/}} + \frac{\varphi_{lkd}^{/r}}{X_{lkd}^{/}}\right) \quad (34)$$

where

$$X_{aq} = \left(\frac{1}{X_{mq}} + \frac{1}{X_{ls}} + \frac{1}{X_{lkq1}^{/}} + \frac{1}{X_{lkq2}^{/}}\right)^{-1} \quad (35)$$

$$X_{ad} = \left(\frac{1}{X_{md}} + \frac{1}{X_{ls}} + \frac{1}{X_{lkd}^{/}} + \frac{1}{X_{lfd}^{/}}\right)^{-1} \quad (36)$$

The electromagnetic torque equation can be given as follows:

$$T_g = \frac{3P}{4\omega_b}\left(\varphi_{ds} i_{qs} - \varphi_{qs} i_{ds}\right) \quad (37)$$

VI. Transmission Line

For modeling transmission line a distributed parameter is used [8]. Figure 8 shows a technique for implementing a transmission line using the single-phase circuit.

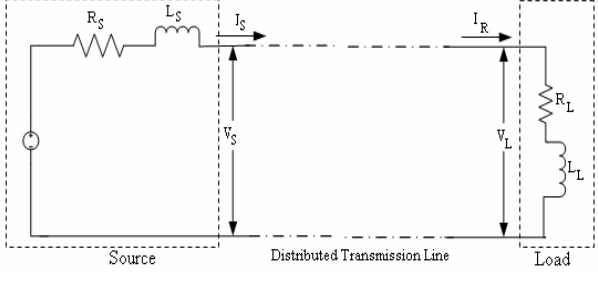


Figure 8. Single-phase line connecting a source to load.

Here at source end, the source end current can be calculated as:

$$I_s = \frac{1}{L_s}\int \left(e - V_s - R_s I_s\right) dt \quad (38)$$

Source end voltage (V_s) can be calculated as:

$$V_s = Zc\, I_s + 2V_{bs} \quad (39)$$

V_{bs} is the backward voltage component at the sending end. At receiving end, the load current can be calculated by integrating the voltage equation for the load which is given as follows:

$$I_R = \frac{1}{L_R}\int (V_R - R_L I_R)dt \quad (40)$$

If V_{fR} is the forward voltage component at the receiving end. Receiving end voltage (V_R) can be calculated as:

$$V_R = 2V_{fR} - Z_C I_R \quad (41)$$

The characteristic impedance of the transmission line can be given by:

$$Z_C = \sqrt{\frac{L}{C}} \quad (42)$$

The attenuation characteristic and transport delay of transmission line can be given as follows:

$$\text{Attenuation factor} = e^{-\frac{R}{2}\sqrt{\frac{C}{L}}x} \quad (43)$$

$$\text{Transmission delay} = x\sqrt{LC} \quad (44)$$

Here x is the length of transmission line.

VII. Simulation Results and Discussion

With the help from equation (1) to equation (44) the complete wind-diesel system is modeled in SIMULINK as shown in figure 9.

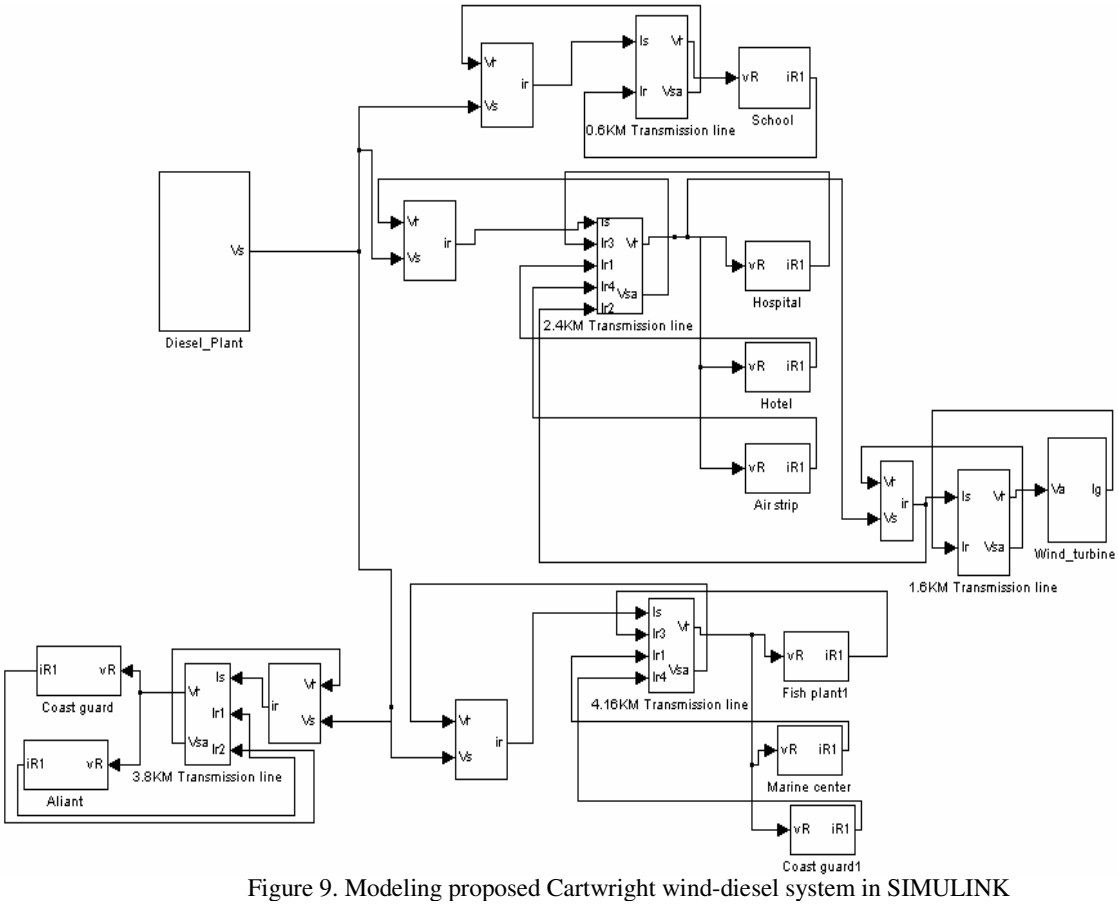


Figure 9. Modeling proposed Cartwright wind-diesel system in SIMULINK

For the simulation wind speed is input to the wind turbine and three phase voltage from grid is applied to the induction machine through a 4km transmission line. The simulation was performed for 10 second time interval and simulation integral and basic mathematical blocks are used instead of derivative term to avoid spikes in signals. All initial transients are included in this paper. Here the wind speed signals are generated using simulink step input block. To introduce power in grid by wind turbine, wind speed was changed from 0 m/s to 13 m/s at the 5th second of simulation time.

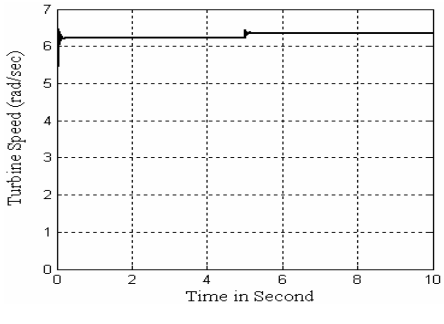


Figure 10. Turbine speed (rad/sec)

At null wind speed, the turbine was rotating as a motor and takes power from the grid. When the wind speed starts increasing from null speed then turbine speed is also increasing. Here the speed of wind turbine can be controlled by adjusting the blade angle. By using a gear box decoupling mechanism the turbine can be detached from the induction machine when the wind speed is lower than the cut-in speed.

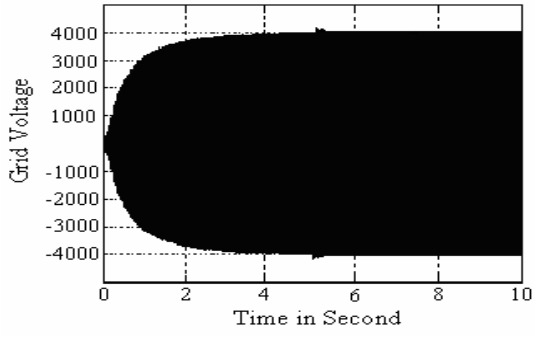


Figure 11. Grid Voltage (V)

From figure 11 we see that there is a voltage fluctuation in the grid voltage. After zoom-in the grid voltage in simulink we get voltage fluctuation about 60V which is shown in figure 12.

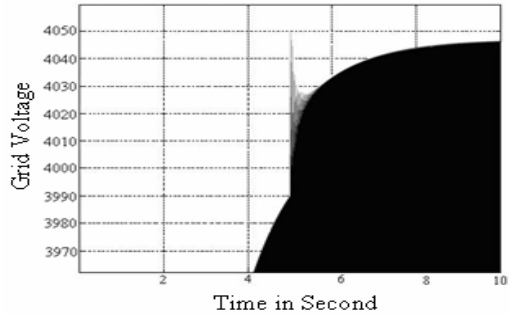


Figure 12. Voltage fluctuation at Grid

From figure 12 we see that at 5th second there is a fluctuation in grid voltage as wind turbine starts feeding power to the grid.

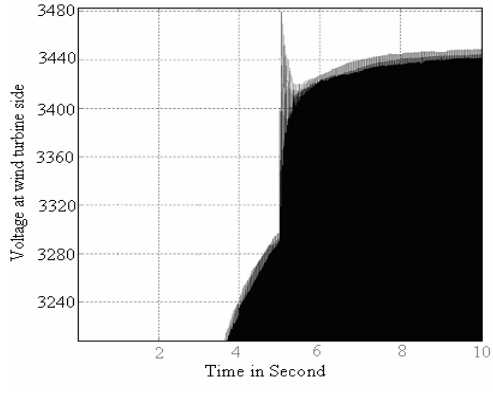


Figure 13. Voltage Fluctuation at wind turbine

From figure 13 we see that there is a fluctuation at wind turbine terminal which is about 180V at 5th second of total simulation time. As there is a 4km length of transmission line between wind turbine and grid so this large fluctuation attenuates while traveling through the transmission line to the grid.

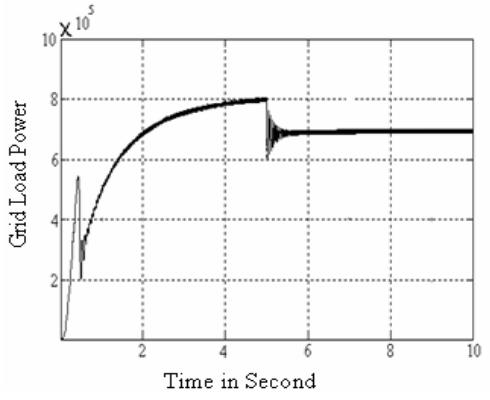


Figure 14. Grid load power flow

Here this is the load power flow of grid. At first grid is supplying about 80kW power to the load. At 5th second of simulation wind turbine starts producing power. In this paper as we are using a 100kW wind turbine so from figure 14 we see that wind turbine is delivering almost 100kW to the grid.

VIII. Conclusion

During simulation we have also observed voltage fluctuations due to different lengths of transmission line. The results are shown in following table.

Table 1 Voltage fluctuation for different length of transmission lines

Wind turbine distances (km)	Voltage fluctuation at grid (V)	Voltage fluctuation at wind turbine (V)
6.5	48	77
8	42	76
10	18	46
13	16	43

The above table shows us that with increasing the length of transmission line voltage fluctuation decreases. So to avoid sudden variation of voltage at the grid terminal wind turbine can be placed at a suitable distance but it can't be placed far away from the grid as attenuation characteristic of transmission line can reduce power flow to the grid. This paper presented modeling and simulation of an isolated hybrid wind diesel system of Cartwright. The above-presented model can be a useful tool for the wind power industry to study the behavior and influence of wind turbines on isolated wind diesel system.

IX. Acknowledgement

The authors would like to thank the National Science and Engineering Research Council (NSERC) Canada for supporting this research.

X. List of symbols and parameters

Table 1 List of symbols:

d	Direct axis
q	Quadrature axis
l	Leakage variable
s	Stator variable
r	Rotor variable
f	Field quantity
m	Magnetizing quantity
k	Damping variable
R	Resistance variable
I	Current variable
V	Voltage variable
ω	Stator angular electrical frequency
ω_b	Rotor angular electrical base frequency
ω_r	Rotor angular electrical speed
X	Reactance variable
L	Inductance variable
φ	Flux variable
λ	Flux linkage variable
p	d/dt
P	No of generator pole
T_g	Electromagnetic torque
T_T	Aerodynamic torque

References

[1] K. Rajambal, B. Umamaheswari and C. Chellamuthu, "Electrical Braking of large wind turbines," Renewable Energy, vol. 30, pp. 2235-2245, Dec. 2005.

[2] Siegfried Heier, *Wind Energy Conversion System,* John Wiley & Sons ltd, 2006.

[3] Tony Burton, David Sharpe, Nick Jenkins, Ervin Bossanyi. *Wind Energy Handbook,* John Wiley & Sons ltd, 2001

[4] Blaabjerg F., Sørensen P., Hansen A.D., Lov F., "Wind Turbine Blockset in Matlab/Simulink-General Overview and Description of Model" ISBN 87-89179-46-3, Institute of Energy Technology, Aalborg University

[5] Ozpineci, B., Tolbert, L.M., "Simulink implementation of induction machine model - a modular approach", Electric Machines and Drives Conference, 2003. IEMDC'03. IEEE International, Volume: 2, pp. 728 – 734, 1-4 June 2003.

[6] K. Rajambal and C. Chellamuthu, "Modelling and Simulation of Grid Connected Wind Electric Generating System," IEEE Trans on Energy Conversion, vol. 3, pp. 1847- 1852, Oct. 2002.

[7] Krause, P.C., Wasynczuk, O. and Sudhoff, S.D."*Analysis of Electric Machinery*", IEEE Press, 2002.

[8] Ong,Chee-Mun, *Dynamic Simulation of Electric Machinery: Using Matlab/Simulink*, Prentice Hall PTR, 1997

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Monte-Carlo and Recency-Weighted Learning Methods for Conjectural Variations in Dynamic Power Markets

P. N. Vali[1], A. R. Kian[2]

[1]Department of Electrical Engineering, KNT University of Technology, Tehran, Iran
[2]Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

E-mail: pnayebvali@gmail.com

Abstract - Conjectural variations based bidding strategies have been proved to be a more appropriate model to analyze bidding profile of players in an electricity market than other game theoretic models. The equilibrium quantities and market clearing prices result from Nash-Cournot equilibrium are far from real markets data. However CV has been criticized for having no definite meaning in static form. In this paper we proposed the dynamic form of quantity setting conjectural variations. Dynamic optimization and conjectures learning which set the collection of dynamic nonlinear state equations are obtained. Monte-Carlo and recency-weighted learning methods are introduced and their effects on equilibrium and MCP are investigated. The results of simulations verify that learning would lead to greater social welfare and more realistic price.

I. Introduction

The competition has been introduced in recent power markets throughout the world, in order to decrease the consumers' price and enhance economic efficiency. The conventional integrated electric utility structures have been restructured and substituted by competitive markets. However, impeding obstacles, like enormous capital investments for new entrants to enter electricity market and time-consuming construction of power plants and also, isolation between consumers and generation companies caused by transmission constraints and transmission losses, have made the new electricity markets more similar to oligopoly markets, than perfectly competitive markets. In oligopoly markets, generation companies can increase their profit through strategic bidding and enforcing market power. Various models are exercised in the literature to model the influence of strategic bargaining, and uncover the ultimate point, in which the market would rest. Dynamic programming, stochastic optimization, Lagrangian relaxation, and some evolutionary methods, are some non game theoretic instances, to be named. Several common methods, observe the strategic bidding as a multi objective decision making in multi agent environment. They predict and evaluate the equilibrium as a set of strategies for agents in the market, such that firms' unilateral deviations are less profitable for themselves (which is known as Nash Equilibrium). Supply function equilibrium is a case in point, where optimal supply functions of firms, that match the aggregate demand, are to be investigated. Cournot model is another widely applied approach in which firms tactically bid on output, to satisfy a predetermined demand function. Cournot's concluding price is hardly justifiable in power markets, due to disregarding rival's supply function. Stackelberg or "leader follower" models categorizes opponents, based on their information. Leaders know how their decisions would affect the followers whose decisions are made, taking leaders' actions for granted. Neglecting the rival's reaction to self alteration of production or supply function is revised in conjecture based methods. Concept of conjectural variations was introduced by Bowely in (1924) and named as conjectural variation by Frisch [1] in (1933). It assumes that oligopolists who choose their actions simultaneously in a single-shot market, each believe that their individual action choices will influence the choices of the other players [2]. Conjectural variation based models have been widely applied in electricity markets modeling [3, 4]. In [5] conjectured supply function has been introduced to shed light on the fact that, firms will adjust production in response to opponents' supply functions altering which are assumed to be linear. General attributes of CV-based methods, make them capable of modeling different types of markets, such as monopoly, perfect competition, and Cournot model. The values of conjectures are mentioned in [6] for different market models. Repeated decisions in markets bore in mind the dynamic conjectural variations [7, 8, and 9] to improve firms' strategic biddings through time. Consistency, existence, and uniqueness of conjectural variations are moot points in the literature, and quite a lot papers are dedicated to explore the proofs, justifications, or modifications to be applied for acquiring apt conjectures. The criticism on inconsistent conjecture variation [10] spurred the current searches for ''Consistent conjectures'' [7, 11]. [11] declares, with quadratic costs for firms, an infinite horizon optimized equilibrium singularly exists with consistent conjectural variations. Objective point of view about conjectural variation stimulates the inspection of its values from the historical market data [8, 12]. Dynamic conjectural variation is discussed in this paper, to put aside criticisms on the defects of static conjectures. Dynamic rules for updating conjectures are presented. The paper is organized as follows. Section 2 presents the

math model for conjectural variation based bidding. Section 3 introduces dynamic model and learning methods for conjectural variations. Section 4 brings forward computer simulation study to elucidate firms' optimal decisions. Conclusions are drawn in Section 5.

II. Mathematical Formulation for Conjectural Variations

As mentioned above one of the broadly applied models, used to model electricity markets is Cournot. But the price that comes out of the Cournot equilibrium is far from the market real price in most cases. Assume an N-firm oligopoly market with price (p) which can be obtained from the total product of the market:

$$p = p(\sum_{i=1}^{n} q_i) = p(Q) \tag{1}$$

Here q_i, means firms' levels of production. And Q shows market total product. Firms endeavor to maximize their profit through setting individual product levels q_i . CV based bidding strategy, can help generation firms improve their bidding and increase their profits in actual electricity markets with imperfect information.

A. Conjectural Variation

In industrial organization theory the CV of a firm is defined as its belief or expectation on rival's reaction to its output. The aim of each firm is to maximize its profit via q_i. The profit of firm i can be calculated through the following equations:

$$\pi_i = Rev_i(q_i) - Cost_i(q_i) \tag{2}$$

$$\pi_i = p * q_i - Cost_i(q_i) \tag{3}$$

The optimal solution should satisfy the first order condition, which implies that marginal revenue should be identical to marginal cost at optimal solution (F.O.C). The price is an explicit function of productions. In conjectural variations theorem it is postulated that productions are implicit functions of each other (in contrast with Cournot model that firm's productions are assumed to be totally independent). Therefore at optimal solution we have:

$$\begin{aligned} MR_i(q_i) &= \frac{\partial(p * q_i)}{\partial(q_i)} \\ &= \left(\frac{\partial(p)}{\partial(q_i)} + \frac{\partial(p)}{\partial(q_j)} * \frac{\partial(q_j)}{\partial(q_i)}\right) * q_i \\ &+ p = MCost(q_i) \end{aligned} \tag{4}$$

The conjecture of firm i , on the response of firm j to its production quantity change is shown as CV_{ij} :

$$CV_{ij} = \frac{\partial q_j}{\partial q_i} \tag{5}$$

Optimizing quantities of firm i can be obtained, if CV_{ij} $(j = 1, \dots, n)$ are known by firm i. $p(Q)$ is assumed to be known by all firms based on historical data, however cost function of each firm is a self-known information.

B. Conjectural Variation Based Bidding Strategy

New method of bidding is proposed for generation firms to improve their strategic behavior through a real electricity market. Inverse demand function is assumed to be linear.

$$p = (e - fQ) \tag{6}$$

(e, f) which get extracted from historical data, are constants known by all firms. It is supposed that each firm has quadratic cost function and rationally behave to maximize its profit. The corresponding optimization problem becomes:

$$\begin{aligned} \max_{q_i} \quad \pi_i &= p(Q)q_i - Cost_i(q_i) \\ &= (e - fQ)q_i - (a_i + b_i q_i \\ &+ \frac{1}{2} c_i {q_i}^2) \end{aligned} \tag{7}$$

The constraints can be regarded as:

$$\begin{Bmatrix} Q = \sum_{i=1}^{n} q_i \\ q_{imin} \le q_i \le q_{imax} \end{Bmatrix} \tag{8}$$

Where a_i, b_i and c_i are coefficients of cost function for firm i. By using the first order condition ($\partial\pi_i/\partial q_i = 0$), solving each equation in terms of q_i , we can acquire:

$$q_i = \frac{e - f\sum_{j=1, j\neq i}^{N} q_j - b_i}{f\left(2 + \sum_{j=1, j\neq i}^{N} CV_{ij}\right) + c_i} \tag{9}$$

As noted down over, we should only know total production, and total conjectures on other firms' changes due to alteration of self production.

$$q_i = \frac{e - f q_{-i} - b_i}{f(2 + CV_i) + c_i} \tag{10}$$

Where

$$q_{-i} = \sum_{j=1, j\neq i}^{N} q_j = Q - q_i \tag{11}$$

And

$$CV_i = \sum_{j=1, j\neq i}^{N} CV_{ij} \tag{12}$$

We can integrate all the equations in matrix form. Then the optimal decisions (q_i) are taken via following equation:

$$A * \vec{q} = B \tag{13}$$

Where A is a $(N+1) \times (N+1)$ matrix of the below form:

$$A = \begin{bmatrix} (f + fcv_1) & 0 \dots 0 & f \\ 0 \dots & (f + fcv_i) \; 0 \dots & f \\ \vdots & \ddots & \vdots \\ 0 \dots & (f + fcv_n) & f \\ 1 \dots . & 1 & -1 \end{bmatrix} \tag{14}$$

And $\vec{q}\ and\ B$ are both $(N+1)\times 1$ vectors which are defined below respectively:

$$\vec{q} = \begin{bmatrix} q_1 \\ \vdots \\ q_i \\ \vdots \\ q_n \\ Q \end{bmatrix} \tag{15}$$

$$B = \begin{bmatrix} e - b_1 \\ \vdots \\ e - b_i \\ \vdots \\ e - b_n \\ 0 \end{bmatrix} \tag{16}$$

Many market structures can be represented by CV based bidding strategies, spreading from perfect competition to monopoly. Perfect competition can be modeled b taking all CV values equal to (-1) (Supposing identical linear cost functions for firms). Cournot model is the case of independent production, where there is no conjecture about other firms' output. Monopoly can be figured as the cooperation of firms. CV values in monopoly are (1) which means firms cooperate together and increase or decrease their production simultaneously and equally (for analogous firms). Other market structures can also be characterized by conjectural variations such as Stackelberg etc. (A comparison of CVBS and different market structures have been represented in [6].)

III. Dynamic Model and Conjectural Variations Learning

Real electricity markets are dynamic cause firms in the market bid over their quantity daily or even hourly. When firms aim to solve their optimization problem, they should know the others quantities. So the firms will assume that their rivals would not change their productions for the new market round. The updating equation for the price of firm i is as follows:

$$q_i^k = \frac{e - f q_{-i}^{k-1} - b_i}{f(2 + CV_i^{k-1}) + c_i} \tag{17}$$

CV values, can be extracted from the data publicized in the market. Several methods for updating conjectural variations could be utilized. Two different methods for updating CV values are introduced in this paper.

A. Monte-Carlo method

In this method the comprehended CV value is not the instantaneous variation, but is the average of all observed variations during market simulation. The updating rule for Monte-Carlo case is equation (18).

B. Recency-Weighted Average

Previous method implicitly assumes that CV is stationary. If we apply the non-stationary format of expectation, we have to consider weighted expectations where recent observed values for CV are more valuable (more weighted), than past CV values. This is sometimes called an exponential, recency-weighted average. Equation (19), present the update rule for the CV values in this case.

$$CV_i^k = \begin{cases} \frac{1}{k} * \frac{q_j^{k-1} - q_j^{k-2}}{q_i^{k-2} - q_i^{k-3}} + \left(\frac{k-1}{k}\right) * CV_i^{k-1} & if\ (q_i^{k-2} - q_i^{k-3} > \varepsilon) \\ CV_i^{k-1} & if\ (q_i^{k-2} - q_i^{k-3} < \varepsilon) \end{cases} \tag{18}$$

$$CV_i^k = \begin{cases} \frac{(1-\lambda)}{(1-\lambda^k)} * \frac{q_j^{k-1} - q_j^{k-2}}{q_i^{k-2} - q_i^{k-3}} + \lambda\left(\frac{(1-\lambda^{k-1})}{(1-\lambda^k)}\right) * CV_i^{k-1} & if\ (q_i^{k-2} - q_i^{k-3} > \varepsilon) \\ CV_i^{k-1} & if\ (q_i^{k-2} - q_i^{k-3} < \varepsilon) \end{cases} \tag{19}$$

The coupled equations of CV and prices make a nonlinear state space system representation. It should be mentioned that when the variation in production falls below the predetermined value (epsilon), then firms would not associate the others firm change to their own alteration.

IV. Computer Simulation Study

This section applies the above mentioned dynamic equations, for simulation of a power market. A duopoly market is considered. The market has been simulated for 20 rounds. Cost function parameters of these firms are listed in Table 1.

Table 1. Coefficients Of Cost Functions			
	a_i	b_i	c_i
Firm 1	0	2	0.005
Firm 2	0	4	0.005

Inverse demand function is assumed to be of the form:

$$p = (100 - Q) \tag{21}$$

Three different cases are studied. First no firm conducts learning and conjectures all are set to zero. Second one of the firms would utilize learning, while the other still maintains its previous strategy. In case 3, both firms conduct learning. Various updating rules for CV values are applied in this case.

A. No Learning

The conjectural variations of firms about each other are all zero. Market will operate according to equation (17). Initial productions of firms are 20 MWh. Generations and Profits of firms are shown in Fig. 1, and Fig. 2, respectively for this case.

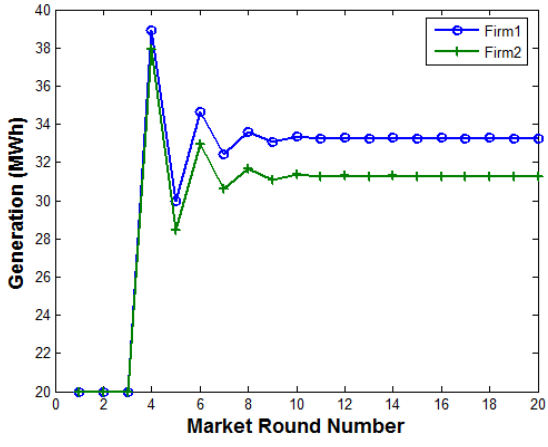


Fig. 1. Generations of firms (case A)

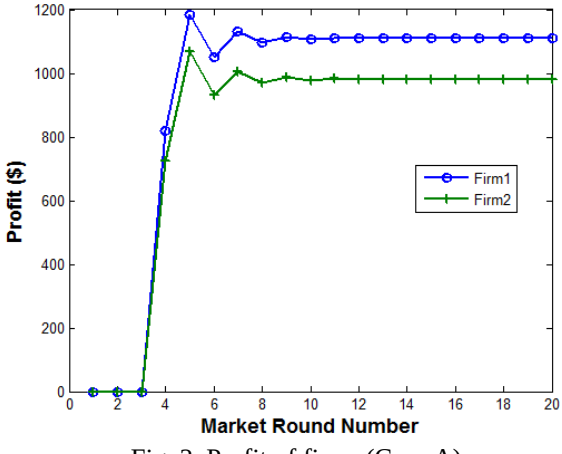


Fig. 2. Profit of firms (Case A)

B. Unilateral Learning

Firm 1 conducts learning in each round of the market, while the other firm maintains its belief that Firm 1 will not react to its bidden production. The generations of firm for this case are shown in Fig. 3. Epsilon in updating rules for CV values is inserted to prevent nonlinear computational difficulties. We can also consider that other firms would not base their decision on little change in opponents' quantities, and the changes in their quantities are related to other market parameters. From now on, wherever this parameter is needed, we take $\varepsilon = 7$.

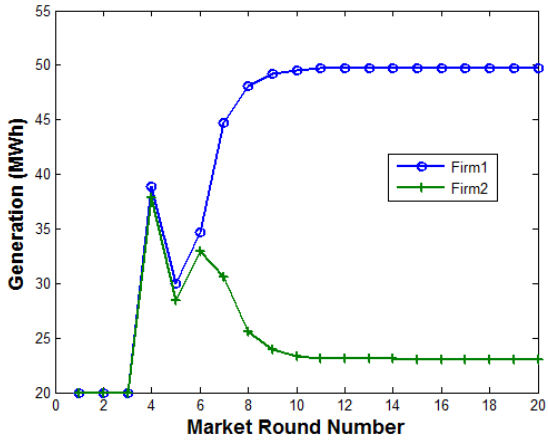


Fig. 3, Generations of firms (case B)

Profits of firms for unilateral learning are shown in Fig. 4.

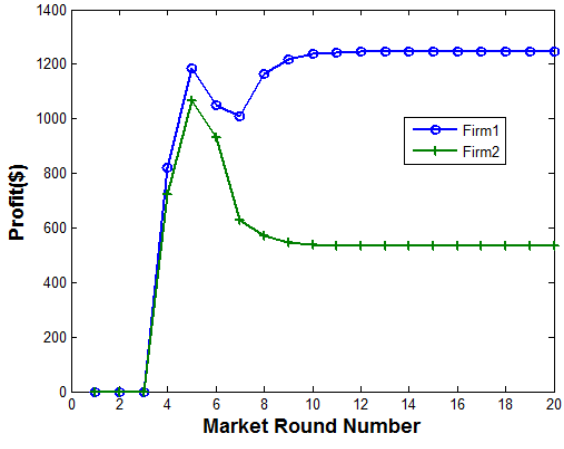


Fig. 4. Profit of firms (Case A)

CV Value that is obtained in this case is as Stackelberg model, where the follower (firm 2), has CV=0, and the leader has CV=-0.5. We should express that learning method in this case is just simple instantaneous observed CV value for each firm. Firm 1 gain more profit than firm 2. It is the effect of learning that gives firm 1 the advantage of being leader, which biases firms to conduct learning.

C. All Firms Learn

We will utilize two different learning methods and simulate the market when all firms learn their CV value. First we used Monte-Carlo learning method. In this method firms will take into consideration all previously observed CV values. The rule for updating CV is as

equation (18). The productions and profits of firms are depicted in Fig. 5 and Fig. 6 respectively.

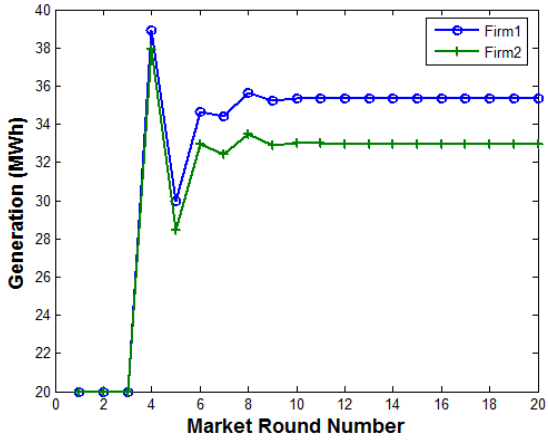


Fig. 5. Generations of firms (Case C1)

We should mention that this way of CV learning would lead to less social welfare but, more profits for the firms in the market. The final CV values for the firms in this case are [-0.1633,-0.1633].

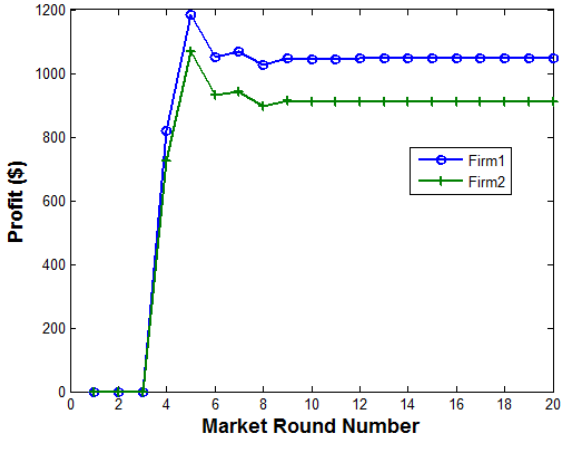


Fig. 6. Profit of firms (Case C1)

Now we change our updating rule and use the recency-weighted average method. Lambda in this method determines the rate of discounting we suppose lambda to be 0.5 for our simulation. The generation and profit of firms are shown in Fig. 7 and Fig. 8 respectively. The final CV value for this case is -0.4207.

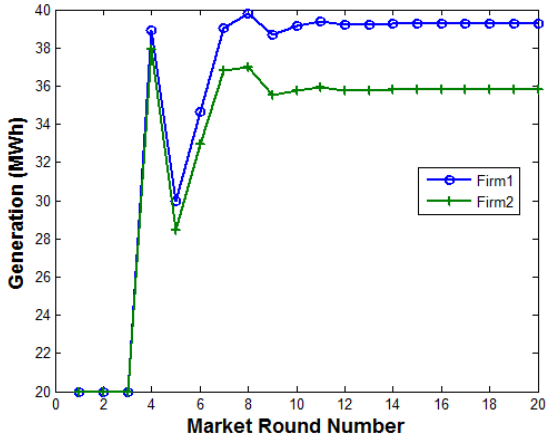


Fig. 7. Generations of firms (Case C2)

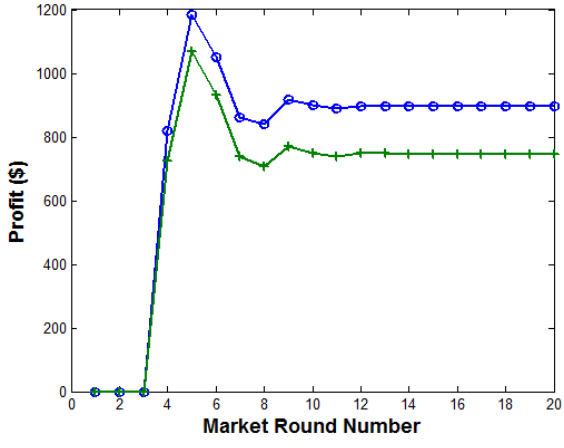


Fig. 8. Profit of firms (Case C2)

This case is more competitive in comparison with case C1. And the social welfare enhances while the firms in the market gain less profit. Table 2 summarizes the results that have been observed in these four different cases. It is shown that with Monte-Carlo method the MCP is higher than case 2, which is equivalent to Stackelberg model. Learning strategy is dominant strategy for firms participating in market although firm's profits get less when all firms in the market learn CV values.

It is called prisoners' dilemma game in the context of game theory. It is more profitable for both firms to refuse to learn but because there exist the possibility of unilateral learning, in this point of mutual strategies, there exist a better strategy for firms and that is to learn, which make the learning dominant strategy for both firms.

Table 2. Profit, Generation, and CV of firms in various cases

	Firm 1			Firm 2			Market Data		
	Profit ($)	Generation(MWh)	CV	Profit ($)	Generation(MWh)	CV	MCP ($/MWh)	Total Generation (MWh)	Total Profit ($)
No learning (A)	1110	33.27	0	981	31.28	0	35.44	64.55	2091
Unilateral learning (B)	1240	46.17	-0.5	619	24.85	0	28.98	71.02	1859
Monte-Carlo	1045	35.36	-0.166	909	32.98	-0.166	31.66	68.34	1954
Recency-weighted ($\lambda = 0.5$)	896	39.25	-0.42	746	35.82	-0.42	24.93	75.07	1642

V. Conclusion

In this paper new model of strategic bidding via learning quantity setting conjectural variations for electricity markets has been brought up. The dynamic approach toward bidding, make it capable of reaching Nash equilibrium in incomplete information. Different learning methods were utilized. And from simulation results it is understandable that every learning method would work better for the consumers than Nash-Cournot outputs (not necessarily Stackelberg outputs.) but there is no chance for firms to learn because it is their dominant strategy. The Monte-Carlo method would not work as effective as recency-weighted average method. Because keeping in mind previous CV values observed, would keep us nearer to the our first guess about the market (which was Cournot initialization in this case.)

References

[1] J. Tirole, The Theory of Industrial Organization, MIT Press, Cambridge, MA, 1989.

[2] J. W. Friedman, and C. Mezzetti, "Bounded rationality, dynamic oligopoly, and conjectural variations," Journal of Economic Behavior & Organization, vol. 49, pp. 287–306, 2002.

[3] A. G. Alcalde, M. Ventosa, M. Rivier, A. Ramos, and G. Relano, "Fitting electricity market: A conjectural variations approach," presented at 14th PSCC, Sevilla, June 2002.

[4] S. L. Haro, P. S. Martin, J. E. de la Hoz Ardiz, and J. F. Caro, "Estimating conjectural variations for electricity market models," European Journal of Operational Research 181, pp. 1322–1338, 2007.

[5] C. J. Day, B. F. Hobbs, and J. S. Pang, "Oligopolistic competition in power networks: A conjectured supply function approach," IEEE Trans. Power Syst., vol. 20, no. 3, pp. 597-607 August 2002.

[6] Y. Song, Y. Ni, F. Wen, Z. Hou, and F. F. Wub, "Conjectural variation based bidding strategy in spot markets: fundamentals and comparison with classical game theoretical bidding strategies," Electric Power Systems Research 67, pp. 45-51, 2003.

[7] J. Itayaa, and K. Shimomura, "A dynamic conjectural variations model in the private provision of public goods: a differential game approach," Journal of Public Economics, vol. 81, pp. 153–172, 2001.

[8] J. D. Liu, T. T. Lie, and K. L. Lo, "An empirical method of dynamic oligopoly behavior analysis in electricity markets," IEEE Trans. Power Syst., vol. 21, no. 2, May 2006.

[9] Y. Song, Y. Ni, F. Wen, and F. F. Wu, "Conjectural variation based learning model of strategic bidding in spot market," Electrical Power and Energy Systems, vol. 26, pp. 797–804, 2004.

[10] T. Lindh, "The inconsistency of consistent conjectures: coming back to Cournot", Journal of Economic Behavior and Organization 18, 1992.

[11] Y. Liu, Y. X. Ni, Felix F. Wu, and Bin Cai, "Existence and uniqueness of consistent conjectural variation equilibrium in electricity markets," Electrical Power and Energy Systems 29, pp. 455–461, 2007.

[12] J. D. Liu, and T. T. Lie, "Empirical dynamic oligopoly behaviour analysis in electricity markets", International Conference on Power System Technology (POWERCON), November 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Research for Data acquisition equipment with micro-Grid system

Tae-young Lee, Kwang-ho Ha, Hyun-jea Yoo, Jong-wan Seo and Myong-chul Shin

School of Information and Communication Engineering, Sungkyunkwan University
300 ChunChundong, Jangangu, Suwon 740-446, Korea
E-mail: samshinsmd@chol.com

Abstract – Micro-Grid system, which is connected with Power system, should not affect on Power quality. Therefore, Micro-Grid system needs to control system to minimize effect of power quality. In this paper represents the measurement equipment for voltage/current of micro-grid system and the result. This system monitors the state of Micro-Gris system, such as steady state, islanding operation and reconnection with conventional power system. The equipment has real-time measure function and communication with PC via USB port. This equipment can use various purposes such as investigation of transient state, monitoring of real-time operation and evaluate the power system quality.

I. Introduction

The Micro-Grid is a small scale power generation system. Some of the Micro-Gird system is connected with distribution line and existing transmission line. Moreover Micro-Grid is power generation system of new form that can the independent operation[1]. However Micro-Grid has some problems. For example, if fault occurred in power system or in Micro-Grid, it may cause the power quality problem. Also it may happen to a trouble between provider and customer. So we need facility for data acquisition and monitoring of micro-Grid system to analyze and estimate. Because of high price, this equipment has difficult in use of small scale load.

Accordingly, in this paper describes that developed equipment of measurement system to simple and cheap for Micro-Gird. The equipment of measurement and storage system is the monitoring at real-time and stores an operational condition. It measures an each voltage and electric current at real-time.

And it uses USB ports and stores the acquired data to hard disk drives in general PC. In generally, the hard disk is cheaper than the flash memory which is inside the instruments. And stores a data the many time. But the data storage capability is bigger than the flash memory. That will be able to store a many data.

In order to monitor the operational condition, it measures currents and voltage at real-time. It measures a voltage and an electric current at real-time. And it confirms the operational condition of the equipment. So the data which is converted with digital 128 counts the above is necessary in 1 period. The many data is needed for measures a voltage and an electric current and in order transmits about three phase power. In power system applied a Micro-Grid, this paper describes acquiring equipment about load, Distributed Generation, Voltage and Current of Power system so that make a model about load and distributed device.

In this paper, in order to storage as possible as many rapidly and many data to use USB pots and in order to store a data in the hard disk. This system acquires voltage and current with 128 samples per a cycle. The acquired data with this equipment is useful to make a model about load and distributed generation system.

II. Micro-Grid System

To realize the emerging potential of distributed generation one must take a system approach which views generation and associated loads as a subsystem or a Micro-Grid[2].

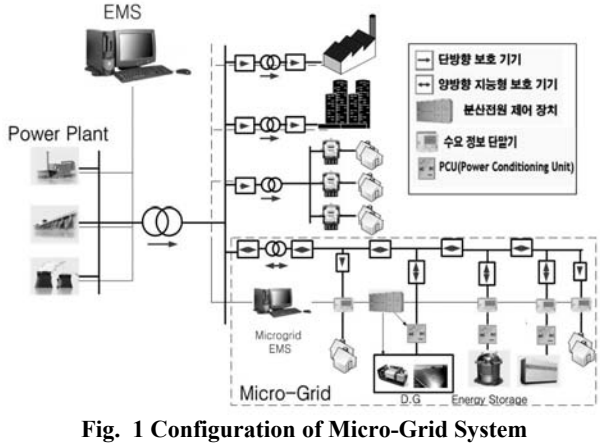


Fig. 1 Configuration of Micro-Grid System

During disturbances, the generation and corresponding loads can separate from the distribution system to isolate the micro-Grid's load from the disturbance (and thereby maintaining service) without harming the transmission grid's integrity.

The difficult task is to achieve this functionality without extensive custom engineering and still have high system reliability and generation placement flexibility.

To achieve this we promote a peer-to-peer and plug-and-play model for each component of the micro-Grid. The peer-to-peer concept insures that there are no components,

such as a master controller or central storage unit that is critical for operation of the micro-Grid.

This implies that the micro-Grid can continue operating with loss of any component or generator. With one additional source (N+1) we can insure complete functionality with the loss of any source. Plug-and-play implies that a unit can be placed at any point on the electrical system without reengineering the controls. Plug-and-play functionality is much akin to the flexibility one has when using a home appliance.

That is it can be attached to the electrical system at the location where it is needed. The traditional model is to cluster generation at a single point that makes the electrical application simpler.

The plug-and-play model facilitates placing generators near the heat loads thereby. Allowing more effective use of waste heat without complex heat distribution systems such as steam and chilled water pipes. This ability to island generation and loads together has the potential to provide a higher local reliability than that provided by the power system as a whole. Smaller units, having power ratings in thousands of watts, can provide even higher reliability and fuel efficiency.

These units can create micro-Grid services at customer sites such as office buildings, industrial parks and homes. Since the smaller units are modular, site management could decide to have more units (N+) than required by the electrical/heat load, providing local, online backup if one or more of the operating units failed. It is also much easier to place small generators near the heat loads thereby allowing more effective use of waste heat.

Basic Micro-Grid architecture is shown in figure 1[2]. This consists of a group of radial feeders, which could be part of a distribution system or a building's electrical system.

There is a single point of connection to the utility called point of common coupling. The non-critical load feeders do not have any local generation. When there is a problem with the utility supply the static switch will open, isolating the sensitive loads from the power grid. Feeder D loads ride through the event. It is assumed that there is sufficient generation to meet the loads' demand. When the Micro-Grid is grid-connected power from the local generation can be directed to feeder D. following section describes the style and format of the submitted paper.

It is to be noted that, submission of paper in the final camera ready format is required. The submitted paper, subject to acceptance on the basis of reviewers' report, will be included in the conference proceeding without any modifications. Your adherence to the format and style specifications described in this section is required to maintain uniformity of appearance throughout the proceedings.

III. Concept and Equipment of Measurement System

A. Concept of Measurement System

The equipment of measurement is composed with measurement system and general PC. The measurement system is composed to a part of measurement and communication. This chapter describes the structure of measurement unit and software in PC side.

B. Equipment of Measurement System

The measurement unit converts the analog voltage and current to digital value, and calculates power and phase. And also, the unit transmits the measured values to general PC. Figure 2 shows the structure of measurement unit.[3]

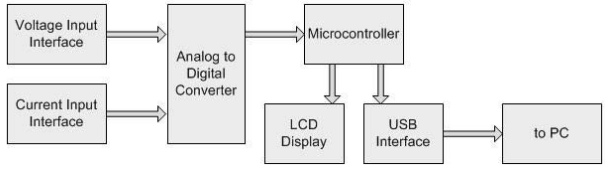


Fig. 2 Structure of the measurement unit

The major function of this unit is that it measures voltage and current, converts the analog value to digital data and transmits to general PC for more complex calculation, applying advanced analysis algorithms.

Measurement function acquires voltage data of 220V/110V at 60Hz and current value from 1mA to 200A. Figure 2 and Figure 3 show voltage and current measurement interface circuit.[3]

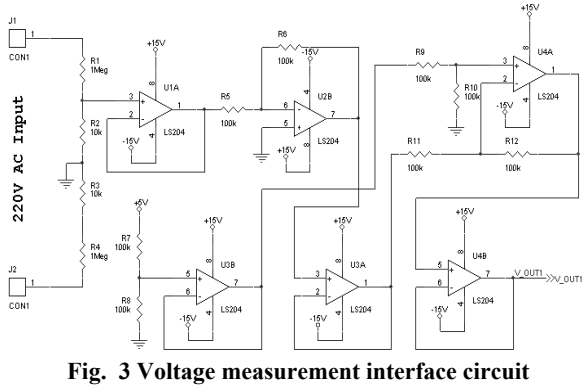


Fig. 3 Voltage measurement interface circuit

Figure 2 shows voltage scaling and polarization conversion circuit. The resolution of ADC is 10-bit; therefore input voltage is equation 1.

$$V_{IN} = 101 \times (5 \times ADC / 1024 - 2.5) \quad (1)$$

If the voltage range of ADC is from 0V to 5V, then the range of input voltage (V_{IN}) is from -252.5V to 252.5V, from the equation 1[3].

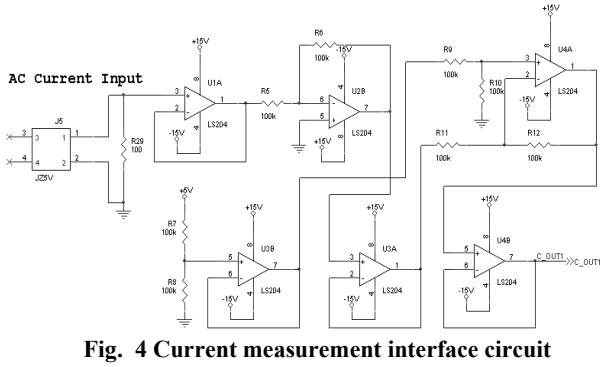


Fig. 4 Current measurement interface circuit

Current transformer (CT) interface circuit is in Figure 4. The ratio of CT is 2500:1. The input voltage of ADC is equation 2, that value depends on current and sensing resistor.

$$V_{ADC} = 2.5 - i_{CT\_OUT} \times R_{SENSE} \times (-1) \quad (2)$$

If the value of sense resistor is 100 ohm, the current is equation 3.

$$i_{IN} = 125/1024 \times ADC - 62.5 \quad (3)$$

From the equation 1, the minimum resolution of voltage is about 0.5V that is 1-bit change of ADC value. And also, from the equation 3, the minimum detectable current variation is 0.122A[3].

The measurement unit collects the data of voltage and current 128 samples per cycle. ADC that is used in this paper is included in MCU (microcontroller unit), which has maximum 8 channels, 10-bit resolution ADC. Then transmit these data to the general PC via USB interface.

IV. System Application of Micro-Grid System about Measurement System

Figure 5 shows skeleton structure of Micro-Grid for photovoltaic(PV) power generation equipment and Power system. And PV system is presented by Micro-Source. The PV and general power source are composed to automatic or manual operation in order to use the switching. This system connects a PT and CT in digital WHM parts and measures a voltage and an electric current. A part the experiment is embodied by our the laboratory where was developed[8].

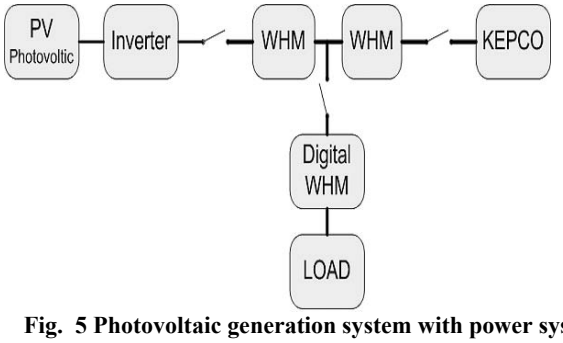


Fig. 5 Photovoltaic generation system with power system measurement interface circuit

The programs in general PC are many for various applications such as EXCEL, ORGIN or MATLAB, etc. The measurement unit transmits the measured and calculated data to the general PC such as voltage, current, power and power factor. Therefore general PC is required bridge program measurement unit with application programs. It calls communication driver. The communication driver stores data from measurement unit to storage device and relays to application program.

The application program displays the measured data, calculated data and apply filter algorithm which is filters such as fast fourier transform (FFT), wavelet and kalman filter and is power protection function such as over/under frequency, voltage and over current detection and output the trip signal for switches.

Figure 6 shows voltage data from measurement unit and Figure 7 shows FFT result with Origin by McroCal's.

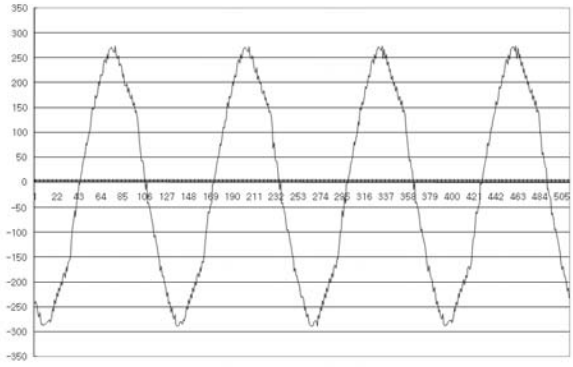


Fig. 6 Voltage data from measurement unit

The data in figure 6 is measured in laboratory environment that has many SMPS devices and dimmer light therefore the data has harmonics and distorted.

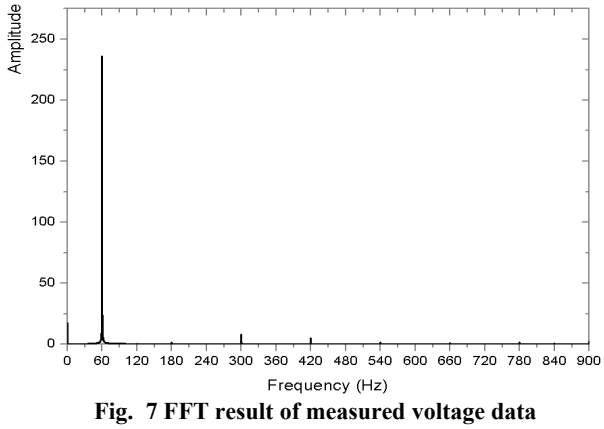


Fig. 7 FFT result of measured voltage data

Figure 7 shows FFT result of measured voltage data. That represents that the measured voltage data contains 5-th and 7-th harmonics. Collected data is stored computer hard disk for a long time. Also we can analysis collected data signal to use data analysis program. We can made power system model for Micro-Grid system analysis using collected data. This model used usefully where Micro-Grid system analysis and prediction.

V. Discussion

The Micro-Grid system is a small scale and located near the power consumer. Some Micro-Grid system is operated with conventional power systems. In this case, some problems can occur such that degradation of the power quality and islanding operation. These problems lead power consumer to argue about a cause with electric power company.

Therefore the need to monitoring and data storage system which is measured from the Micro-Grid is increase. However this system is so expensive.

In this paper propose the Micro-Grid monitoring and data storage equipment, that is a simple and chipper than conventional systems. This equipment measures the voltage/current at real-time and sends to the general PC with USB communication port. If the over voltage, over current and under voltage condition is occurred, the equipment detect and notice this situation to users. And also, the equipment sends the measured data to the PC. The general PC calculates more complicate operations, which are analysis of harmonics and transform the data with wavelet or FFT. This equipment uses hard disk in the general PC to store the data instead of flash memory that leads to record more and long time duration data. This system acquires voltage and current with 128 samples per a cycle. The acquired data with this equipment is useful to make a model about load and distributed generation system.

VI. Conclusion

This paper represents the development procedure and analysis example that is real-time voltage/current measurement, storage and analysis equipment. This equipment works with general PC and its hard disk for data storage to long time duration. This system is simple and chipper than conventional data acquisition equipment with flash memory storage system. The first aim of this equipment is monitoring the state of Micro-Grid connected power system. Moreover stored data can use to make model of Micro-Grid system. The stored data is sampled 128 times every cycle. Therefore this data is useful to make model of power system, which consists of Micro-Grid and conventional Power system.

By using described equipment in this paper, we can acquire a data of load connected between PV and Power system. If we acquire materials about lots of case, we will make the precise model of Micro-Grid connected power system.

References

[1] KWang-Myuong Son, Kye-Byung Lee, "A Study On the Stbility of Micro-Grid System", Journal of the Korean institute of illuminating and electrical installation engineers, 21(7), pp.46-53 August 2007

[2] Microgrid: A Conceptual Solution, Robert H. Lasseter, Paolo Piagi, University of Wisconsin-Madison, Madison, Wisconsin, PESC'04 Aachen, Germany 20-25 June 2004

[3] Research for Experimental Equipment of Efficient Power Education, Jong-Wan Seo, Myong-Chul Shin, ICEE2008, 8 JUN 2008

[4] The status of Electrical education course, Hae-won Yang, Magazine of KIEE, 1598-4613, No. 37, Vol. 11, pp.9-16, 1988

[5] The status and direction of Electrical Education, Heung-seok Yang, Magazine of KIEE, 1598-4613, No. 37, Vol. 11, pp.4-8, 1988

[6] Electrical education the point of view with Power Suppliers, Dae-ho Chung, Magazine of KIEE, 1598-4613, No. 44, Vol. 8, pp.27-29, 1995

[7] The direction of electrical engineering education at the industrial area, Jong-gu Kim, Magazine of KIEE, 1598-4613, No. 44, Vol. 8, pp.30-31, 1995A. B. Author, C. Author, and D. E. Author, "Title of conference paper," in Proc. ICECE 2002, Dhaka, Bangladesh, pp. 101–104, 20-22 December 2002.

[8] Research for Implementation of Real-time Power Measurement and Storage System, Jong-wan Seo, Hee-Seok Suh, Myong-Chul Shin, Journal of KIIEE, No. 21, Vol. 7, pp.29-36, 2007. 8

[9] The Present Situations of Engineering Education and Accreditation System in Korea, Sung-gun Kang, Tae-Cheon Rho, Seung-Yeon, Hahm, Cheong-Sig Kim, Journal of kseett, No. 9, Vol. 2, pp.21-33, 2006.6

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Furan Measurement in Transformer Oil by UV-Vis Spectral Response Using Fuzzy Logic

Sin Pin Lai, Dr. Ahmed Abu-Siada, and Prof. Syed Islam

Department of Electrical and Computer Engineering, Curtin University of Technology, Perth 6845, Australia
sinpin.lai@postgraduate.curtin.edu.au

Abstract – Furan derivatives (2-acetyl furan, 2-furfural, 5-methyl-2-furfural, 5-hydroxymethyl 2-furfural, and 2-furfurylalcohol) presence in transformer oil are the key indicators of the solid dielectric degradation. In recent technology, such deterioration is measured according to ASTM D 5837 standard. However, the method requires very expensive and sophisticated equipment like High-Performance Liquid Chromatography (HPLC) or Gas Chromatography-Mass Spectrometry (GC/MS). This paper proposes a low cost and convenient approach in furan contents estimation based on UV-Vis spectrometry and Fuzzy Logic Theory.

Index Terms- Furans; Furfural; UV-Vis Spectral Response; Fuzzy Logic

I. Nomenclature

ASTM	American Society for Testing and Materials
GC/MS	Gas Chromatography-Mass Spectrometry
HPLC	High-Performance Liquid Chromatography
UV-Vis	Ultraviolet-to-Visible

II. Introduction

Furan derivatives are the major degradation products from the chain scission of cellulosic insulating paper that presence in the insulation oils of operational transformers as shown in Figure 1. The state of degradation in the paper determines its mechanical properties and hence its ability to effectively operate as an insulator. As the paper is not accessible in an operating transformer, sampling of oil provides a convenient approach to access the condition of the paper [1]. Furanic compounds are identified and quantified by HPLC or GC/MS in accordance to ASTM D 5837 [2]. Figure 2 shows those furanic compounds that are detectable in transformer oil that has been depolymerises from the deterioration of the cellulose chains.

HPLC and GC/MS are accurate and reliable in identifying and quantifying furan contents. However, the method is expensive and required trained personnel in conducting the experiments. This paper proposes a potential alternative in estimating furan contents at lower cost and convenient approach compare to conventional method.

III. Experimental Setup

Series of experiments were conducted on short term laboratory accelerated aged and in-service transformer oil. The experiment had been done in accordance to the standard of IEC 61125 oxidation stability test of oil, ASTM D 923 standard practices for sampling electrical insulating liquids, ASTM D 5837 (modified) for furanic compounds measurements and ASTM E 275 as a guidance in generating experimental setup for UV-Vis spectral response analysis.

A. IEC 61125 Unused Hydrocarbon-based Insulating Liquids-Test Methods for Evaluating the Oxidation Stability

Laboratory aged insulating oil is prepared by utilising the heating process available in IEC 61125. Section of new paper (length 280mm) were cut and wrapped around copper strips (3mmx10mm). They were then impregnated in 25ml of new transformer oil (shell Diala B). All samples were heated up to 100˚C±0.5˚C in a thermostatically-controlled aluminium alloy block heater for 7 days. Oxygen flow at 1l/h is supplied into each tube to further accelerate the aging process.

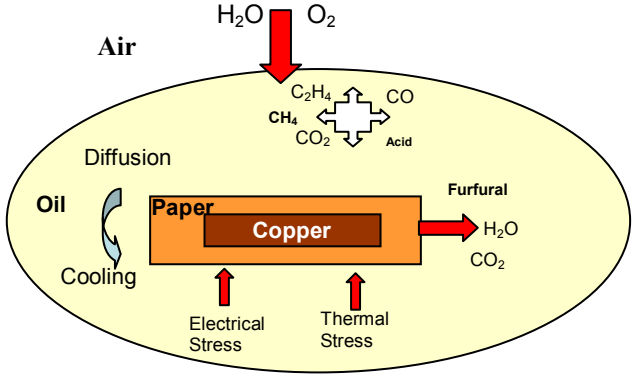


Fig. 1 General schematic representation of degradation mechanism in transformer insulation [3]

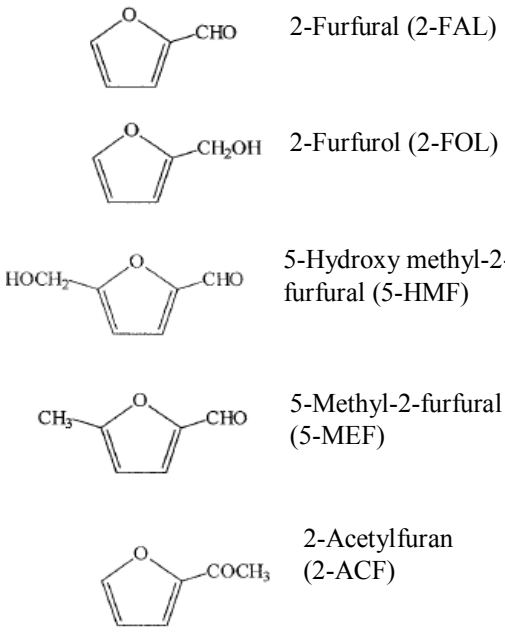


Fig. 2 Furanic compounds detectable in transformer oil that has been in contacted with degraded cellulosic insulation [1]

B. ASTM D 5837 (modified) Standard Test Method for Furanic Compounds in Electrical Insulating Liquids by Gas Chromatography-Mass Spectrometry

All oil test samples were tested for products from the degradation of cellulosic materials from the paper. Four samples of in-services transformer oil originated from the South East Asia country, which believed to contain high level of furan concentrations, were also taken for furan contents test. All samples were prepared in accordance to standard ASTM D 5837. The extracted portion is then injected into a GC/MS system for furan derivatives identification and quantification.

C. ASTM E 275 Standard Practice for Describing and Measuring Performance of Ultraviolet, Visible, and Near-Infrared Spectrometers

All samples that were tested for furan contents by GC/MS were also tested with UV-Vis spectrometer. The experiment procedure was set up in reference to ASTM E 275. The wavelength of spectral response to be investigated was between 200nm to 1100nm. All the oil test sample used in the experiment were all handled according to the standard of ASTM D 923 as it is very essential from the standpoint of preserving the originality of oil quality.

IV. Result and Discussion

The samples of laboratory accelerated aged transformer oil and four samples of chosen in-service transformer oil with different furan concentration were tested in GC/MS and UV-Vis spectrometry analysis. Table I provides the furan derivatives measurement by GC/MS.

It can be seen from Table I that the laboratory aged transformer oil for 7 days is insufficient to generate a significant amount of furan concentration to be detected. Four samples of in-service transformer oil with high furan concentration collected from 150KVA power transformer age between 10 to 20 years old were selected for this research. Table I shows that the main constituent of furan derivative is 2-furfural (2-FAL). It is evident that 2-FAL is the highest solubility in the oil and consistent with the possibility of using it as an oil degradation indicator in the diagnosis of transformer cellulosic insulation condition monitoring [4].

TABLE 1 Test Result of Furan Derivatives Concentration in Part Per-Million (ppm)

Test Sample	2-FAL	2-FOL	2-ACF	5-MEF	5-HMF
Laboratory aged	<0.01	<0.01	<0.01	<0.01	<0.01
In-service	3.1	<0.01	0.01	0.01	<0.01
In-service	5.1	<0.01	0.02	0.01	<0.01
In-service	10.0	<0.01	0.03	0.03	<0.01
In-service	15.0	0.01	0.05	0.05	<0.01

The same samples of oil and new transformer oil were tested with UV-Vis spectrometer. The results show an excellent consistency between the UV-Vis spectral responses and the furanic compound concentrations as shown in Fig. 3. Figure 3 shows that the absorption peaks and bandwidth increase along with the increasing of 2-FAL concentration. The spikes that arise at wavelength 350nm are attributed to the fact that furanic derivatives have no absorption properties at this wavelength.

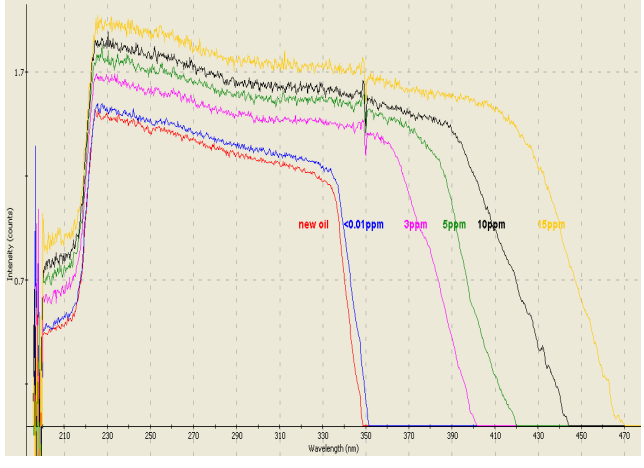


Fig. 3 UV-Vis Spectral response of transformer oil with 2-FAL concentration at 0, <0.01, 3, 5, 10 and 15 parts per-million (ppm)

Dilution test was conducted on the available collection of transformer oil to generate oil sample at various furans concentration. The dilution test was made based on the equation:

$$\frac{Desired\ ppm}{Original\ ppm} \times 8.89g \qquad (1)$$

The equation is calculated based on the transformer oil density of 0.889g/ml at 20°C. New samples were prepared in 10ml volumetric flask that weighted on a

laboratory weighing scale and the diluted sample is as shown in Table 2.

Table 2 Diluted sample at other furan concentration

Origin	Diluted Sample
3ppm	0.5ppm, 1.0ppm
5ppm	4ppm
10ppm	7ppm
15ppm	11ppm,12ppm, 3ppm

Table 2 above shows the oil sample at other furans concentration diluted from the available collection of transformer oil. These samples of oil were scanned with UV light and projected the UV-Vis spectral response as illustrated in Figure 4.

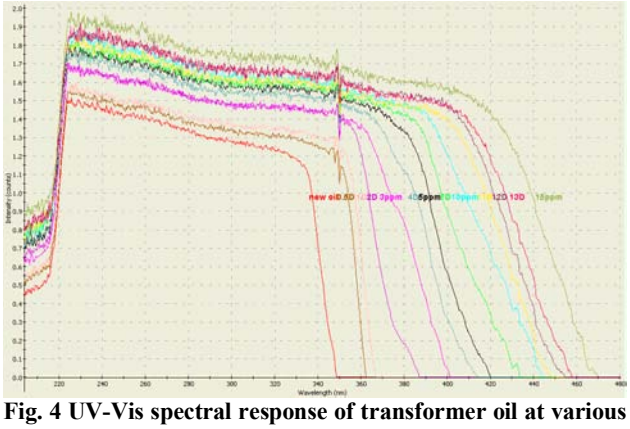


Fig. 4 UV-Vis spectral response of transformer oil at various furan concentrations

Figure 4 above illustrates the UV-Vis spectral response of all the insulating oil test samples. As more than 99% of furan contents was dominated by 2-FAL as mentioned in Table I, it is therefore logical to state the UV-Vis spectral response will be mainly contributed by 2-FAL. In consistent with Figure 3, Figure 4 has shown that spectral response bandwidth and first absorption peak increased in correlation with the furan contents. These characteristic have indicated that the wavelength bandwidth and the absorption peak provide an excellent indication of furans concentration as proven in the Figure 5 below.

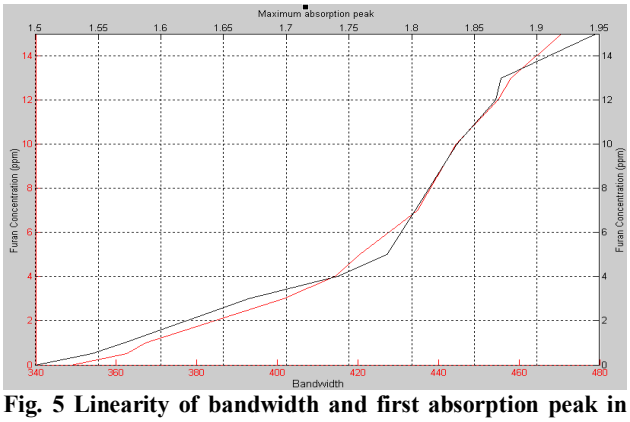


Fig. 5 Linearity of bandwidth and first absorption peak in indicating furan concentration level

Figure 5 clearly demonstrates the correlation between each parameter with respect to furan contents and also the linearity of slope for both curves. The red line is the wavelength bandwidth with respect to furan contents; whereas, the dark line is the first absorption peak at different furan concentration levels. Such characteristics provide useful parameters to generate a software model that able to estimate furan contents at much more convenient approach compare to conventional method.

V. Fuzzy Logic Modelling

Fuzzy logic in simpler sense is a "soft-computing" methodology that has variable whose values are words that based on fuzzy if-then rule. Such theory is particularly useful in developing the software model to estimate the furan concentration based on the input of wavelength bandwidth and first absorption peak.

Due to the fact that the UV-Vis spectral response varied slightly in accordance to ambient temperature, test samples were scanned repeatedly in the morning with room temperature of 20°C and in the afternoon of 25°C. Therefore the range of variation for each parameter at its corresponding furan concentration was taken into account in designing the software model.

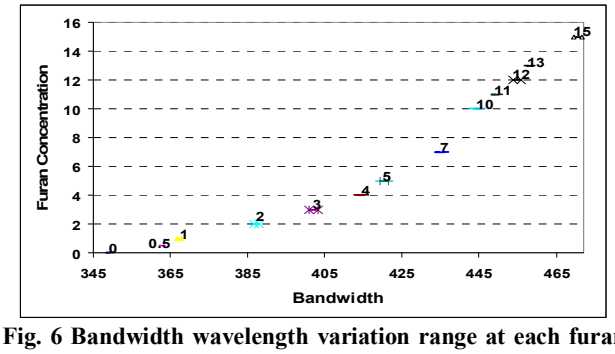


Fig. 6 Bandwidth wavelength variation range at each furan concentration at 20°C and 25°C

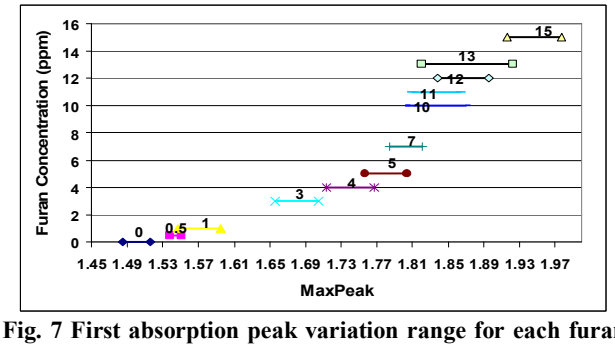


Fig. 7 First absorption peak variation range for each furan concentration at 20°C and 25°C

Figure 6 and 7 illustrate the variation range for bandwidth wavelength and first absorption peak at each level furan concentration at 20°C and 25°C. By including the variation range in the designing the model, it increases the credibility of the furan content estimation model.

Furan content estimation Fuzzy Logic model was developed in according to Fuzzy Inference flow chart as shown in Figure 8. The model built has an interface as shown in Figure 9. The model is able to aggregate the corresponding estimated furan concentration with more

than 95% accuracy based on all the rules that were predetermined with respect to the inputs from the bandwidth wavelength and first absorption peak. The three-dimensional curve that represents the mapping from bandwidth wavelength and absorption peak to furan concentration is shown in Figure 10.

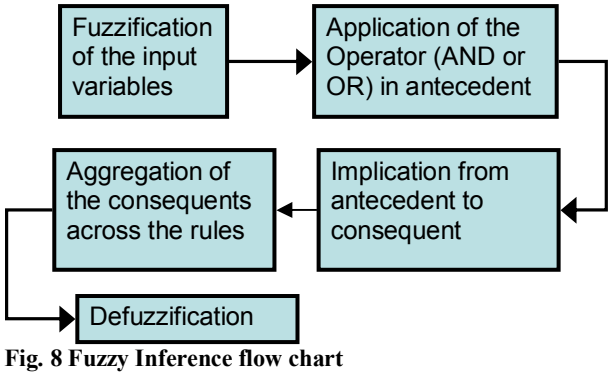


Fig. 8 Fuzzy Inference flow chart

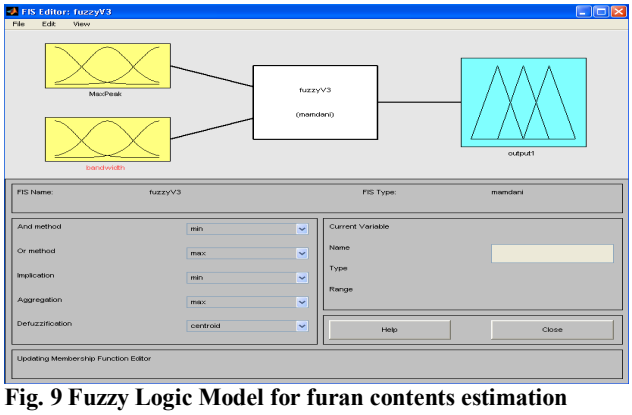


Fig. 9 Fuzzy Logic Model for furan contents estimation

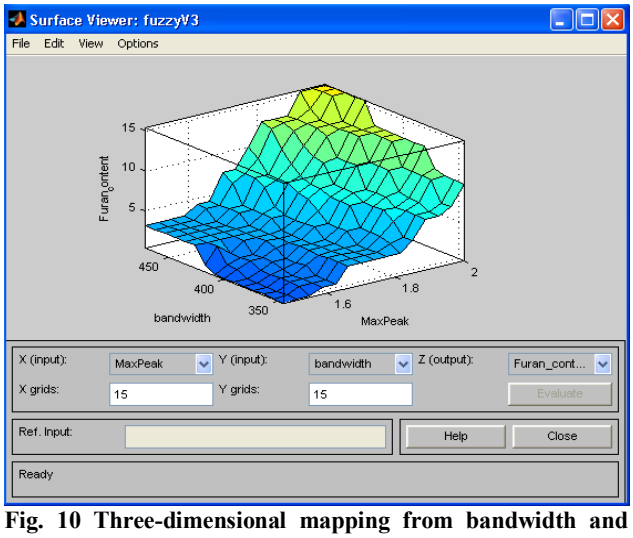


Fig. 10 Three-dimensional mapping from bandwidth and absorption peak to furan concentration

VI. Comparison between ASTM D 5837 and UV-Vis Spectral Response Analysed with Fuzzy Logic

HPLC or GC/MS used in ASTM D 5837 has been very accurate and reliable in identifying and quantifying furan derivatives. However, the method is very expensive, time-consuming and trained personnel are required in conducting the experiment. With UV-Vis spectral response analysis in combination with Fuzzy Logic Model, it provides an alternative for Utilities Company in conducting furan test themselves without outsourcing the test to other company that would lower the cost of maintenance. The benefits in using the method suggested in this paper with comparison to GC/MS is summarised in Table 3.

TABLE 3 Comparisons between GC/MS and UV-Vis Spectrometry with Fuzzy Logic

Method	GC/MS	UV-Vis Spectrometer
Test method	Requires trained person to prepare the sample and handle the experiment. Oil samples are required to be pre-treated with acetone	Doesn't require trained person to conduct the experiment Oil samples don't have to be pre-treated with any chemical reagent
Time taken to conduct the test	>4hrs	Instantly
Price of Equipments	>AUD $50,000	< AUD $12,000
Accuracy	Very accurate. Able to identify the concentration for every single furan derivatives	Shows the response as a whole. However, 99% of furan concentration is determined by 2-furfural which mean the test is very reliable and accurate.

VII. Conclusion

In this paper, the UV-Vis spectral response is proved to be correlated with the furanic compounds concentration. With its advantages in providing more convenient, low cost and instant result in furan measurement, this method has the potential to be commercialized and serves as a reliable and alternative diagnostic technique for oil's furan concentration.

VIII. Acknowledgement

The author gratefully acknowledges the contributions of Gary Lenco (TACS) and Ming Zaw (SGS) for their supervision and facilities provided in this research. This research is supported by Cooperative Research Centre for Integrated Engineering Asset Management (CIEAM).

References

[1] G. C. M. A. W. T. John Scheirs, "Origin of furanic compounds in thermal degradation of cellulosic insulating paper," *Journal of Applied Polymer Science,* vol. 69, pp. 2541-2547, 1998.

[2] ASTM, "Standard Test Method for Furanic Compounds in Electrical Insulating Liquids by High-Performance Liquid Chromatography (HPLC)," *D 5837 - 05,* 2005.

[3] P. J. Baird, H. Herman, and G. C. Stevens, "Non-destructive condition assessment of insulating materials in power transformers," 2005, pp. 425-428 Vol. 2.

[4] G. C. W. T. M. A. John Scheirs, "Study of the mechanism of thermal degradation of cellulosic paper insulation in electrical transformer oil," *Die Angewandte Makromolekulare Chemie,* vol. 259, pp. 19-24, 1998.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Application Specific Integrated Circuit for Optimization of Fixed Polarity Reed-Muller Expressions

Tahseen Kamal [a], Mozammmel H. A. Khan [b]

[a] Department of Electrical and Electronic Engineering, East West University
[b] Department of Computer Science and Engineering, East West University
43 Mohakhali, Dhaka-1212, Bangladesh
E-mail: tahseen@ewubd.edu, mhakhan@ewubd.edu

Abstract - EXOR-based logic circuits have become more popular than AND-OR circuits because they have some specific advantages over AND-OR realizations. Two-level AND-EXOR logic is one of the EXOR-based logics, which is also known as Reed-Muller logic. A Fixed Polarity Reed-Muller (FPRM) expression is one of the seven classes of AND-EXOR logic expressions. An FPRM expression is canonical and uses a fixed polarity for each variable. An *n*-variable function has 2^n different polarity vectors; consequently, there are 2^n different FPRM expressions. The expression with minimum number of products is the minimum FPRM expression. Therefore, the minimization problem of FPRM expressions is to find a polarity vector that produces an FPRM expression with minimum number of products. There are many software methods for FPRM minimization which are sequential in nature and require exponential execution time. In this work an ASIC has been developed to minimize 3-variable FPRM expressions which is parallel in nature and requires constant time. This ASIC takes the minterm coefficients of a Boolean function as input. It generates all the polarity vectors for a three variable function and determines the optimum polarity and corresponding FPRM coefficients.

I. Introduction

The most popular way a Boolean function can be represented is a *truth table* representation. The size of the *truth table* increases exponentially with the increase of *n* (number of variables in the function). Another commonly used approach is the *AND-OR* representation, also known as the *Sum-of-Products (SOP)* representation which is more compact than the *truth table* representation. During the last two decades, researchers focused their eyes extensively on realizing logic functions using *EXOR-based* circuits which is more compact than the *AND-OR* realization. For example, for representing a parity function an *AND-OR* representation takes 2^{n-1} product terms, whereas *AND-EXOR* representation takes *n* product terms [1].

Logic circuits may be minimized as AND-OR expression using established techniques such as the K-map, Quine-McCluskey method, Espresso, etc. The starting point, generally, is the SOP forms, and the aim is to reduce the number of terms/literals. The minimization of Boolean functions can also be done as AND-EXOR expressions. One of the AND-EXOR expressions is the Fixed Polarity Reed-Muller (FPRM) expression which has the property that the polarity of a variable remains same throughout the expression, which eases the implementation of the expression in VLSI. The aspects for which researchers are interested in working on FPRM minimization are that Boolean matching, symmetry of Boolean functions can be detected and also Boolean functions can be classified using FPRM representation as a tool [2]. If an FPRM expression is an *n* variable function then there are 2^n different polarity vectors. So there are 2^n distinct FPRMs for an *n* variable function. Different expressions will have different number of products. An expression with minimum number of products is the minimum FPRM expression for a given function. Therefore, the minimization problem of FPRM expressions is to find a polarity vector that produces an FPRM expression with minimum number of products. Many exact and heuristic software approaches are available to minimize FPRM expressions for both completely and incompletely specified functions [3, 4, 5, 6, 7, 8, 9]. FPRM expressions have signal processing applications also [10, 11, 12]. For real time applications the software approaches are not efficient as they all take exponential time for computation. The application specific integrated circuit (ASIC) may be an alternative approach, which will take constant time to generate FPRM coefficients.

A semicustom IC was designed by Almaini to generate optimum generalized Reed-Muller (GRM) expansions [3]. The paper [3] explains the theory and design of a semicustom integrated circuit (IC) for the generation of the optimum polarity of a given Boolean function. Given the minterm coefficients of a Boolean function, the chip computes coefficients of all the fixed polarities of the generalized Reed-Muller (GRM) expansions, and identifies the polarity with the least number of terms.

Drechsler [4] proposed a Genetic Algorithm for minimization of FPRM. Experimental results show that for

up to 15 variables the proposed Hybrid GA gives as many as product terms as the exact algorithms give but require much less CPU seconds. Drechsler [5] also proposed a Fast OFDD (Ordered Functional Decision Diagrams) based minimization of FPRM expressions. The experimental results show that it performs very fast. For all functions for which the OFDD can be constructed the minimum FPRM can be obtained. This can be done for (some) functions with several hundred variables.

Khan [6] proposed mapping of fixed polarity Reed-Muller coefficients from minterms and the minimization of FPRM expressions. In paper [6] an efficient and simple algorithm for mapping FPRM coefficients from the on-set minterms of the function for a given polarity vector is presented. Another heuristic algorithm for finding an optimal polarity vector from the on-set minterms that produces the near minimum FPRM expression is also presented. The maximum and the average computational time of these algorithms are $O(n \cdot 2^n)$.

Another exact algorithm, *Sympathy* has been proposed by Drechsler [7] to minimize FPRMs for symmetric functions and the required computation time is polynomial.

An Exact minimization of Fixed Polarity Reed-Muller expressions for Incompletely Specified Functions has been proposed by Debnath [8]. A method for the exact minimization of FPRMs for three-variable switching function has been proposed in the paper [8]. The implementation results shown in paper [8] prove that the algorithm works favorably for many functions with eight or fewer variables and with any number of unspecified minterms.

In this paper the development of an ASIC has been depicted that minimizes 3-variable FPRM expressions. This ASIC is parallel in nature and requires constant time to produce result. The ASIC takes the minterm coefficients of a Boolean function as input. It generates all the polarity vectors for a 3-variable function and determines the optimum polarity and corresponding FPRM coefficients.

II. Background on FPRM expressions

The positive and negative Davio expansions of a logic function are defined as follows. An arbitrary *n*-variable function $f(x_1, x_2, ..., x_n)$ can be expanded using the following expansions.

$f(x_1, x_2, ..., x_n) = f_0 \oplus x_i f_2$ (positive Davio expansion)

$f(x_1, x_2, ..., x_n) = f_1 \oplus \overline{x_i} f_2$ (negative Davio expansion)

where, $f_0 = f(x_1, ... x_{i-1}, 0, x_{i+1}, ..., x_n)$,

$f_1 = f(x_1, ... x_{i-1}, 1, x_{i+1}, ..., x_n)$ and $f_2 = f_0 \oplus f_1$.

In positive Davio expansion the variable x_i appears as only x_i whereas in negative Davio expansion the variable x_i appears as only $\overline{x_i}$.

If we use either the positive or the negative Davio expansion for each variable, we can represent a logic function by a Reed-Muller tree as shown in Fig. 1.

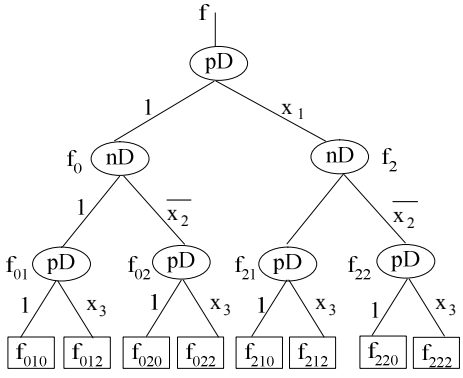


Fig. 1 A Reed-Muller tree for 3-variable function

Fig. 1 shows an example of a Reed-Muller tree for a 3-variable function *f*, where the symbols *pD* and *nD* denote the Positive and the Negative Davio expansions, respectively. In this tree, variable x_1 and x_3 use the Positive Davio expansion and variable x_2 uses the Negative Davio expansion. The expression corresponding to this tree is,

$$f = f_{010}\, 1 \cdot 1 \cdot 1 \oplus f_{012}\, 1 \cdot 1 \cdot x_3 \oplus f_{020}\, 1 \cdot \overline{x_2} \cdot 1 \oplus f_{022}\, 1 \cdot \overline{x_2}\, x_3 \oplus f_{210}\, x_1 \cdot 1 \cdot 1 \oplus f_{212}\, x_1 \cdot 1 \cdot x_3 \oplus f_{220}\, x_1\, \overline{x_2} \cdot 1 \oplus f_{222}\, x_1\, \overline{x_2}\, x_3 \quad (1)$$

In equation (1) we have the following observations:

- For a subscript $i \in \{0, 1\}$ of a coefficient *f*, the corresponding literal of the associated product term appears as $x_i^* = 1$.
- For a subscript $i = 2$ of a coefficient *f*, the corresponding literal of the associated product term appears as $x_i^* \in \{ x_i, \overline{x_i} \}$ depending on the expansion used.

In this work the polarity of an uncomplemented variable is represented by 0 and that of a complemented variable by 1. The reverse polarity convention is also used in the literature, but throughout this work we will follow this convention.

If we replace coefficient *f* with *b*, subscript 1 with 0, subscript 2 with 1, and literals x_i and $\overline{x_i}$ with x_i^* in equation (1), we have the following expression:

$$f = b_{000} \oplus b_{001} x_3^* \oplus b_{010} x_2^* \oplus b_{011} x_2^* x_3 \oplus b_{100} x_1^* \oplus b_{101} x_1^* x_3^* \oplus b_{110} x_1^* x_2^* \oplus b_{111} x_1^* x_2^* x_3^*$$

where, $x_i^* = \begin{cases} x_i & \text{if polarity of } x_i \text{ is } 0 \\ \overline{x_i} & \text{if polarity of } x_i \text{ is } 1 \end{cases}$

This expression uses fixed polarity for a given variable and is called a fixed polarity Reed-Muller (FPRM) expression. This expression is canonical for a given polarity

vector of the variables. In general, the FPRM expression for a logic function is represented as follows:

For example, the FPRM expression for a 3-variable function $f(x_1, x_2, x_3)$ with polarity vector p=(010) can be represented as,

$$F(x_1, x_2, x_3) = b_{000} \oplus b_{001} x_3 \oplus b_{010} \bar{x}_2 \oplus b_{011} \bar{x}_2 x_3 \oplus b_{100} x_1 \oplus b_{101} x_1 x_3 \oplus b_{110} x_1 \bar{x}_2 \oplus b_{111} x_1 \bar{x}_2 x_3$$

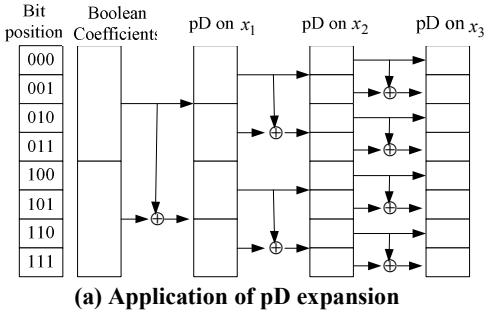
As specific example,

$$F(x_1, x_2, x_3) = \bar{x}_1 x_2 \oplus x_2 \bar{x}_3 \oplus \bar{x}_1 \bar{x}_3$$

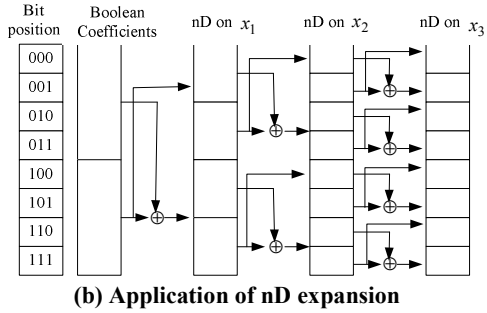
is an FPRM expression for the function $f(x_1, x_2, x_3) = \sum_m (1, 4, 5, 7)$ with the polarity vector p=(101).

III. RTL Level Architecture of the Developed ASIC

In Fig. 2 the computation of the expansions on each variable for both positive and negative Davio expansions are shown. Fig 2(a) shows the computation of the cofactors when the pD expansion is used for each variable x_1, x_2 and x_3. Fig. 2(b) shows the computation of cofactors when the nD expansion is used for each variable x_1, x_2 and x_3. In this work the expansions are used on the variables in the similar approach shown in Fig. 2 depending on the polarity of the variables. If the polarity is 0 then pD expansion is used and if the polarity is 1 then the nD expansion is used.



(a) Application of pD expansion



(b) Application of nD expansion

Fig. 2 Computation of pD and nD expansions for a 3-variable function

For example, for a function $f(A, B, C)$ the input vector $b = [1, 0, 0, 1, 1, 0, 1, 0]^T$ and polarity vector cp=[010] then the transformations on each variable are shown in Fig. 3. Here, the polarity of variable A is 0, B is 1 and C is 0, respectively. So the transformation on A and C is done using *pD* expansion and the transformation on B is done using *nD* expansion. The last column gives the FPRM coefficients of the given function for polarity vector [010].

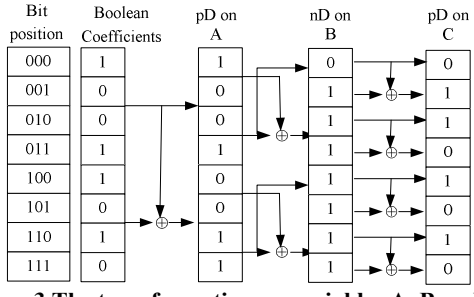


Fig. 3 The transformation on variables A, B and C

The RTL design of the developed ASIC for minimizing FPRM expressions is shown in Fig. 4. This a block diagram of the developed ASIC that minimizes FPRM expressions for 3-variable functions. The eight input coefficients $b_0 ... b_7$ are applied to the inputs of the FPRM converter *convert*. Fig. 5 shows the block diagram of the converter.

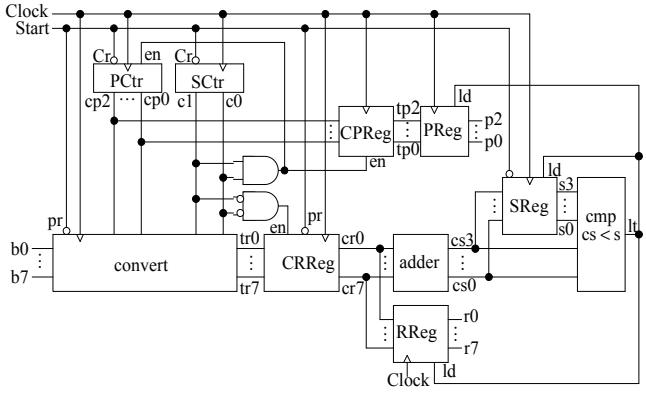


Fig. 4 The block diagram of the developed ASIC to minimize FPRM expressions

The converter consists of $n \cdot 2^{n-1}$ 2-to-1 multiplexers and the same number of EXOR gates where *n* is the number of variables in the function. The converter also has 2^n number of (*n*+1)-to-1 multiplexers. Here, the polarity vector of an FPRM expression works as address lines of the 2-to-1 multiplexers. And thus provide either pD or nD expansion on each variable. Here, if the value of polarity $cp_i\ [where\ 0 \le i \le n-1]$ is 0 then pD expansion is used and if the value of polarity cp_i is 1 then nD expansion is used. As the address lines to the 2-to-1 multiplexers are the polarity vector cp_i, the value of cp_i multiplexes the input lines to the outputs of the converter.

For the first variable, if $cp_2 = 0$ then the 2^{n-1} inputs go directly to the converter output register *tr* ($0 \ldots 2^{n-1}$-1) and the EXOR of the rest 2^{n-1} inputs and the first 2^{n-1} inputs go to *tr*($2^{n-1} \ldots 2^n$-1). If $cp_2 = 1$, then the

EXOR of the first 2^{n-1} and last 2^{n-1} inputs go to the register *tr* ($2^{n-1} \ldots 2^n$ -1) and the last 2^{n-1} inputs go to *tr*($0 \ldots 2^{n-1}$ -1) directly. In this way, for the second variable the inputs to the 2-to-1 multiplexers are the values of the vector *tr*. Now, if $cp_1 = 0$ then ($0 \ldots 2^{n-2}$ -1) of inputs go directly to the register *tr*($0 \ldots 2^{n-2}$ -1) and the EXOR of ($0 \ldots 2^{n-2}$ -1) and ($2^{n-2} \ldots (2^{n-2} + 2^{n-2}$ -1)) inputs go to *tr*($2^{n-2} \ldots (2^{n-2} + 2^{n-2}$ -1)). Accordingly, the transformations on the other variables are done.

The *convert* produces 2^n bit FPRM coefficients for each polarity vector. For an *n* variable function we have 2^n polarity vectors and thus 2^n sets of FPRM coefficients.

In Fig. 4 *PCtr* and *SCtr* are three bit and two bit counters, respectively. The counter PCtr provides 2^n numbers of *n* bit polarity vectors to *convert*. For a three variable function the converter takes four states to generate FPRM coefficients, where the counter SCtr counts states for each polarity vector. At the first state the inputs are loaded into the converter, and then at each three of the following states the transformation on each variable is done. That is, an *n* variable function will need *n+1* states for the converter to generate FPRM coefficients for a particular polarity vector.

In Fig. 4 *CRReg* is a 2^n bit register which initially holds 2^n 1's. Then after the generation of each set of FPRM coefficients CRReg is loaded with the FPRM coefficients. That is, the converter passes the FPRM coefficients to CRReg.

The *adder* in Fig. 4 is a module that adds the bits of the FPRM coefficients to determine the number of 1's in the FPRM coefficients. As we are intended to find the FPRM expressions with least number of product terms thus least number of 1's, we keep record of number of 1's in each set of FPRM coefficients. The adder computes the number of 1's in the FPRM coefficients for each polarity vector.

The register *RReg* in Fig. 4 holds a set of FPRM coefficients which has the least number of 1's. If the *i*-th polarity vector produces FPRM coefficients which has lesser number of 1's than the FPRM coefficients produced by the i+1-st polarity then RReg holds the FPRM coefficients produced by the *i*-th polarity.

Another register in Fig. 4 *CPReg* holds the polarity vectors. When the converter computes the FPRM coefficients for a polarity vector then CPReg is loaded with the next polarity vector.

PReg register in Fig. 4 holds the value of the polarity which produces FPRM coefficients with least number of 1's. The process of keeping record of the polarity vector is the same as the process which stores the FPRM coefficients with least number of 1's in RReg.

The register *SReg* in Fig. 4 holds the value of number of 1's in the FPRM coefficients. Initially this register is loaded with the value of 2^n, as the FPRM coefficients can have at most 2^n 1's for an *n* variable function. Again the process of storing this value is same as the process which stores the FPRM coefficients with least number of 1's in RReg.

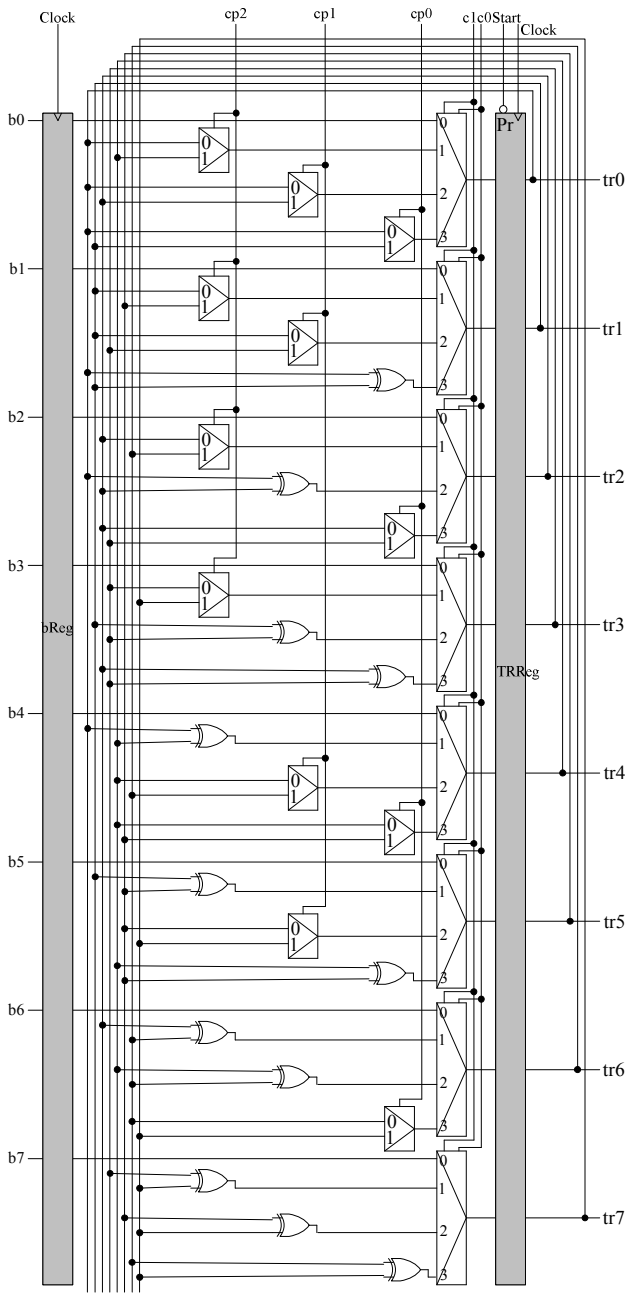


Fig. 5 The Converter that converts an 8-bit Boolean function to 8-bit FPRM coefficients for a particular polarity vector

The *cmp* module used in Fig. 4 is a comparator. This module compares the number of 1's in the FPRM coefficients produced by the polarity vector being used presently (*cs*) with the FPRM coefficients' number of 1's (*s*) produced by some other polarity vectors previously. If $cs < s$ then the output of the comparator (the line *lt*) goes high. This output *lt* enables the registers *RReg*, *PReg* and

SReg. After the computation of FPRM coefficients for 2^n polarity vectors we get the FPRM coefficients with the minimum number of 1's (*r*) in register *RReg*, the polarity vector (*p*) which produced *r* in register *PReg* and we get the number of 1's in *r* in register *SReg*.

In Fig. 4 all the flip-flops are positive edge-triggered and the total process is *reset* with the negative edge of *start*. The transition between states is shown in Fig. 6.

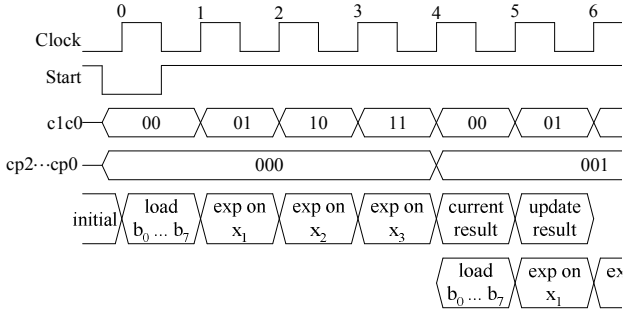


Fig. 6 The pipelining of the minimization process

IV. FPGA Implementation of the ASIC for FPRM Minimization

We have used Verilog as the HDL to describe the developed design and the Quartus II 4.2 software to synthesize the design. The design has been implemented using Altera UP3 FPGA Education Kit. Table 1 shows the maximum frequency and logic elements required to implement our circuit using different devices of different families available in Quartus II 4.2.

Table 1 Compilation Reports

Family	Device	Total logic elements	Clock Setup time
Cyclone	EP1C6Q240C8	89 / 5,980 (1 %)	129.62 MHz
Cyclone II	EP2C5T144C6	85 / 4,608 (1 %)	182.08 MHz

The simulation results for an input function are shown in Fig. 7 and 8. Here, the signal *A* is the minterm coefficients of the input function, signal *r* is the FPRM coefficients, signal *p* is the polarity vector corresponding to *r* that produces *A* and signal *s* shows the value of number of 1's in *r*. From both the figures we can observe that the number of 1's in the signal *r* is varying for different polarity vectors. The signal *p* holds the best polarity vector. As shown in Fig. 7 and 8 the minimum of number of 1's in *r* comes for the polarity vector [110].

V. Conclusion

There are seven types of AND-EXOR logic expressions, Fixed Polarity Reed-Muller (FPRM) expression is one of them. FPRM expressions have the property that the polarity of a variable remains same throughout the expression, which eases the implementation of the expression in VLSI. For an *n*-variable function, there are 2^n possible FPRM expressions having different number of products and number of literals. So, finding out the minimum FPRM expression for a given Boolean function is very important.

There are many software methods for FPRM minimization. But for real time problems the software approaches are not applicable as they all take exponential time for computation. In this work an approach to minimize FPRM expressions using hardware and implemented in FPGA has been outlined. In real life problems finding Boolean equivalence of a function is important. The FPRM representation is a tool to find Boolean equivalence or Boolean matching. Our designed ASIC will provide the FPRM coefficients and the optimum polarity vector for a particular Boolean function of three variables within constant time.

This design is able to minimize FPRM expressions of three variable single-output, fully-specified functions and is very small in size. Further works may include,

- To parameterize the design so that the number of variables the ASIC can handle is $n>3$.
- To develop the design for handling multi-output and incompletely specified functions.
- To develop the design for optimization of other AND-EXOR expressions such as pseudo Reed-Muller, Kronecker, pseudo Kronecker, etc.

References

[1] T. Sasao, *Logic Synthesis and Optimization*, Kluwer Academic Publisher, 1993.

[2] C. –C. Tsai and M. Marek-Sadawska, "Boolean functions classifications via fixed polarity Reed-Muller forms", IEEE Transaction on Computers, vol. 46, no. 2, pp. 173-186, 1997.

[3] A. E. A. Almaini, "A semicustom IC for generating optimum generalized Reed-Muller expansions", Microelectronics Journal, vol. 28, no. 2, pp. 129-142, February 1997.

[4] R. Drechsler, B. Becker and N. Göckel, April, "A Genetic Algorithm for minimization of Fixed Polarity Reed-Muller expressions", Computers and Digital Techniques, IEE Proceedings, vol. 147, no. 5, pp. 349 – 353, October 2000.

[5] R. Drechsler, M. Theobald and B. Becker, "Fast OFDD based minimization of fixed polarity Reed-Muller expressions" IEEE Transactions on Computers, vol. 45, no. 11, pp. 1294-1299, 1996.

[6] M. M. H. A. Khan and M. S. Alam, "Mapping of fixed polarity Reed-Muller coefficients from minterms, and the minimization of fixed polarity Reed-Muller expressions", International Journal of Electronics, vol. 83, no. 2, pp. 235-247, 1997.

[7] R. Drechsler and B. Becker, "Sympathy: Fast Exact Minimization of Fixed Polarity Reed-Muller Expressions for Symmetric Functions." Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions, vol. 16, no. 1, pp. 1-5, January 1997.

[8] D. Debnath and T. Sasao, "Exact minimization of fixed polarity Reed-Muller expressions for incompletely specified functions," Asia and South Pacific Design Automation Conference (ASP-DAC'2000), Yokohama, Japan, pp. 247-252, 2000.
[9] Ph. W. Besslich, "Spectral processing of switching functions using signal-flow transformations", in (M. Karpovsky ed.) Spectral Techniques and Fault Detection, (Orlando, FL: Academic Press), pp. 91-141, 1985.
[10] E. M. Clarke, K. L. McMillan, X. Zhao, M. Fujita, and J. Yang, "Spectral transforms for large Boolean functions with applications to technology mapping," in Proc. Design Automation Conference, Dallas, Texas, USA, pp. 54–60, June 1993.
[11] Lun Li, M. Thornton, M. Perkowski, "A Quantum CAD Accelerator Based on Grover's Algorithm for Finding the Minimum Fixed Polarity Reed-Muller Form", 36th International Symposium on Multiple-Valued Logic, pp. 33, May 2006.
[12] R. S. Stankovic and B. J. Falkowski, "Spectral interpretation of the fast tabular technique for fixed-polarity Reed-Muller expressions", International Journal of Electronics, vol. 87, no. 6, pp. 641-648, 2000.

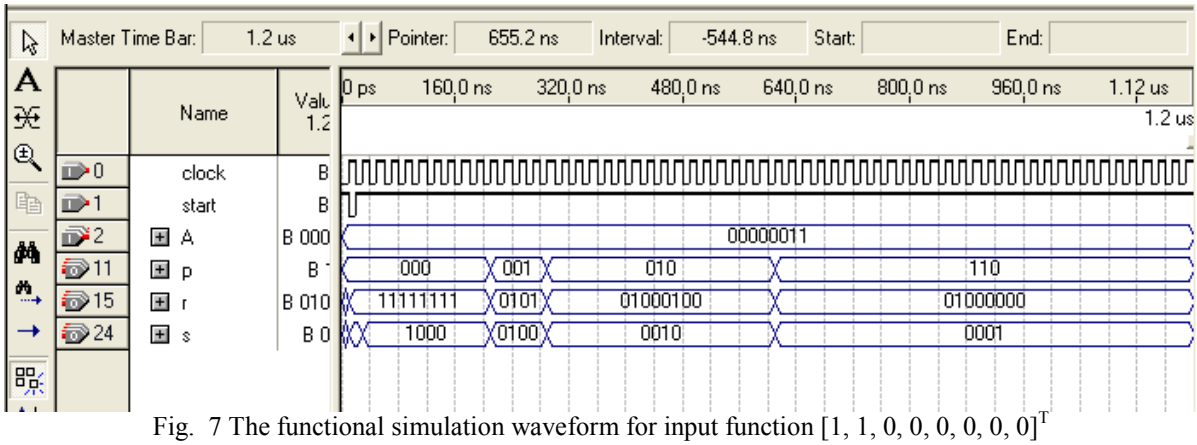


Fig. 7 The functional simulation waveform for input function $[1, 1, 0, 0, 0, 0, 0, 0]^T$

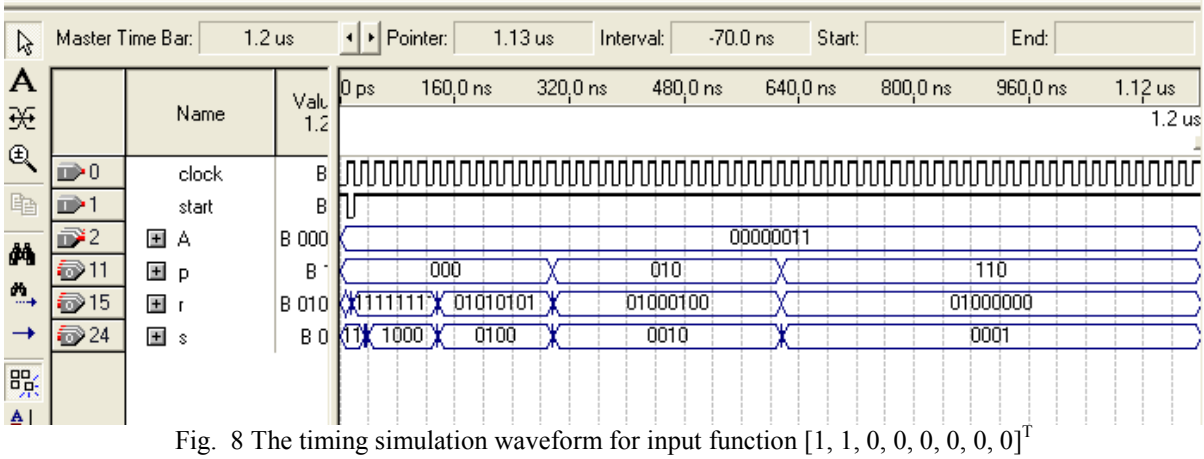


Fig. 8 The timing simulation waveform for input function $[1, 1, 0, 0, 0, 0, 0, 0]^T$

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Micro Heat Pipes – A Promising Means of Thermal Solution for Desktop Computers

Ahmed Imtiaz Uddin[1] and Chowdhury Md. Feroz[2]

Department of Mechanical Engineering, Bangladesh University of Engineering and Technology, BUET, Dhaka -1000, Bangladesh.
E-mail: [1]aimtiaz_me@yahoo.com; [2]cmferoz@me.buet.ac.bd

Abstract - Heat transfer performance of parallel micro heat pipes (MHPs) used for the cooling of desktop computer processor has been investigated in this paper. MHPs consists of six single tube micro heat pipes connected by a copper block at the evaporator section and fifteen parallel copper sheets used as external fins at the condenser section. Ethanol was used as working fluid. The copper block is placed above the heat source (on the top of the processor) and the condenser section is provided with external fins perpendicular to the MHPs. Heat transfer characteristics of MHPs are determined experimentally, based on the principle of phase change of the working fluid. The experimental results show that, the maximum and steady state temperature of the processor has been reduced significantly by using micro heat pipes instead of the cooling fan. Addition of fan at the condenser section of MHPs gives much low and steady state temperature of the processor.

I. Introduction

In the demand of high performance computers, more and more power performance processors had been developed for use in desktops. The drawback in the higher performance processor is that they have larger heat generation. This creates a challenge in providing thermal solution in desktop computers since high performance processor and additional features such as more transistors and cache memory leads to higher total power dissipation in desktop processors. It is even more challenge to provide the thermal solution without compromising the performance.

With the continuing increase of CPU and system power dissipation in personal computers including notebooks, desktops, workstation and servers the last decade has witnessed an increasing use of and interest in heat pipes in personal computers for meeting cooling requirements. The development of high-end and compact computers has resulted in a considerable rise in the power dissipation tendency of their microprocessors. At present, heat released by the Central Processing Unit (CPU) of a desktop and server computer is 80 to 130 W and of notebook computer is 25 to 50W [1]. In the latter case, the heating area of the chipset has become as small as 1–4 cm. This problem is further complicated by both the limited available space and the restriction to maintain the chip surface temperature below 100oC [2]. It is expected that conventional cooling fan system will not be able to meet the futuristic thermal needs of the next generation computers. Other technologies like liquid cooling and thermoelectric coolers have good potential but still create major integration, reliability and cost issues. With the development in the two-phase heat transfer systems and porous media technology, heat pipes have come up as a potential candidate to meet these challenging needs. Notebook computers involved the first high volume use of heat pipes when Intel introduced the Pentium ® TCP packages in 1994 [3]. The main reason for the use of heat pipes is the Pentium ® power dissipation level and the limitation and constraints of space and weight in notebooks. Compared to metal plates or heat sinks, heat pipes offer excellent thermal performance with much less weight and can spread the heat away from the CPU to other areas where the heat can be rejected. Today, Pentium ® based notebooks and sub-notebooks are estimated to use several millions of heat pipes annually based on the PC based notebook volume.

The performance of natural convection heat sinks is directly dependent on the effective surface area: more effective surface area results in better performance. To enhance the heat transfer, an additional cooling fan is used with the aluminium heat sink. The increase of the microprocessor speed and number of transistors cramped into the processor core silicon die has continuously driven up its power dissipation. Heat sink sizes have been increasing in personal computers, from the 2”× 2” aluminium extrusion heat sinks for i486 to the 3”× 3” heat sinks for Pentium ® and even large heat sinks for the latest Pentium ® II microprocessors. Heat pipes, as higher level thermal solutions are naturally being investigated as the potential thermal solutions for these systems [4].

Another severe problem of today’s processor cooling fan is the generation of noise. Much effort has been made in recent years to minimize noise generated by CPU cooling fans, a fact that has been demonstrated by the popularity of variable and low speed fans coupled with efficient CPU heat sink designs. Even with the adjustable fans generating lower noise at lower speeds, the main noise sources in a computer system are fans and hard drive. Therefore, the best way to eliminate the noise is to remove these sources. As it is impractical to get rid of the hard drives, it seems like a good idea to cool the CPU without a fan. After looking at products based on heat pipe technology, such as

Zalman's graphics card coolers, a good idea could be to try passive CPU cooling utilizing heat pipes [5]-[7].

The heat pipe can, even in its simplest form, provide a unique medium for the study of several aspects of fluid dynamics and heat transfer and it is growing in significance as a tool for use by the practicing engineer or physicist in applications ranging from heat recovery to precise control of electronic equipments. Normally for these equipments heat pipes of diameter 3 to 6 mm and length less than 400 mm are preferred [8]. Most preferable length is 150 mm [9]. The heat pipe applications for cooling computer CPU was started in the last decade and now 98% of notebooks PCs are cooled by using heat pipes.

The concept of micro heat pipe (MHP) was introduced by Cotter in 1984 [10]. MHP is a very small heat pipe that has a diameter between about 100 micrometers and 2 millimetres (0.004 and 0.08 inch) and a length of several centimetres. Studies on the application of heat pipes having the diameter of 3 or 4 mm for cooling of the notebook PC CPU has been actively conducted by the American and Japanese enterprises specializing in the heat pipe recently [7], [11], [12]. An experimental study is performed by Tanim et al. [13] to investigate the performance of cooling desktop processors using miniature heat pipes of 5.78 mm ID and a length of 150 mm with respect to the normal fanned CPU unit. So far no investigation has been conducted for cooling desktop processor with MHPs. The concept in this experiment is to draw the heat from the CPU into one end of micro heat pipes while making the other end of the MHPs as extended fins of copper plate to expel the heat into the air. Finally the performance of the MHPs in cooling desktop processor is investigated with respect to the miniature heat pipe and conventional fanned CPU cooling system.

II. Experimental Setup and Test Procedure

The experimental setup for this study is mainly consists of four parts – parallel MHPs, a desktop computer, temperature measurement system and cooling system. Six MHPs are placed parallel to each other for cooling purpose. Every MHP has an inner diameter of 1.8 mm and outer diameter of 2.8 mm having a length of 150 mm. There are three sections in every MHP: evaporator, adiabatic and condenser. Ethanol is used as the working fluid in this experiment.

The condenser sections of MHPs are made of copper sheets of 67mm×50mm (thickness 0.5mm) placed parallel as extended fins at a constant interval of 5 mm as shown in Fig. 1. Plates are welded with the MHPs for better heat transfer. As there is space constrain inside the CPU, the MHPs are bend at 90o in adiabatic section. The evaporator sections of MHPs are inserted in to the grooves of copper blocks shown in Fig. 2, which are placed on the top of the processor to remove the generated heat. Two copper blocks of 67mm×50mm×8mm are made very precisely to mate with the MHPs. Grooves are cut inside the blocks. The blocks are precise in dimension and surfaces are finished highly to reduce the contact resistance as well as to increase the heat transfer rate.

Heat is generated in the processor which is conducted through the copper blocks to the evaporator sections of MHPs where working fluid absorbs the heat and rejects that in the condenser sections. Before bending, wick of stainless steel of 200 mesh are inserted into the MHPs. After inserting the wick, MHPs are bent to the desired angle. The different sections of a parallel MHP is shown in Fig. 3. One end of the MHPs is sealed and ethanol as working fluids with charge ratio 0.9 is poured into that. Nine calibrated K-type (Φ = 0.18 mm) thermocouples are attached at the wall of each MHP using adhesive to measure the wall temperature: four units at the evaporator section, one unit at the adiabatic section and four units at the condenser section. Locations of thermocouples connected on different points along the length of the MHP are shown in the Fig. 4. The surface temperature of the processor is also measured by four K-type thermocouples. All thermocouples are connected with a digital temperature indicator (YF-160A, type-K thermometer, Made in Taiwan) through selector switches. Experiment is conducted in two arrangements: Firstly the cooling system by using MHPs, shown in Fig. 5 and secondly with the MHPs, an additional cooling fan placed at condenser section as shown in Fig. 6 to enhance heat transfer rate. Processor surface temperature and wall temperatures of MHPs are recorded for 150 minutes at an interval of 10 minutes. Experimental parameters and configurations of the desktop computer that is used in this experiment are given in Table 1 and Table 2, respectively.

Table 1 Experimental parameters.

Parameters	Condition
Number of the heat Pipes	6
Diameter of the heat pipe (mm)	ID- 1.8 OD- 2.8
Length of the heat pipe (mm)	150
Length of the evaporator section (mm)	50
Length of the adiabatic section (mm)	30
Length of the condenser section (mm)	70
Working fluid	Ethanol
Dimension of the copper block (mm)	67×50×8
Dimension of the copper sheet(mm)	67×50×0.5
Charge ratio	0.9
Wick (SS)	200 mesh

Table 2 Configuration of Desktop computer.

Components	Specification
Processor	Ali M1542A1
Fan	Power logic- DC brushless Fan, Model- PL80S12H-1; DC-12V, 0.18A
Ram	16 MB
Hard disk	Seagate; Model ST34321A; 10GB
Power box	115/230 VAC,15 A/ 10 A

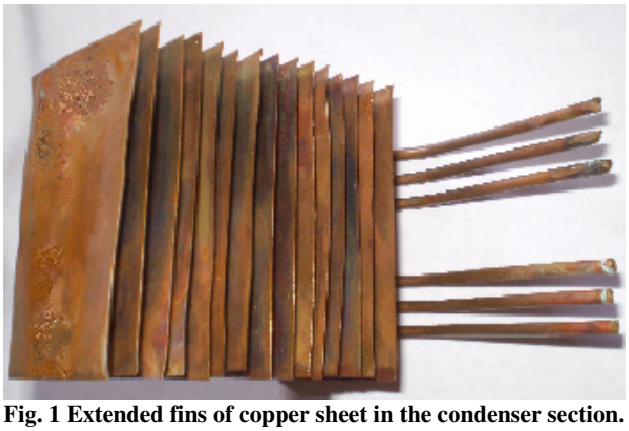


Fig. 1 Extended fins of copper sheet in the condenser section.

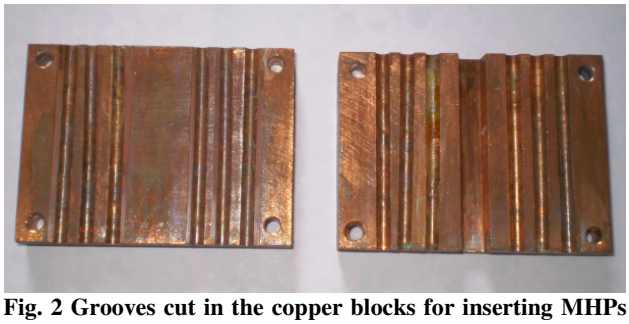


Fig. 2 Grooves cut in the copper blocks for inserting MHPs evaporator section.

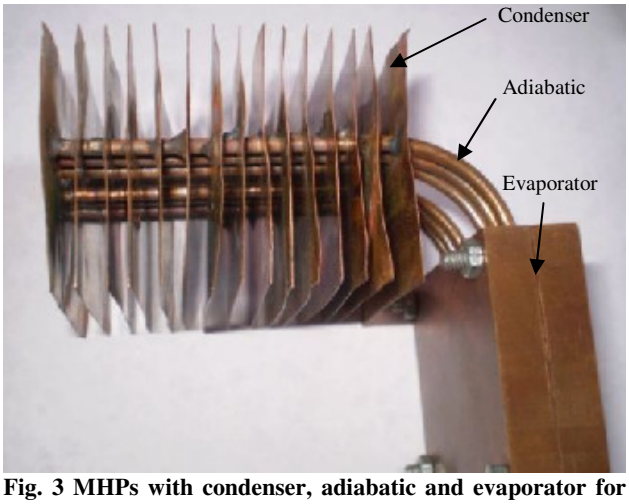


Fig. 3 MHPs with condenser, adiabatic and evaporator for cooling desktop processor.

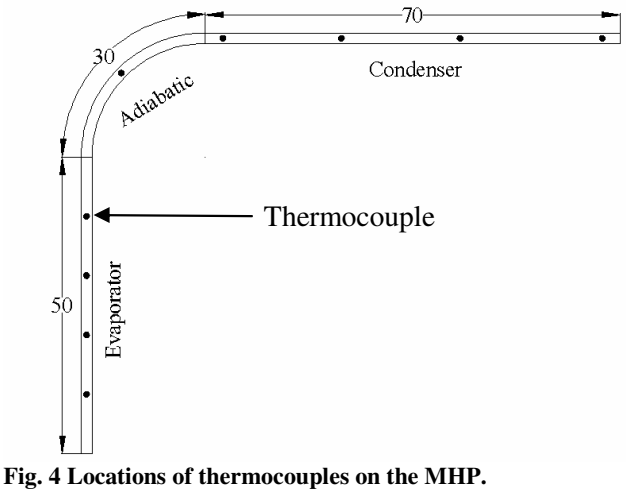


Fig. 4 Locations of thermocouples on the MHP.

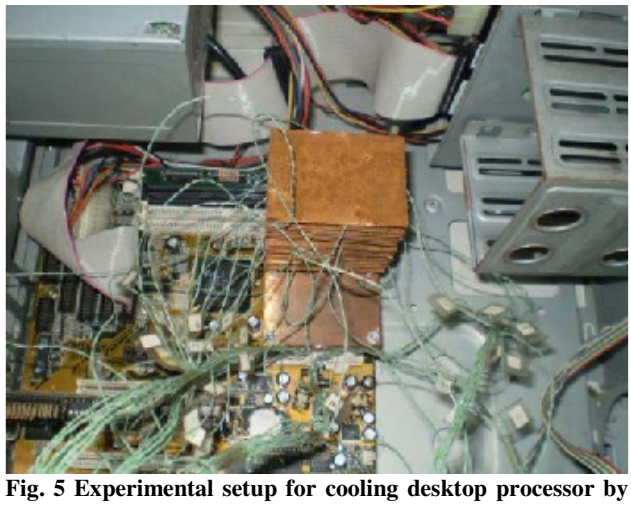


Fig. 5 Experimental setup for cooling desktop processor by using MHPs.

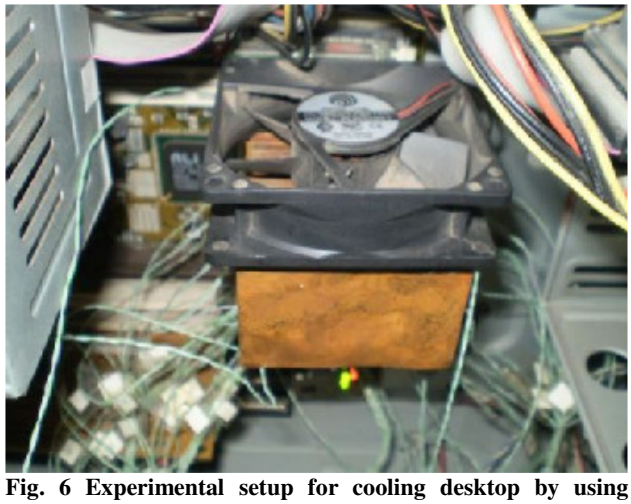


Fig. 6 Experimental setup for cooling desktop by using MHPs with cooing fan.

III. Test Results and Discussions

As heat load is applied to the evaporator section, the temperature of the evaporator rises and results in the vaporization of the working fluid. This vaporization of liquid absorbs heat from the evaporator section which causes a lower working temperature in the processor.

Fig. 7 shows the variation of the processor surface temperature with time. Results of Tanim et al. [13], is included in this plot to compare their results with results of the present study. The figure indicates that:

- The maximum temperature on the processor surface by using conventional cooling fan and aluminum heat sink is found 90.9ºC.
- Replacement of the cooling fan and aluminium heat sink with six MHPs of 1.8 mm ID is efficient as it can reduce the maximum processor surface temperature to 79.8 ºC which shows much lower value than that of four heat pipes of 5.78 mm ID without cooling fan [13].
- Addition of a cooling fan in the condenser with six MHPs gives much better result as it can reduce the processor surface temperature to 72.3ºC.

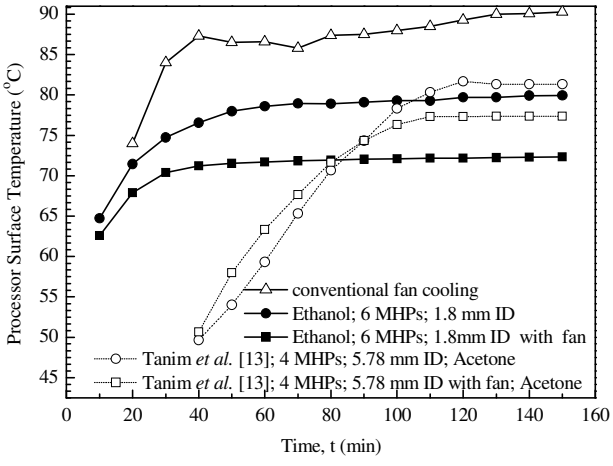


Fig. 7. Variation of processor surface temperature with time.

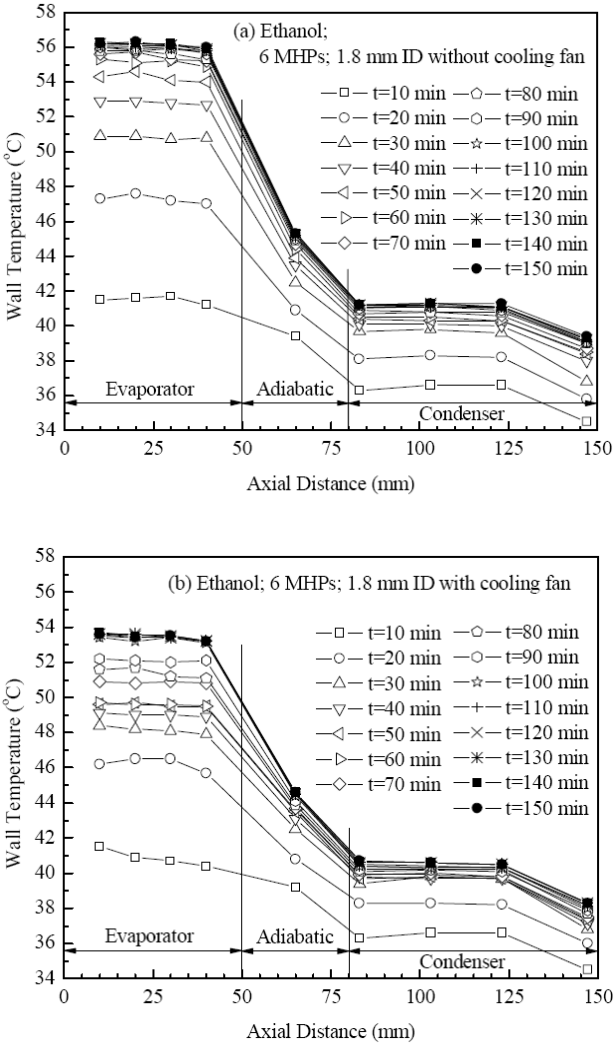


Fig. 8 Temperature profile along the length of the MHPs.

Fig. 8 shows the axial wall temperature distribution along the length of the MHP for the time duration of 150 minutes. For six MHPs without fan, the maximum wall temperature in the evaporator section raises to 56.5 ºC, shown in Fig. 8 (a). The maximum wall temperature in the evaporator section of MHP is 53.8 ºC, when a cooling fan is used with the MHPs, shown in Fig. 8 (b). Using MHPs, the steady state temperature in the evaporator is attained approximately after 80 minutes from the start of the CPU which can be reduced to approximately 60 minutes by using a cooling fan with MHPs. Result shows that natural convection cooling by using only extended fins in the condenser of MHPs is not efficient enough to cool the high performance desktop processors.

Variation of the axial wall temperature along the length of the MHP with and without the cooling fan for a transient time t=30 minutes after the start of CPU is shown in Fig. 9. The temperature in the evaporator section is almost uniform. This uniformity of temperature in the evaporator and condenser sections indicates the reliability of using MHP for the cooling of desktop processors.

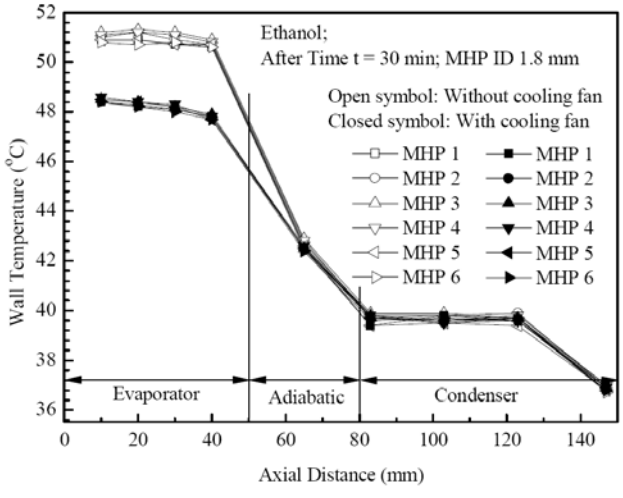


Fig. 9. Temperature profile of the six MHPs along their length after a time t=30 min.

IV. Conclusion

The following conclusions can be drawn form the experimental study of the performance test of MHPs in cooling of desktop computer processors are:

- Addition of a fan at the condenser section of MHPs provides maximum low temperature of the surface of the processor.
- Steady state temperature in the evaporator sections of MHPs with fan is attained much earlier than that of MHPs without a fan.
- Insignificant fluctuation of temperature along the length of the MHPs during the collection of temperature data indicates the stability and consistency of the system.

References

[1] M. Mochizuki, Y. Saito, V. Wuttijumnong, X. Wu, and T. Nguyen, "Revolution in fan heat sink cooling technology to extend and maximize air cooling for high performance processors in laptop/desktop/server application," in Proc. IPACK 2005, San Francisco, CA, Jul. 17–22, 2005, [CD ROM].

[2] I. Saucius, R. Prasher, J. Chang, H. Erturk, G. Chrysler, C. Chiu, and R. Mahajan, "Thermal performance and key challenges for future CPU cooling technologies," in Proc. IPACK 2005, San Francisco, CA, Jul. 17–22, 2005, [CD ROM].

[3] H. Xie, M. Aghazadeh, W. Lui, and K. Haley, "Thermal solutions to pentium processors in TCP in notebooks and subnotebooks," IEEE Transactions on components, Packaging and Manufacturing Technology. Part A, Vol. 19, No. 1, pp54-65, March, 1996.

[4] H. Xie, A. Ali, and R. Bathia, "The use of heat pipes in personal computers," IEEE Inter Society Conference on Thermal Phenomena 1998; pp 442-448.

[5] B. Babin, R. Peterson, G. P., D. Wu, "Steady state modeling and testing of a micro heat pipe," ASME J. of Heat Transfer, Vol. 112, No. 3, August 1990, pp. 595-601.

[6] J. Zhou., Z. Yao, and J. Zhu, "Experimental investigation of the application characters of micro heat pipe," Proc. 8th Int. Heat Pipe Conf., Beijing, China, 1992.

[7] K. Eguchi, M. Mochizuki, K. Mashiko, K. Goto, Y. Saito, A. Takamiya, and T. Nguyen, "Cooling of CPU using micro heat pipe," Fujikura Co., Technical Note, Vol. 9, 1997, pp. 64~68.

[8] S. Yoshiaki, M. Takase, M. Tanabe, N. Teruo, I. Kinoshita, I. Tadashi, K. Namba and S. Masahiro, "A junction block incorporating a micro heat-pipe," Furukawa Review, No. 18. 1999.

[9] K. S. Kim, S. H. Moon, and C. Gi. Choi, "Cooling characteristics of miniature heat pipes with woven wired wick," in Proc. of the 11th international Heat Pipe Conference – Tokyo 1999, pp 239 – 244.

[10] T. P. Cotter, "Principles and prospects for micro heat pipes", in Proc. of the 5th International Heat Pipe Conference – Tsukubs, Japan, 1984, pp. 328~335

[11]H. Xie, M. Aghazadeh and J. Togh, "The use of heat pipes in the cooling of portables with high power packages," Thermacore Co., Technical Note.

[12]M. Mochizuki, K. Mashiko, T. Nguyen, Y. Saito, and K. Goto, 1997, "Cooling CPU Using Hinge Heat Pipe," Heat Pipe Technology, Pergamon , 1997, pp. 218~229.

[13]T. R. Tanim., T. Hussain and C. M. Feroz, "Cooling of desktop computer using heat pipes," in Proc. ICME'07, Dhaka, Bangladesh, Dec. 29-31, 2007. [CD ROM, ICME07-TH-01]

ACKNOWLEDGEMENT

This work was supported by Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh. The authors are also grateful to the Department of Mechanical Engineering and Directorate of Advisory Extension and Research Services (DAERS), BUET, Dhaka for providing the facilities to carryout the experiment.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Low-Cost Realization of Quantum Ternary Adder Using Muthukrishnan-Stroud Gate

Md. Mehedi Hasan

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
E-mail: mehedibuet@gmail.com

Abstract –Ternary quantum computing technology is a promising technology for future computer systems. In any computer system, an adder circuit is an essential circuit that performs arithmetic and logic operation of the computer. So, for the ternary quantum computer system, an efficient and low-cost realization of ternary adder is very necessary. This paper introduces a novel realization of quantum ternary adder which reduces the realization cost extensively compared to the previous realization.

I. Introduction

Quantum computing technology is one of the most promising technologies for future computer systems [1]. As the quantum computer is based on the reversible logic, the power dissipation will be extremely low compared to the counterpart CMOS computer [2]. This computer is able to perform vast amount of computation [1, 3] and also able to solve some of the most unfeasible problems in computer science such as large integer factorization, discrete logarithms etc [4].

For realization of quantum circuits, three-valued logic (ternary logic) is a very good choice as the amount of information per bit for ternary logic is more than that for two-valued (binary) logic. For this reason, ternary quantum logic synthesis is a very attractive research area in the recent years [5-9].

Oracle is the most vital part in quantum computer system as ALU in the classical computer. Oracle consists of arithmetic, logic, and other units. Ternary adder is the part of the oracle that can perform the arithmetic and many logic operations. So, designing a ternary full-adder efficiently is very important for the realization of the quantum computer.

In literature, some realizations of quantum ternary full-adder are proposed [7-9]. The realization [7] uses 10 Generalized Ternary Gates (GTG) for realization; [8] uses 96 elementary gates and [9] uses 50 elementary Muthukrishnan-Stroud (MS) gates (MS gate is described in section II). To reduce the realization cost (cost is measured in terms of the number of the elementary or basic gates) more, in this paper, we have proposed an efficient and low-cost realization which uses 19 MS gates only. Compared to the recent realization [9] which also uses MS gates, the realization cost is reduced by 62%.

The rest of the paper is organized as follows. Section II reviews a group of quantum ternary basic gates (called MS gate) which has been used to realize the adder in this paper. Section III presents the proposed realization of ternary full-adder and also analyzes the simulation result. Finally, section IV provides the conclusion.

II. Background

Muthukrishnan and Stroud proposed a group of 2-quidit (quantum digit) primitive ternary gates which are ion-trap realizable [10]. Basically the success in the realization of these gates gives the hopes of building practical quantum computer. This group of gates is known as Muthukrishnan-Stroud (MS) gate.

Fig. 1 shows the MS gate for ternary system. Here, A and B are the controlling input and the controlled input respectively while P and Q are the outputs. If the logic value of A equals 2, the value of Q will be the quantum Z-transform of the input B and if A is not equal to 2, then $Q=B$. Different Z-transforms are shown in Table I. For example, if Z-transform is +2, $A=2$ and $B= 0$, Q will be 2.

III. Proposed Realization of Ternary Full-adder

Table 2 shows the truth table of the ternary full-adder. Here, A, B are the two input that are to be added, C_{in} is the carry input while C_{out} and S are the carry-output and sum respectively.

Fig. 2 shows the proposed realization of the ternary full-adder. The inputs other than A, B, and C_{in} which are known as the ancillary or auxiliary inputs are used for the help of the synthesis. As seen from the figure, we have used only ion-trap realizable MS gates.

The operation of the proposed circuit is verified by the simulation using MATLAB. Table 3 shows the simulation result. The result shows the logic condition of the different nodes for the different nodes for different inputs.

To explain the operation and result, let us take an example when the inputs are $A=1$, $B=2$, and $C_{in}=0$. As the Z-transform of +1 (with the controlling input of 2) is applied to the input $A=1$, the node A_1 will be 2. Similarly, A_2 becomes 1 for the transform +2. A_2 is the input of the MS gate (+2) whose controlling input is B (=2). So $A_3= 0$. By

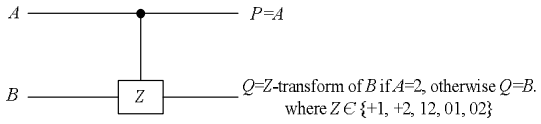


Fig. 1 Quantum ternary MS gate.

Table 1 Values after quantum *Z*-transform.

Z-Transform	*Q* when *B*=		
	0	1	2
+1	1	2	0
+2	2	0	1
12	0	2	1
01	1	0	2
02	2	1	0

the same way, A_4, A_5, A_6, A_7, and A_8 becomes 0, 0, 0, 1, and 0 respectively. A_8 gives the sum (S). To examine the logic condition of the line of input B, $B_1= 0$ and $B_2=2=B$ as this is the reversible circuit. Again, $C_1=1$ and $C_2=0=C_{in}$. The logic conditions AX_1, AX_2, AX_3, and AX_4 are 0, 1, 1, and 1 respectively. As AX_6, AX_7 becomes 0 and 1 respectively, AX_5 outputs 1. The node AX_5 gives the carry-out. The operation of the circuit for the other inputs can be explained in the similar way.

It is seen from the Fig. 2, the cost (number of elementary gates) of the proposed realization is 19 whereas the cost of the realization of [9] is 50. Therefore, our realization requires less elementary gates than that of [9] and reduces the realization cost by 62%.

IV. Conclusion

In this paper, we have proposed a novel circuit for the realization of ternary full-adder. For realization, we have used the MS primitive gate which is practically realizable. Here, the implementation cost is reduced from 50 to 19 (62% reduction). The proposed circuit presents a low cost, simple, and practically realizable scheme for the quantum ternary full-adder.

Acknowledgment

The author would like to thank Asif I. Khan, of Bangladesh University of Engineering and technology for the wealth of good ideas and advice that he has provided.

Table 2 Truth table of ternary adder.

A	*B*	C_{in}	C_{out}	*S*
0	0	0	0	0
0	0	1	0	1
0	0	2	0	2
0	1	0	0	1
0	1	1	0	2
0	1	2	1	0
0	2	0	0	2
0	2	1	1	0
0	2	2	1	1
1	0	0	0	1
1	0	1	0	2
1	0	2	1	0
1	1	0	0	2
1	1	1	1	0
1	1	2	1	1
1	2	0	1	0
1	2	1	1	1
1	2	2	1	2
2	0	0	0	2
2	0	1	1	0
2	0	2	1	1
2	1	0	1	0
2	1	1	1	1
2	1	2	1	2
2	2	0	1	1
2	2	1	1	2
2	2	2	2	0

References

[1] M. Nielsen, I. Chuang, *Quantum Computation and Quantum Information,* Cambridge University Press, 2000.

[2] C. H. Bennet, *Logical reversibility of computation,* IBM Res. Develop. 17 (1973) 525-532.

[3] M. Hirvensalo, *Quantum Computing,* Springer Verlag, 2001.

[4] P. Shore, "Algorithms for quantum computation: discrete logarithms and factoring," in Proc. 35th Ann. Sym.Foundations of Computer Science, Santa Fe, NM, USA, pp. 124-134, November 1994.

[5] M. H. A. Khan, M. A. Perkowski, M. R. Khan, and P. Kerntopf, "Ternary GFSOP minimization using Kronecker Decision Diagram and their synthesis with quantum cascades," J. Multiple-valued Logic and Soft Computing, vol. 11, pp. 567–602, 2005.

[6] D. M. Miller, G. Dueck, and D. Maslov, "A synthesis method for MVL reversible logic," in Proc. 34th IEEE Int. Symp. Multiple-valued Logic, Toronto, Canada, pp. 74-80, 2004.

[7] M. H. A. Khan, "Quantum Realization of Ternary Adder Circuits," in Proc. 3rd Int. Conf. Electrical and Computer Engineering, Dhaka, Bangladesh, pp. 141-144, 2004.

[8] D. M. Miller, G. Dueck, and D. Maslov, "Synthesis of quantum multiple-valued circuits," J. Multiple-valued logic and Soft computing, vol. 12, 2006.

[9] M. H. A. Khan and M. A. Perkowski, "Quantum ternary parallel adder/subtractor with partially-look-ahead carry," *J. System Architecture,* vol. 53, pp. 453-464, 2007.

[10] A. Muthukrishnan and C. R. Stroud, "Multivalued logic gates for quantum computation," Physical *Review A,* vol. 62, pp. 052309/1-8, 2000.

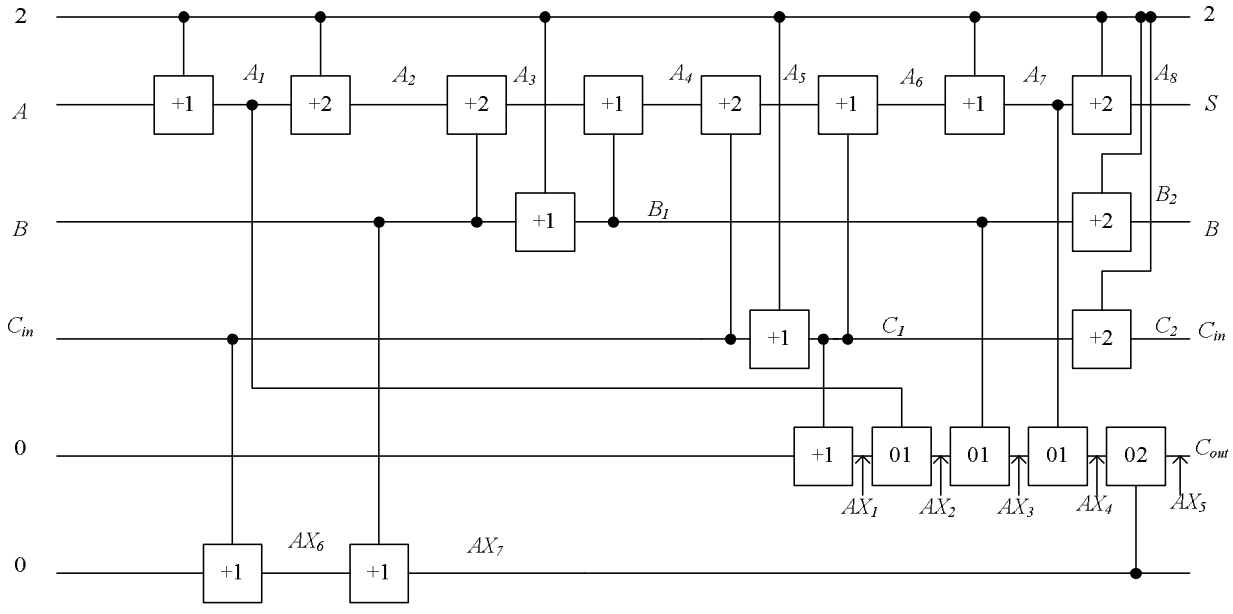


Fig. 2 The proposed realization of quantum ternary full-adder.

Table 3 Simulation result for the proposed realization.

A	B	C_{in}	A_1	A_2	A_3	A_4	A_5	A_6	A_7	S	B_1	B_2	C_1	C_2	AX_1	AX_2	AX_3	AX_4	C_{out}	AX_6	AX_7
0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	1	2	1	1	0	2	1	1	1	1	0	0	0	0
0	0	2	1	0	0	0	2	2	0	2	1	0	0	2	0	0	0	0	0	1	1
0	1	0	1	0	0	1	1	1	2	1	2	1	1	0	0	0	1	0	0	0	0
0	1	1	1	0	0	1	1	2	0	2	2	1	2	1	1	1	0	0	0	0	0
0	1	2	1	0	0	1	0	0	1	0	2	1	0	2	0	0	1	1	1	1	1
0	2	0	1	0	2	2	2	2	0	2	0	2	1	0	0	0	0	0	0	0	1
0	2	1	1	0	2	2	2	0	1	0	0	2	2	1	1	1	1	1	1	0	1
0	2	2	1	0	2	2	1	1	2	1	0	2	0	2	0	0	0	1	1	1	2
1	0	0	2	1	1	1	1	1	2	1	1	0	1	0	0	1	1	0	0	0	0
1	0	1	2	1	1	1	1	2	0	2	1	0	2	1	1	0	0	0	0	0	0
1	0	2	2	1	1	1	0	0	1	0	1	0	0	2	0	1	1	1	1	1	1
1	1	0	2	1	1	2	2	2	0	2	2	1	1	0	0	1	0	0	0	0	0
1	1	1	2	1	1	2	2	0	1	0	2	1	2	1	1	0	1	1	1	0	0
1	1	2	2	1	1	2	1	1	2	1	2	1	0	2	0	1	0	1	1	1	1
1	2	0	2	1	0	0	0	0	1	0	0	2	1	0	0	1	1	1	1	0	1
1	2	1	2	1	0	0	0	1	2	1	0	2	2	1	1	0	0	1	1	0	1
1	2	2	2	1	0	0	2	2	0	2	0	2	0	2	0	1	1	1	1	1	2
2	0	0	0	2	2	2	2	2	0	2	1	0	1	0	0	0	0	0	0	0	0
2	0	1	0	2	2	2	2	0	1	0	1	0	2	1	1	1	1	1	1	0	0
2	0	2	0	2	2	2	1	1	2	1	1	0	0	2	0	0	0	1	1	1	1
2	1	0	0	2	2	0	0	0	1	0	2	1	1	0	0	0	1	1	1	0	0
2	1	1	0	2	2	0	0	1	2	1	2	1	2	1	1	1	0	1	1	0	0
2	1	2	0	2	2	0	2	2	0	2	2	1	0	2	0	0	1	1	1	1	1
2	2	0	0	2	1	1	1	1	2	1	0	2	1	0	0	0	0	1	1	0	1
2	2	1	0	2	1	1	1	2	0	2	0	2	2	1	1	1	1	1	1	0	1
2	2	2	0	2	1	1	0	0	1	0	0	2	0	2	0	0	0	0	2	1	2

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Novel C-Testable Design for H.264 Integer Motion Estimation

Po-Yu YEH, Bo-Yuan YE, Sy-Yen KUO, and Shyue-Kung LU

Department of Electrical Engineering
Graduate Institute of Electronics Engineering, BL522
National Taiwan University
Taipei, Taiwan 106
sykuo@cc.ee.ntu.edu.tw

Abstract - H.264/AVC is the latest video compression standard with highest coding efficiency, and the chip-area are increased significantly, especially the Integer-Motion-Estimation (IME) block. Thus the testability of H.264-IME is becoming more and more important. Currently, the scan-chain with Automatic Test Pattern Generation (ATPG) method is very popular for testing H.264-IME block, but the test time usually increases as the design grows. In this paper, a C-testable DFT (Design-for-Testability) scheme at bit-plane level is proposed by using the Iterative-Logic-Array (ILA) architecture for the largest part in H.264-IME block. A simple BIST (built-in self-test) circuit is also proposed due to the ILA architecture, and the number of test pattern (NTP), hardware overhead (HO) and delay-time overhead (DTO) are only about *192*, *4.70%* and *5.56%* respectively. The proposed DFT scheme reduces the test time and test cost significantly.

I. Introduction

H.264/AVC is the latest video compression standard developed by Joint Video Team (JVT). In real-time video applications, the multiple-reference-frames (MRF) and variable-block-size (VBS) properties of H.264-IME not only provide better compression efficiency, but also increase the chip-area rapidly. Thus how to test such a large design is becoming more and more important. In current VLSI industry, ATPG (Automatic Test Pattern Generation) for scan-chain DFT is very popular and useful for general applications. But the test-cost and test-pins may grow as chip gate-count increases. There are already several relative researches on testing ME block for previous MPEG standards [1-3]. In [1], a test scheme that uses TVDG (Test Vector Dependence Graph) on bit-sliced array is proposed. It can improve the test efficiency and fault coverage with a few test patterns. A BIST circuit for testing the ME process-element (PE) is proposed in [2]. A system-level error tolerance approach is proposed in [3] to analysis and compensate the performance degradation. Unfortunately, these methods above don't deal with the thorny VBS and incremental word-length problems for the latest H.264 standard. Thus, the general scan-chain method currently is still very popular for H.264-IME.

It is well-known that the general logic testing problem is NP-complete [4]. But to test a circuit with repeated cells such as H.264-IME block, the novel Iterative-Logic-Array (ILA) testing scheme may be helpful. In the past, an ILA is C-testable with a constant number of test patterns regardless of the number of the cells [5-15]. The most important condition for a C-testable ILA is that the single cell function is bijective, where the bijective property means that the input/output function of the cell is one-to-one mapped. However, the traditional stuck-at and the single cell fault models (SCFM) may not be sufficient in current VLSI industry. Thus more comprehensive sequential fault models are established [16-19]. In [19], a novel Realistic Sequential Cell Fault Model (RS-CFM) is proposed, and it is a more comprehensive, cell-level model suitable for ILA testing. In [14], it shows that a constant number of test vectors are sufficient for fully testing a *k*-dimensional ILA for sequential faults if the cell function is bijective. Furthermore, this concept is extended under RS-CFM fault model [20]. Based on these researches, we will just focus on how to meet the conditions of C-testable ILA under SCFM for simplification. Then this ILA will still be C-testable for sequential faults under RS-CFM. In order to meet the requirements of C-testable ILA under SCFM, each module (cell) should have the bijective property.

In this paper, an effective ILA test scheme at bit-plane level is proposed for testing the largest part in H.264-IME block. Furthermore, it's very easy to design the built-in-self-test (BIST) circuit due to the C-testable ILA structure. The NTP, HO and DTO of the proposed DFT are only about *192*, *4.70%*, and *5.56%* respectively, and the test time and cost are reduced significantly. The ILA test scheme and original H.264-IME design is reviewed in section II and III respectively. The effective DFT design and the BIST circuit for the largest part of H.264-IME block are proposed in section IV and V individually. In section VI, the performance results are analyzed. Finally, the conclusions are given in section VII.

II. Review of the ILA Architecture

Based on SCFM, we assume that a module's behaviour is invariant over time, even if it is faulty. A faulty module's function may deviate from the correct one in any manner, as long as it remains combinational. That is, we are testing for permanent combinational faults only. As long as each module in ILA has same number of I/O pins and conforms to bijective property, the NTP of an ILA

will be a constant same as testing a single module. Take Fig. 1 for an example, N bijective modules with n-bit I/O pins are cascaded one after another as an 1-D ILA, and the optional D-cells (D-Flip-Flop, DFFs) are for pipelining purpose. For the simplest case without any D-cells, if 2^n exhaustive test patterns (ETPs) are input to I_0, then *Module-1* will also output 2^n ETPs on I_1 due to the *1*-to-*1* mapped property. Similarly, the ETPs will be propagated from I_0, $I_1 \ldots I_{N-1}$ to I_N. So all modules are fully tested with only 2^n ETPs. If each module's output pins are fully or partially pipelined with equal or less than n-bit D-cells, each input pattern on I_0 should be fixed for N clock cycles until correct output test pattern appears on I_N. Therefore, the controllability and observability for each module in ILA are achieved easily no matter the D-cells exist or not.

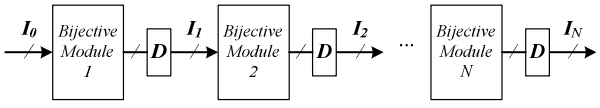


Fig. 1: 1-dimensional iterative logic array.

III. Original H.264-IME Architecture

The widely used Full-Search ME (FSME) algorithm [21] and M-parallel 2D SAD (Sum of Absolute Difference) tree architecture [22] are used in practical H.264-IME design (Fig. 2), where M is the number of parallel structures that can be adjusted until the throughput of block-matching is large enough. This architecture consists of register buffer (*RegBuf*) and M block-matching-modules (*BMMs*). The *RegBuf* can be separated into current, reference and prepare pixel arrays. These pixel arrays are made of DFFs with the capabilities of shifting *1*-pixel upward/downward and M-pixel leftward. Assume the depth of each pixel is d-bit, and there are total $16\times(31+2M)$ pixels in *RegBuf*. Each *BMM* will perform one 16×16 macro-block (MB) matching and then output *41* SAD values (SADs). The M *BMM* modules will do M block-matching operations between current MB (Cur-MB) and reference MBs (Ref-MBs) simultaneously.

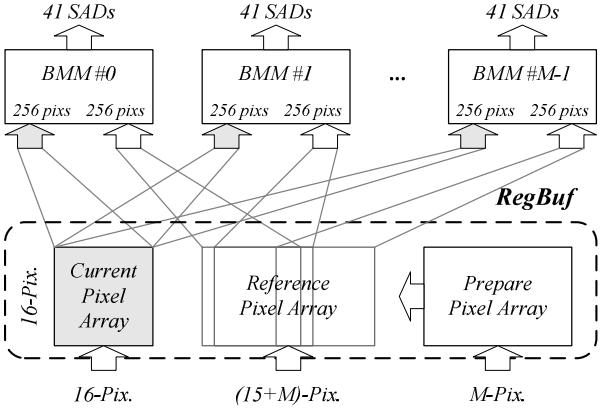


Fig. 2: The *M*-parallel 2D SAD tree architecture.

As shown in Fig. 3, a *BMM* module could be separated into two parts: the first part is called SAD_{b16} that computes *16* basic SADs (*S0~S15*), and the second part is called SAD_{v25} which accumulates all basic SADs to generate the other *25* VBS SADs (*S16~S40*).

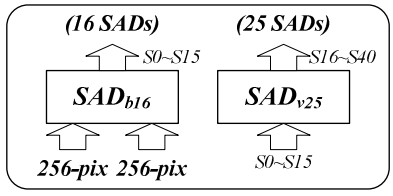


Fig. 3: A BMM consists of SAD_{b16} and SAD_{v25} modules.

IV. The DFT for the SAD_{b16} Part

The H.264-IME block can be divided into *RegBuf*, SAD_{b16} and SAD_{v25} parts. For each part, there are many repeated cells or modules, thus the ILA test schemes can be applied on these parts individually. The SAD_{b16} part is not only the largest part, but also the most irregular part in H.264-IME block. Therefore, only the DFT scheme for the SAD_{b16} part will be discussed in this paper for the limited space. Due to the word-level half-adder (HA) and full-adder/subtractor (FA/FS) cells are the main elements in H.264-IME block, the useful HA/FA/FS bijective cells with scalable word-length are proposed first.

A. Scalable HA/FA/FS Cells

The gate-level circuits of *1*-bit *HA*, *tFA*, and *tFS* cells are shown in Fig. 4(a), (b) and (c) respectively, and they only consist of inverter, NAND and Exclusive-OR (XOR) logic gates. The transistor-count of every logic gate is labelled on its symbol, thus there are total *18*, *42* and *44* transistors in *1*-bit *HA*, *tFA* and *tFS* cells respectively. Notice that the *tFA*/*tFS* cells are almost the same as general used *1*-bit FA/FS cells except the bypassed output *bo*. These *1*-bit cells are the basic units of word-level cells. The w-bit HA (HA_w) and FA/FS (FA_w/FS_w) cells are shown in Fig. 4(d) and (e). In Fig. 4(d), the lateral I/O function of HA_w cell can be expressed as $Ao=(Ai+ci)_{mod}\ 2^w$, where the expression $(X)_{mod}2^w$ will return the remainder when the bus X divides 2^w. If the carry-in ci of HA_w cell is *0*, the output bus Ao will be equal to input bus Ai. Obviously, it is a bijective cell that satisfies *1*-to-*1* mapped I/O function. On the other hand, if the carry-in is assigned to *1*, it is still bijective due to the permutation I/O function. Take $w=2$ for example, if the exhaustive values *0*, *1*, *2* and *3* are inputted into Ai, the Ao bus will output *1*, *2*, *3*, and 0, respectively. Thus we know the permutation function $Ao=(Ai+ci)_{mod}\ 2^w$ is bijective. The number of test pattern for each ci value is 2^w, thus the total NTP is 2^{w+1}. The FA_w/FS_w cells are shown in Fig. 4(e), notice that the carry-in of FA_w and FS_w cells are always fixed at *0* and *1* respectively. Both of them can be modified to become bijective tFA_w/tFS_w cells by using basic *tFA*/*tFS* cells (i.e. $Bo=Bi$). Similar to the HA_w case, for each fixed value of Bi, the mapping from Ai to Ao is permutation again. Thus the I/O function $\{Ao,Bo\}=\{(Ai \pm Bi)_{mod}\ 2^w, Bi\}$ is still bijective, where the comma in the $\{...\}$ expression is to combine several bits or buses into an integrated bus. Take *1*-bit tFA_w as an example, if the value pairs *0/0*, *0/1*, *1/0* and *1/1* are input to *Ai/Bi*, the output buses *Ao/Bo* will output *0/0*, *1/1*, *1/0* and *0/1* respectively. The NTP for tFA_w/tFS_w cells are both 2^{2w}, and notice that the proposed cells are all bijective with scalable word-length w.

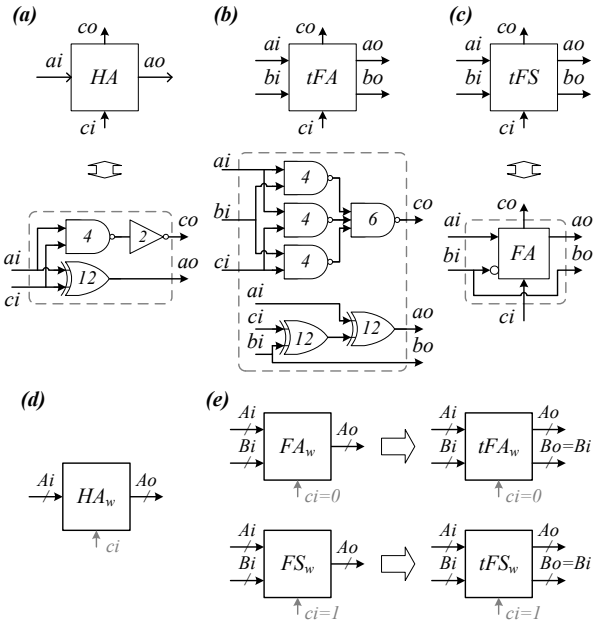


Fig. 4: (a) *1*-bit *HA* cell. (b) *1*-bit *tFA* cell. (c) *1*-bit *tFS* cell. (d) HA_w cell. (e) FA_w/FS_w and tFA_w/tFS_w cells.

B. C-Testable ILA for SAD_{b16}

Every SAD_{b16} module is composed of *16* independent SAD_{4x4} modules, and these SAD_{4x4} modules calculate the basic SADs *S0~S15*. The original circuit of SAD_{4x4} module is shown in Fig. 5. The input buses *C0~C15* and *R0~R15* indicate the *d*-bit pixel values from Cur-MB and Ref-MB respectively. There are *16* AD_w and *15* FA_w cells in the SAD_{4x4} module, where the AD_w cell calculates the absolute difference value between two pixels. The notation *w* for AD_w/FA_w cells represents the *w*-bit computation, which varies from *d* to *d+4*. The *15* FA_w cells with incremental word-lengths act as an adder-tree to accumulate the results of all AD_w cells, thus the final output bus *S* will be *(d+4)*-bit wide. Every AD_w/FA_w cell's output should be pipelined with *D*-cells in high-speed applications, and they are omitted in figure for simplification.

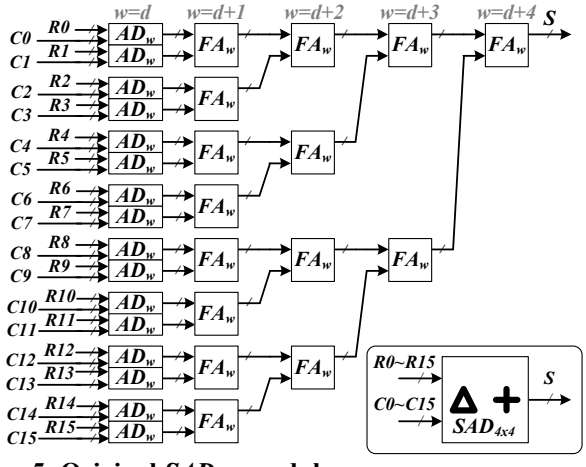


Fig. 5: Original SAD_{4x4} module.

Obviously, the SAD_{4x4} module can't satisfy the bijective property due to the different I/O pin numbers. The NTP for whole SAD_{4x4} module is 2^{32d}, which is an enormous number and the incremental word-length should be solved under ILA-based test scheme. At first, the original AD_w cell at bit-plane level should be introduced. Fig. 6(a) and (b) show the symbol and detail circuit for AD_w module respectively, where the input signals *Ri* and *Ci* are both assumed to be *w*-bit wide. The AD_w module consists of a *(w+1)*-bit FS_w cell, *w*-bit inverter/multiplexer, and one *w*-bit HA_w cell for subtraction and absolute-value operations. The bus *E* will be a negative number as the MSB (most significant bit) output signal *sn* equals to *1*, and then the 2's complement operation is performed for the output bus *A* (the bus *E* is bitwise inverted as $\overline{E}$ and added by *1*). On the contrary, the output *A* will be equal to *E* if *sn* equals to *0*.

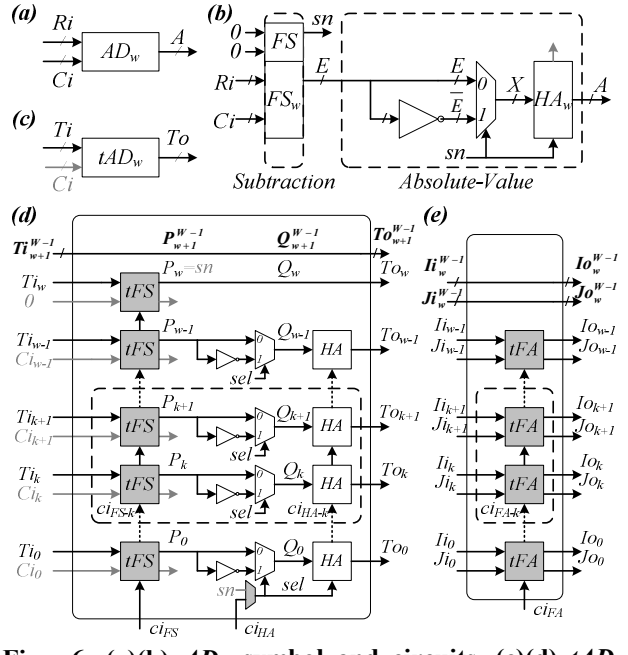


Fig. 6: (a)(b) AD_w symbol and circuits. (c)(d) tAD_w symbol and circuits. (e) The bit-plane level tFA_w cell.

In order to solve the incremental word-length problem, the number of I/O pins for all cells will be kept the same. Fig. 6(c) shows the symbol of testable tAD_w cell, and the bit-plane level circuits for tAD_w and tFA_w cells are proposed in Fig. 6 (d) and (e) respectively. For both cells, there are only *w*-bit actual operation circuits for *W*-bit input buses, where *W* is equal to *d+4* and *w* will vary between *d* to *W* (*w* is equal to *d* for all tAD_w cells). The notations Ti_k and Ti_L^H indicate the *k-th* bit and the sub-bus from *L-th* bit to *H-th* bit in *Ti* bus respectively. Most redundant input pins are bypassed to output directly. See the tAD_w cell in Fig. 6(d), the FS_w cell is replaced with tFS_w cell (composed of *tFS* cells) and extra one *2*-to-*1* multiplexer is inserted for connecting ci_{HA} signal to *sel* in test mode. Fig. 6(e) indicates all tFA_w cells in adder-tree with various word-lengths from *d+1* to *d+4*. For the ILA architecture at bit-plane level, all tAD_w/tFA_w cells are partitioned into *W* bit-planes and they can be tested from *0-th* to *(W-1)-th* bit-plane individually. In test mode, the input buses *Ii* and *Ji* of tFA_w cells may be connected to the outputs of two tAD_w cells (*To*), or the outputs of another tFA_w cell (*Io*/*Jo*). The global carry-in signals $\{ci_{FS}, ci_{HA}, ci_{FA}\}$ can control the carry-ins for all tFS_w, HA_w and tFA_w cells respectively, and the signals $\{ci_{FS\text{-}k}, ci_{HA\text{-}k}, ci_{FA\text{-}k}\}$

represent the carry-in signals for *k-th* bit-plane. When testing *k-th* bit-plane, the Ci_k signal can be controlled directly by the shift-DFFs in Cur-MB, but all carry-ins $\{ci_{FS-k}, ci_{HA-k}, ci_{FA-k}\}$ may not be accessed directly. Next, it will be shown that not only all carry-ins for *k-th* bit-plane will be controllable, but also all carry-outs for *k-th* bit-plane will be observable. For the controllability of the carry-ins in *k-th* bit-plane, the global carry-ins $\{ci_{FS}, ci_{HA}, ci_{FA}\}$ can only be *{0,1,0}* or *{1,0,1}* two cases. As $\{ci_{FS}, ci_{HA}, ci_{FA}\}$ are set to *{0,1,0}* (or *{1,0,1}*), the lower bit-planes' input signals Ti_0^{k-1}/Ci_0^{k-1} for all tAD_w cells also have to be fixed at *0s*/*1s* (or *1s*/*0s*); where *0s* and *1s* indicate all-zeros and all-ones buses respectively. Relative signals for the two cases are listed as below:

$$\begin{aligned} &Vertical: \ \{ci_{FS}, ci_{HA}, ci_{FA}\} = \{ci_{FS-k}, ci_{HA-k}, ci_{FA-k}\} \\ &\qquad = \{0,1,0\} \ (or \ \{1,0,1\}) \\ &Lateral: \ Ti_0^{k-1} = \overline{Ci_0^{k-1}} = P_0^{k-1} = To_0^{k-1} = Ii_0^{k-1} = Ji_0^{k-1} \\ &\qquad = Io_0^{k-1} = Jo_0^{k-1} = 0s \ (or \ 1s) \quad (Q_0^{k-1} = 1s) \end{aligned} \tag{1}$$

Therefore, for the *0-th~(k-1)-th* bit-planes, not only lateral outputs will be equal to lateral inputs, but also vertical inputs and outputs will be equal in the two cases. Therefore the goal to propagate and control all carry-ins for *k-th* bit-plane is achieved. Since the controllability of $\{Ci_k, ci_{FS-k}, ci_{HA-k}, ci_{FA-k}\}$ is achieved, the lateral function for each tAD_w/tFA_w cell should be bijective for testing *k-th* bit-plane in *1*-D ILA architecture. Obviously, if ci_{FA-k} is fixed, the *k-th* bit-plane of tFA_w cell can be regarded as a smaller bijective tFA_w cell or a bypassed cell. Thus the mapping from $\{Ii_k, Ji_k\}$ to $\{Io_k, Jo_k\}$ is always bijective. In fact, the bijective property also can be found in the *k-th* bit-plane in tAD_w cell.

Property 1:

In Fig. 6 (d), if $\{ Ci_k, ci_{FS-k}, sel, ci_{HA-k}\}$ are all fixed, the mapping $Ti_k \Rightarrow To_k$ is bijective for every bit-plane (where the "$\Rightarrow$" notation indicates the mapping direction).

Proof:

Obviously, when *k* is between *w* to *W-1*, the mapping from Ti_k to To_k should be bijective due to the bypass function or the bijective *tFS* cell with fixed $\{Ci_k, ci_{FS-k}\}$. When *k* is between *0* to *w-1*, the mappings $Ti_k \Rightarrow P_k$ and $Q_k \Rightarrow To_k$ are both bijective because of the bijective *tFS*/*HA* cells with fixed $\{ Ci_k, ci_{FS-k}, ci_{HA-k}\}$. If *sel* is fixed at *0 (*or *1*), the mapping from P_k to Q_k is still bijective due to the bypass (or inverted) function. Thus the mappings from Ti_k, P_k, Q_k to To_k are all bijective. Consequently, the bijective property is proved for each bit-plane. □

Notice that as we test *k-th* bit-plane, any faulty carry-outs $\{ci_{FS-(k+1)}, ci_{HA-(k+1)}, ci_{FA-(k+1)}\}$ will be observed via the lateral outputs of *(k+1)-th* bit-plane. Moreover, no carry-out should be detected if there is no *tFS* or *tFA* cell in the *(k+1)-th* bit-plane. Finally, the C-testable ILA structure for $tSAD_{4x4}$ module is shown in Fig. 7. The AD_w and FA_w cells are replaced with tAD_w and tFA_w cells respectively. There are *23* additional *W*-bit multiplexers for changing circuits to straightforward data path in test mode, and there is also extra *2*-to-*1* multiplexer in each tAD_w cell. Thus there are total *(23W+16)* additional *2*-to-*1* multiplexers in $tSAD_{4x4}$ module.

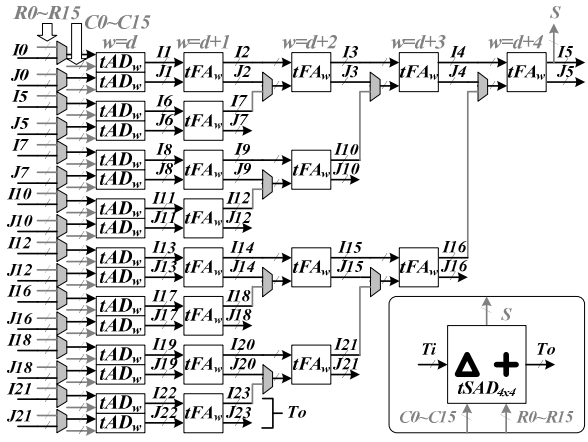


Fig. 7: The C-testable ILA for SAD_{4x4} module.

For each bit-plane, the tAD_w and tFA_w cells can be viewed as *23* bijective stages. As Ci_k and $\{ ci_{FS-k}, ci_{HA-k}, ci_{FA-k}\}$ are fixed, only 2^2 ETPs are needed to input from $\{I0_k,J0_k\}$ (Ti_k), $\{I1_k,J1_k\}$, $\{I2_k,J2_k\}\ldots\{I22_k,J22_k\}$, to $\{I23_k,J23_k\}$ (To_k). For all possibilities of Ci_k and vertical carry-ins (2^2 cases), the NTP for fully testing *k-th* bit-plane is only 2^4. Thus the total NTP for the *W* bit-planes in $tSAD_{4x4}$ module is only $W \times 2^4$. Since the bijective $tSAD_{4x4}$ module at bit-plane level is proposed, all independent modules can be cascaded as a larger ILA ($tSAD_{b16}$) as shown in Fig. 8.

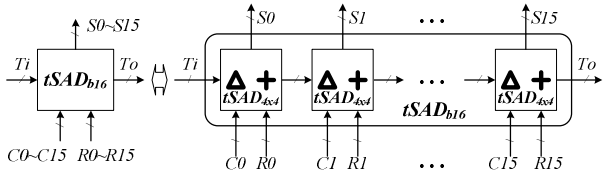


Fig. 8: The $tSAD_{b16}$ module with ILA architecture.

V. Build-In-Self-Test

Due to the property of test pattern propagation, to test all modules in an ILA needs only one test pattern generator (TPG) and one output response analyzer (ORA). Notice that the HO of the BIST can be ignored as the number of modules increases. Since the ILA test scheme for the $tSAD_{b16}$ module is proposed already, the BIST circuits can be added easily as shown in Fig. 9. For entire $tSAD_{b16}$ part, the TPG and ORA will generate input test patterns and compare output patterns respectively. Finally, the ORA will output the test-error signal *Err* to indicate the test is passed or not. Due to the NTP of proposed DFT schemes are reduced significantly, it is very easy to design the TPG/ORA by using simple LFSRs (Linear Feedback Shift Register) for compressing test sequences. The proposed BIST circuit is quite simple and effective to test itself inside the chip. Without extra test equipments, it reduces test time and costs significantly.

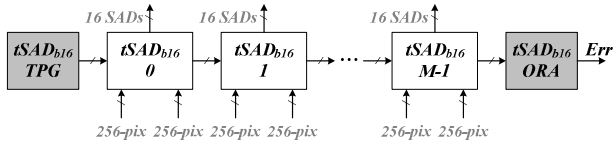


Fig. 9: The BIST for entire $tSAD_{b16}$ part.

VI. Performance Analysis

In this section, the NTP, HO and DTO of the $tSAD_{b16}$ module are discussed. For simplicity, the hardware area is estimated by transistor-count only, and the delay time of all logic gates (multiplexer, inverter, *NAND* and *XOR* gates) are assumed the same as *1* delay-gate. Let the NTP, HO, and DTO for certain module *X* be expressed as *NTP*(*X*), *HO*(*X*) and *DTO(X)* respectively. And the expressions *DT*(*X*), *T*(*X*), and *OT(X)* represent the delay-time, number of transistors, and overhead-transistors of module *X*, individually. In Fig. 7, there are total extra *23W+16* multiplexers in $tSAD_{4x4}$ module, so the HO and NTP can be inferred as below. (Assume *Inv* and *Mux2* are the cell names of inverter and *2*-to-*1* multiplexer respectively.)

$$\begin{aligned} T(AD_w) &= (d+1)\times T(FS) \\ &\quad + d\times[T(HA)+T(Inv)+T(Mux2)+T(DFF)] \\ T(SAD_{4x4}) &= 16\times T(AD_w) \\ &\quad + \sum_{i=1}^{i=4} 2^{4-i}\times(d+i)\times[T(FA)+T(DFF)] \\ HO(tSAD_{4x4}) &= \frac{OT(tSAD_{4x4})}{T(SAD_{4x4})} = \frac{(23W+16)\times T(Mux2)}{T(SAD_{4x4})} \\ NTP(tSAD_{4x4}) &= W\times 2^4\,|_{W=d+4} = 192\,|_{d=8} \end{aligned} \quad \textbf{(2)}$$

Before estimating the DTO of the $tSAD_{4x4}$ module, the delay time of the basic arithmetic cells should be inferred first. In Fig. 4(a)~(e), the critical paths of the HA_w/tFA_w /tFS_w cells include from carry-in *ci* to the MSB of *Ao* bus (through *w-1* ripple carries), thus the total delay time can be estimated as below:

$$\begin{aligned} DT(HA_w) &= [DT(NAND)+DT(Inv)]\times(w-1) \\ &\quad + [DT(XOR)]\times 1 = 2w-1 \\ DT(tFA_w) &= [2\times DT(NAND)]\times(w-1) \\ &\quad + [2\times DT(XOR)]\times 1 = 2w \\ DT(tFS_w) &= DT(Inv)+DT(tFA_w) = 2w+1 \end{aligned} \quad \textbf{(3)}$$

Notice the tFS_w cell has one more delay-gate than the tFA_w cell due to the inverted *Bi* bus. In order to estimate the DTO of the $tSAD_{4x4}$ module in Fig. 7, it is necessary to find the extra delay on the critical path between two sequential cells (DFFs). Because all tAD_w and tFA_w cells in $tSAD_{4x4}$ are pipelined with DFFs, the critical path should be in the tAD_w cell. See Fig. 6(d), the extra multiplexer (for selecting *sn*) and the lateral input multiplexer (not shown in figure) will both induce extra delay time. Thus the DTO of tAD_w and $tSAD_{4x4}$ can be estimated:

$$DTO(tSAD_{4x4}) = DTO(tAD_w) = \frac{DT(Mux2)+DT(Mux2)}{DT(tFS_w)\,|_{w=d+1} + DT(Inv)+DT(Mux2)+DT(HA_w)\,|_{w=d}} \quad \textbf{(4)}$$

Obviously, the NTP, HO and DTO for the $tSAD_{4x4}$ and $tSAD_{b16}$ modules will be the same due to the ILA architecture. If the depth *d* of input pixel is *8*-bit, the performances are listed in Table. 1. The NTP, HO and DTO for the $tSAD_{b16}$ part are only *192*, *4.70%* and *5.56%* respectively.

Table. 1: Performance results for the $tSAD_{b16}$ module.

Full-Search (d=8-bit)	*NTP*	*HO*	*DTO*
$tSAD_{b16}$	*192*	*4.70%*	*5.56%*

VII. Conclusion

In this paper, an effective testing scheme is proposed at bit-plane level for the largest part of H.264-IME block. Due to the ILA architecture, the number of test pattern is a constant even if there are more and more modules to test. The NTP, HO and DTO of the proposed test schemes are only *192*, *4.70%*, and *5.56%* respectively, and the test-time and cost are reduced significantly.

References

[1] W. P. Marnane, W. R. Moore, "Testing a motion estimator array," Proceedings of the International Conference on Application Specific Array Processors (1990), pp. 734-745, 5-7 September 1990.

[2] D. Li, M. Z. Hu and O. A. Moharned, "Built-In Self Test Design of Motion Esitmation," Regular Session H: System Testing & Verification, p.349-352, 2004.

[3] C. Hyukjune, A. Ortega, "Analysis and testing for error tolerant motion estimation," 20th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'05), pp. 514-522, 3-5 October 2005.

[4] H. Fujiwara and S. Toida, "The complexity of fault detection problems for combinational logic circuits," in IEEE Trans. Computers, Vol. C-31, No. 6, pp. 555-560, June 1982.

[5] W. H. Kautz, "Testing for faults in combinational cellular logic arrays," Proc. 8th Annu. Symp. Switching, Automata Theory, 1967, pp. 161-174.

[6] P. R. Menon and A. D. Friedman, "Fault detection in iterative arrays," IEEE Trans. Computers, Vol. C-20, pp. 524-535, May 1971.

[7] A. D. Friedman, "Easily testable iterative systems," IEEE Trans. Computers, Vol. C-22, pp. 1061-1064, Dec. 1973.

[8] T. Sridhar and J. P. Hayes, "Design of easily testable bit-sliced systems," IEEE Trans. Computers, Vol.C-30, No. 11, pp. 842-854, November 1981.

[9] R. Parthasarathy and S. M. Reddy, "A testable design of iterative logic arrays," IEEE Trans. Computers, Vol.C-30, No. 11, pp. 833-841, November 1981.

[10] E. M. Aboulhamid and E. Cerny, "Built-in testing of one-dimensional unilateral iterative arrays," IEEE Trans. Computers, Vol. C-33, No. 6, pp. 560-564, June 1984.

[11] A. Vergis and K. Steiglitz, "Testability conditions for bilateral arrays of combinational cells," IEEE Trans. Computers, Vol. C-35, No. 1, pp. 13-26, January 1986.

[12] F. Lombardi, "On a new class of C-testable systolic arrays," Integration, Vol. 8, pp. 269-283, 1989.

[13] C. W. Wu and P. R. Cappello, "Easily testable iterative logic arrays," IEEE Trans. Computers, Vol. C-31, No. 6, pp. 640-652, May 1990.

[14] C. Y. Su and C. W. Wu, "Testing Iterative Logic Arrays for Sequential Faults with a Constant Number of Patterns," IEEE Trans. Computers, Vol. 43, No. 4, pp. 495-501, April 1994.

[15] S. K. Lu, C. W. Wu, and S. Y. Kuo, "Enhancing testability of VLSI arrays for fast Fourier transform," IEE Proc. E, Vol. 140, No. 3, pp. 161-166, May 1993.

[16] J. Galiay, Y. Crouzet, and M. Vergiault, "Physical versus logical fault models in MOS LSI circuits, impact on their

testability," IEEE Trans. Computers, Vol. 29, No. 6, pp. 527-531, June 1980.
[17] A. K. Pramanick and S. M. Reddy, "On detection of delay faults," Proc. IEEE Int'l Test Conf., pp. 845-856, 1988.
[18] G. L. Smith, "Model for delay faults based upon path," Proc. IEEE Int. Test Conf., pp. 342-349, 1985.
[19] M. Psarakis, D. Gizopoulos, A. Paschalis, and Y. Zorian, "Sequential fault modeling and test pattern generation for CMOS iterative logic arrays," IEEE Trans. Computers, Vol. 49, No. 10, pp. 1083-1099, October 2000.
[20] S. K. Lu, "Delay Fault Testing for CMOS Iterative Logic Arrays with a Constant Number of Patterns," IEICE Trans. Inf. &Syst., Vol. E86-D, No.12 December 2003.
[21] T. Komarek, P. Pirsch, "Array architectures for block matching algorithms," IEEE Trans. Circuits and Systems, Vol. 36, Issue 10, pp. 1301-1308, October 1989.
[22] C. Y. Chen, S. Y. Chien, Y. W. Huang, T. C. Chen, T. C. Wang, and L. G. Chen, "Analysis and Architecture Design of Variable Block Size Motion Estimation for H.264/AVC," IEEE Trans. Circuits and Systems, 53(2): 578-593, February 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Modified Physical Configuration to Compensate Parasitic Effects in High Speed Systems

Saad Bin Abul Kashem[1], Salahuddin Raju[2], and Md Ishfaqur Raza[3]

[1,3]Department of Electrical and Electronic Engineering.
East-West University, Bangladesh
[2]Department of Electrical and Electronic Engineering.
American International University-Bangladesh, Bangladesh
E-mail: [2]raju@aiub.edu, [3]iraza@ewubd.edu

Abstract – With systems operating at higher frequencies, parasitic effects are taking larger shares of the voltage and timing budget of a circuit. Also smaller devices have shrunken landscape for components, increasing coupling between critical physical features. These undesired loading on otherwise uniform transmission lines introduce impedance discontinuities which degrades signal quality and strains the performance metrics. This paper introduces a concept of compensating coupling effects by modifying transmission line physical characteristics. The modification in the line dimensions is calibrated to compensate the lumped equivalent of the coupling effect. High speed system spice simulation and S-parameter analysis has demonstrated the effectiveness of this methodology.

I. Introduction

The effect of parasitics in the performance of a circuit is a function of the frequency content of the signals in the system. With microwave circuits operating at GHz frequencies and the evolution of typical circuit applications into GHz ranges, the management of the impact of parasitic effect is vital to the efficient operations of these circuits. Size of circuits is also decreasing rapidly resulting in very close proximity of components in a circuit. Busy board layouts force the placements of vital traces next to pins, solder balls, and vias [1]. In a much smaller landscape, as in nanometre scale integrated circuits, with decreasing sizes of transistors, increased number of devises in a chip, and lower power supply voltage, the issue of coupling becomes more acute. Due to smaller dimensions and crowded landscape, metal in VLSI circuits are placed closer, increasing capacitive coupling. This results in the inevitable coupling between metal layers, wires, and vias.

With the power supply decreasing, the budget of the device that is attributed to coupling noise is shrinking. Overall, the effect of coupling is set to grow precipitously with advances in board and circuits designs. Often architects and designers provide and follow guidelines designed to reduce coupling effects. However, too often, physical constraints make it impossible to follow those design rules. These constraints, such as pin pitch and wire pitch are not negotiable and hence, coupling and crosstalk become unavoidable, straining on the budget of the circuit and the device, penalizing performance of the circuit [2].

The impact of coupling to a transmission line is the change in the characteristic impedance of the transmission line. The discontinuity in the impedance is due to the increase in reactive loading of the line at the point of cross coupling between the line and the parasitic component. The increased reactive loading changes the transmission line impedance at the coupling point. This issue is typically tackled with design rules that require minimum separation between lines and components. Some designs require select material, such as low-k dielectric in higher metal layers in VLSI designs to reduce coupling coefficient between lines.

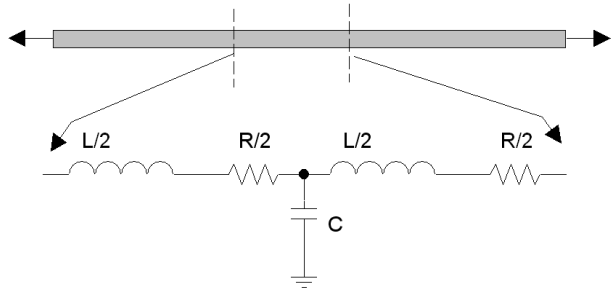


Fig. 1. T model representation of a transmission line

This paper presents a concept to compensate the impact of coupling by modifying transmission line physical design. Spice modelling of a transmission line system along with s-parameter analyses is done to demonstrate the compensation technique.

II. Compensation Technique

In engineering analysis at radio and microwave frequencies, a transmission line is defined by the telegrapher's equation instead of circuit theory, as the component dimensions become comparable to signal wavelength. With frequencies in the range of 0.1-10 GHz, typical components lengths that can be considered to be lumped will be no longer then several millimetres long. Fig. 1 shows the pi model representation of a segment of a transmission line [3].

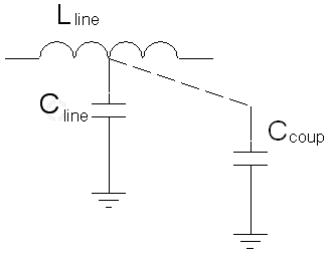


Fig. 2. Capacitive coupling of a transmission line

The characteristic impedance is a function of the resistance R, inductance L, conductance G, and capacitance C, all in per unit length values. At high frequencies, both the resistive and conductive element is neglected as those parameters are negligible compared to the reactive elements. The impedance is written as,

$$Z = \sqrt{\frac{R + jwL}{G + jwC}} \approx \sqrt{\frac{L}{C}}$$

The characteristic impedance of a segment of the transmission line considered as a lumped element is defined by the inductance L_{line} and capacitance C_{line} of that segment.

$$Z_{uncpl} = \sqrt{\frac{L_{line}}{C_{line}}}$$

When the line is placed close to a device or component, the coupling component C_{coup} is introduced, which is in this case capacitive. The impedance of the segment Z_{total} of the line will be reduced because of the increased capacitance C_{total}, which is equal to $C_{line} + C_{coup}$.

$$C_{total} = C_{line} + C_{coup}$$

$$L_{coup} = L_{line}$$

$$Z_{total} = \sqrt{\frac{L_{line}}{C_{total}}}$$

$$Z_{total} = \sqrt{\frac{L_{line}}{C_{line} + C_{coup}}}$$

In this paper, a compensation technique is introduced where the line capacitance is reduced to compensate for the coupled capacitance. With the reduced line capacitance, the total capacitance will now equal the initial line capacitance. The overall impedance of that segment of the transmission line will be corrected and thus discontinuity removed.

III. Transmission Line Analysis

Consider a differential strip line, which is built of two parallel and equally spaced lines, both of uniform width and cross section. It is assumed that one of the lines is adjacent to a pin which has resulted in coupling between one of the differential line and the pin.

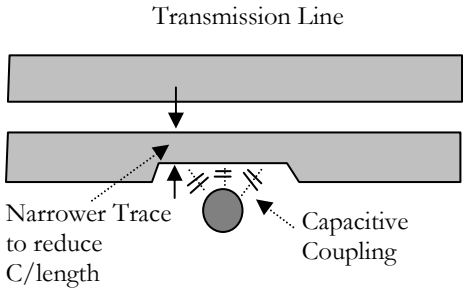


Fig. 3. Narrowed trace to compensate for coupling.

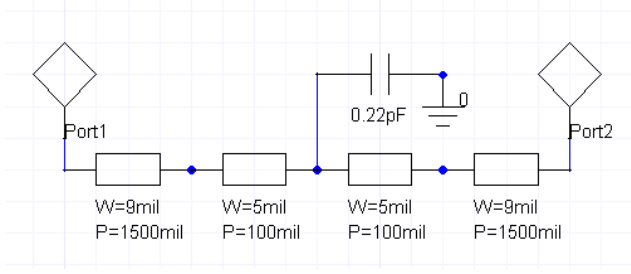


Fig. 4. Model for S-parameter analysis of a capacitively coupled micro strip transmission line

A segment of trace next to the trace is narrowed, as shown in Fig. 3. The narrowed section will have lesser per unit length capacitance. The length of the segment and the shrinking of the width are selected to obtain a reduction in the line capacitance which will equal the added pin coupling. The narrowed trace thus compensates the additional coupling of the pin. The impedance of the line segment is expected to remain uniform across the length of transmission line.

This compensation technique is only valid in the range of frequencies where the coupled segment of the transmission line is significantly smaller then the wavelength of the signal. A general rule for this is length < λ/20. As the system frequency increases and the wavelengths decreases, the ratio of the wavelength to component size decreases. However, for microwave frequencies and the feature sizes involved, this concept can easily be applied. To be considered as lumped at 10GHz, the size of the feature should be smaller then 1.5 mm, which is less then 1/3rd the dimensions of a via.

IV. Frequency Domain Simulation

The characteristic impedance Z_O of a micro strip transmission line is a function of the dielectric constant ε_r and height H of the substrate, the width of the trace W, and the thickness of the copper T. The impedance along with the per unit length capacitance and inductance is given by [4],

$$Z_O = \frac{87}{\sqrt{\varepsilon_r + 1.41}} \ln\left(\frac{5.98H}{0.8W + T}\right)$$

Consider a setup for measuring the s-parameter matrix of a micro strip transmission line (see Fig. 4). The line is terminated at both ends to match impedance at 50 Ohms. For a 9 mil wide transmission line, the characteristic impedance and all other information is given in Table 1.

Table 1: Characteristic parameters of a micro strip line

Dielectric Constant Er	4.4
Dielectric Thickness (mils) H	6.0
Trace Width (mils) W	9.0
Trace Spacing (mils) S	8.0
Finished Copper Weight (mils) T	1.4
Impedance Zo (Ohms)	51.56
Inductance/Inch Lo (nH/in)	7.244
Capacitance/Inch Co (pF/in)	2.725
Propagation Delay Tpd (nS/in)	0.1406

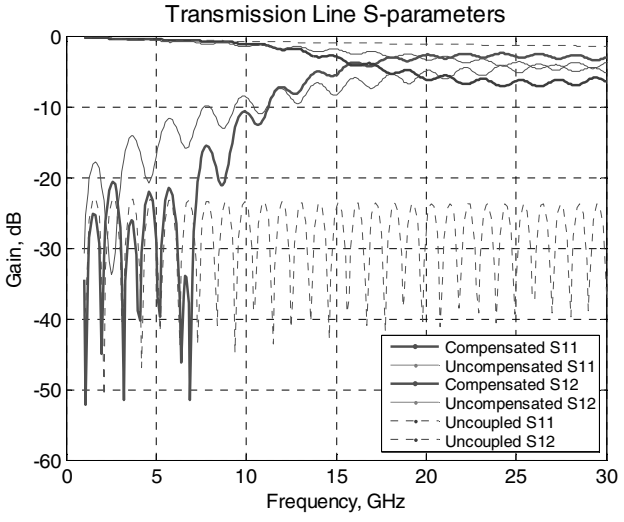


Fig. 5. Transmission line frequency response (modified trace width = 5 mil, reduced from 9 mil)

In this experiment, a 0.220 pF capacitance is coupled to the transmission line. The effect of the parasitic is demonstrated in both Fig. 5 and 6. The S11, reflection parameter at port is calculated in the simulation and illustrated. It is seen that across the large bandwidth (0-30 GHz), S11 is consistently below 20 dB, which is a rather well matched transmission line. With the coupling included the S11 is severely degraded (closer to 0dB). At higher frequencies, around 5 GHz, the S11 has increased to -10dB, indicating significant reflection and impedance mismatch due to the impedance discontinuity.

Compensation technique is then used to demonstrate the proposed concept. Two cases where the trace width at the coupled segment is reduced to 5 and 6 mils are presented here. At 5 mils width, it is calculated that the reduction in the trace capacitance is similar to the amount of coupling capacitance added to the segment. With the compensation added, in Fig. 5, the S11 parameter accounting for reflection at the coupled segment has significantly improved up to 8 GHz, where S11 has dropped by about 12 dB when comparing with the coupled data. This demonstrates the effectiveness of the proposed scheme. The analysis is done again with a trace width of 6 mil. The result is shown in Fig. 6.

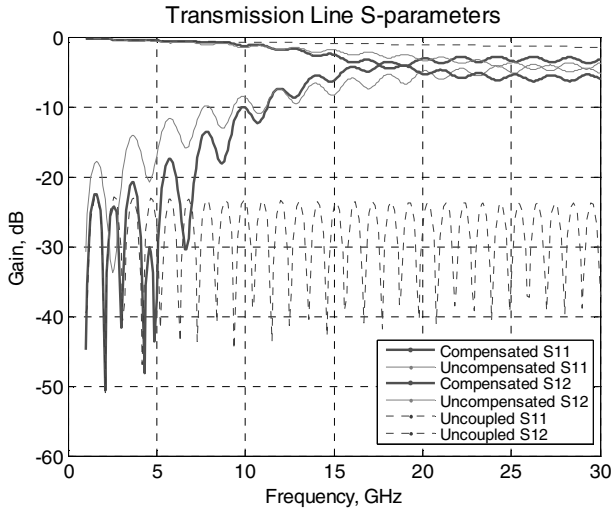


Fig. 6. Transmission line frequency response (modified trace = 6 mil)

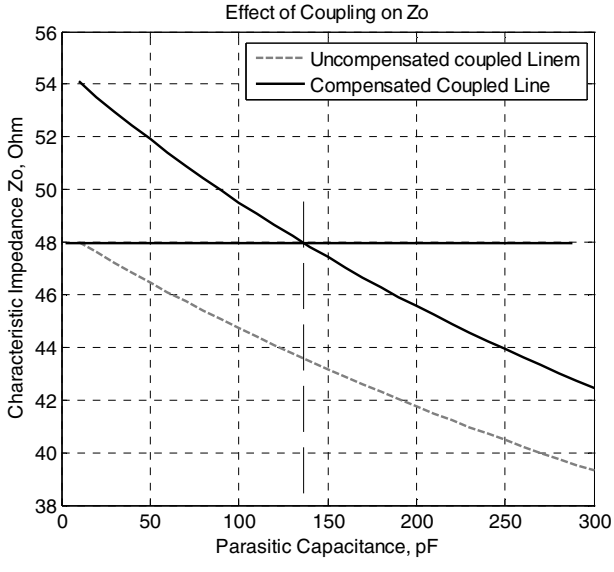


Fig. 7. Change in Zo with coupling

It is shown that the S11 parameter is improved from the coupled case, reducing by about 5 dB. This demonstrates the need to calculate the added capacitance and how the transmission line segment should be modified to compensate the coupled effect, as shown in Fig. 7. For this analysis, the capacitance and inductance of the micro strip is calculated using the following equations, which states the per unit length capacitance of transmission line,

,

$$C_O = \frac{0.67(\varepsilon_r + 1.41)}{\ln\left(\frac{5.98H}{0.8W + T}\right)}$$

V. Time Domain Simulation

The impact of impedance discontinuity is observed in the form of degraded signal integrity in a high speed system. The degradation of the signal quality translates in to smaller eye opening, which results in higher bit error rates in data transmissions. The discontinuity results in both reduced transmitted signal and greater reflections back to the source.

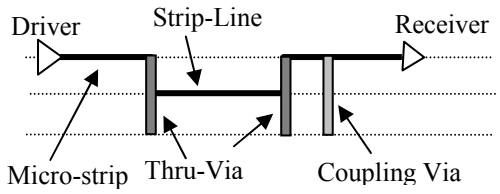


Fig. 8. System for signal analysis in time-domain

Table 2: Strip line characteristic parameters

Dielectric Constant Er	4.4
Dielectric Thickness (mils) H	12.0
Trace Width (mils) W	8.0
Trace Spacing (mils) S	5.0
Finished Copper Weight (mils) T	1.4
Impedance Zo (Ohms)	52.13
Inductance/Inch Lo (nH/in)	9.534
Capacitance/Inch Co (pF/in)	3.508
Propagation Delay Tpd (nS/in)	0.1775
Differential Impedance Zdiff (Ohms)	89.64

The reflections not only translate into higher standing wave ratio and losses, but also intersymbol interference, where reflection from a symbol or data affects the reliability of later data symbols. Degraded transmission translates into reduced signal swings at the receiver which will result in the receiver recording erroneous data.

In this simulation analysis, a strip line transmission line system is modeled (as shown in Fig. 8). The system is driven by a differential signal. In this case, a via was placed outside the differential line pair. The coupling thus only affects one of the wires in the differential line. To avoid complexity, the coupling capacitance between via and adjacent trace of via is considered at the center point of the trace segment. The coupled capacitance is calculated considering the interface between the trace and via and using the equation, where A is the area of the interface and d is the gap (a crude approximation),

$$C_0 = \frac{\varepsilon A}{d}$$

The model that was used in the simulation are defined by parameters in Table 2. The impedance, capacitance, and inductance of the differential strip line are given by

$$Z_O = \frac{60}{\sqrt{\varepsilon_r}} \ln\left(\frac{1.9(2H+T)}{0.8W+T}\right)$$

$$C_0 = \frac{1.41\varepsilon_r}{\ln\frac{3.81H}{0.8W+T}}$$

$$L_0 = 0.001 C_0 Z_0^2$$

$$Z_{diff} = 2\,Z_0\left(1 - 0.347e^{\left(-2.9\,\frac{S}{H}\right)}\right)$$

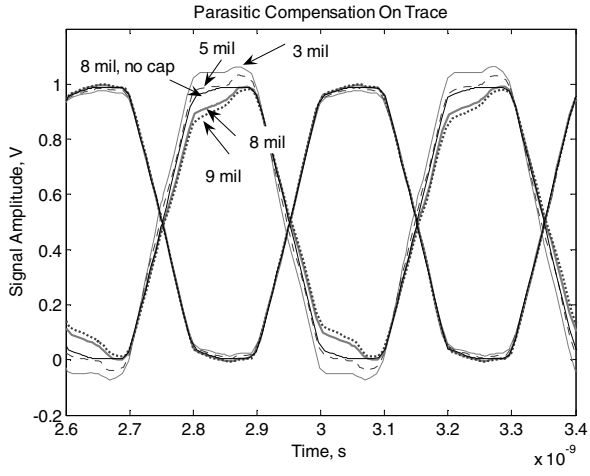


Fig. 9. Changes in signal shape due to different amount of compensation by trace narrowing.

The result of the time-domain signal integrity analysis is shown in Fig. 9. The waveform in this data is observed before the coupled segment, which highlights the effect of coupling on the reflection coefficient, instead of transmission coefficient. As can be seen in the signal data swing, the best coupling compensation is achieved with the 5 mil trace. It should be noted that without compensation the effect of reflection reduces the signal strength, which is due to the negative reflection coefficient, as a result of impedance being lower then the characteristic impedance. The impedance is lower due to the increase in overall capacitance of the segment with the addition of the coupled capacitance.

VI. Conclusion and Future Work

The data has shown very conclusively the benefit of transmission line modifications to compensate the effect of parasitic capacitance. It is also observed that the advantages are bandwidth constrained. For the dimensions discussed in this paper, the proposed methodology is good up to frequencies around 12 GHz. It can be further improved if the interface between the coupling feature and the transmission line is minimized. The application of this methodology should be considered for inclusion into CAD tools where designers would automatically modify transmission line definitions to compensate for coupling. For future work, the methodology should be demonstrated in the lab. The proposed should also be extended to microwave design where susceptibility to impedance discontinuity is more significant.

References

[1] C. R. Paul, *Introduction to Electromagnetic Compatibility,* Wiley Series in Microwave and Optical Engineering, Jan 2006.

[2] H. W. Johnson, *High Speed Signal Propagation: Advanced Black Magic*, Prentice Hall PTR, 1st edition, February 24, 2003.

[3] D. M. Pozar, *Microwave Engineering,* John Wile & Sons, Inc, 2nd edition, 1998.

[4] IPC-2251 Document, "Design Guide for the packaging of High Speed Electronic Circuits," 2003

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Improved VLSI Circuit Performance using Localized Power Decoupling

Laila S. Sraboni, Ophelia Mohaimen, Rezwana H. Mustazir, and [1]S. M. Salahuddin, and [2]Md Ishfaqur Raza

Department of Electrical and Electronic Engineering.
East-West University, Bangladesh
E-mail: [1]sms@ewubd.edu, [2]iraza@ewubd.edu

Abstract – Power droop in the silicon is a major cause for system performance degradation. Higher frequency of operation and reduced power levels are limiting the timing and voltage budget, which is designed in circuits to account for system noise, which includes voltage drooping due to inductive losses. Novel techniques are evolving to compensate for these losses at all levels, starting from motherboard, package, down in to silicon. Due to lack of available space and design constraints, decoupling at the die level is very limited. In this paper a proposal is made to provide for decoupling at the CMOS levels, right where the power is needed. Advanced technology for DRAM capacitors is proposed for use in this paper for the decoupling strategy. Simulation of sub 100 nm multi-metal layer circuit demonstrates the advantage of proposed localized decoupling.

I. General Information

CMOS dissipates power only when it switches from high to low or vice versa. The simultaneous switching of millions/billions of transistors draws huge currents from the system DC power supply (Fig. 1). The speed at which the system responds to this need is in-sufficient to drive the device efficiently. As process geometries shrink and devices are driven at higher frequencies, adequate and fast power supply is starting to become a gating factor to system performance. Inherent parasitic components in the power transmission line, from the DC-DC converter to the CMOS devices, are delaying the response time of the power supply. Without the appropriate voltage levels at the gates and the drains of the devices, the devices are able to perform.

Different technologies are available to help the system provide the necessary power to the circuit. For motherboards and packages, low inductance capacitors are available which are easily integrated into a system [1]. Advanced decoupling technology, embodied in devices such as LICA (Low Inductance Capacitor Array), Monolithic Capacitors (e.g.-X7R,X5R) , Super Capacitors have extremely low internal resistance or ESR and ESL, high efficiency (up to 97-98%), high output power, extremely low heating levels, and improved safety. Currently, the elements of monolithic ceramic capacitors are joined to each other by solder layers and are stacked on each other to minimize the value of equivalent series resistance and equivalent series inductance.

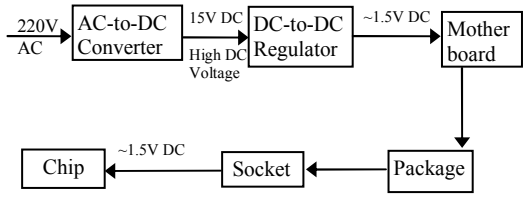


Fig. 1. System Power Distribution

Small deviations from the ideal behavior of a device can become significant when circuits are operating under 'fast corner' conditions, i.e. high frequency, high current, or temperature extremes. However, all these low ESR/ESL technology can only serve to compensate the power droop due to motherboard, package, and connector level impedance. These new decoupling technologies are not enough to the increasing frequencies and large current switching. Challenge remains trying to provide for decoupling at the integrated circuit level. In this paper, a methodology to introduce decoupling right at the transistor is proposed. The crux of the concept is to use DRAM capacitor technology, such as trench and stacked capacitors, as a local capacitance bank to deliver power to the devices right when the switching takes places, without being delayed by the multiple levels of metal layer parasitics.

II. CMOS Power Requirements

The power requirements and dissipation in CMOS can be divided into static and dynamic components. The static component is independent of the clock. The main component in static power consumption is leakage power. The focus of this paper is not on leakage power, but on the power that is a function of clock frequency, or more correctly stated, is a function of the switching of the CMOS circuit transistors. This power is widely known as dynamic power and is composed of the short circuit power and the swirching power, P_s.

$$P_s = \alpha C_L V_{dd}^2 f$$

where α is the activity factor, C_L is the transistor load capacitance (~another CMOS gate input capacitance), V_{dd} circuit power supply, and f is the signal frequency.

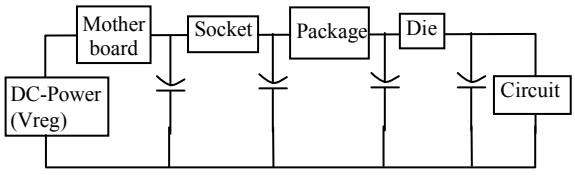


Fig. 2. System level power delivery decoupling solution.

In dynamic power the harmonic content of the current drawn by the circuit is of vital importance for analysis in this paper. The equation for dynamic power drawn does not relate to the speed at which this power is sourced or sink-ed. Rather, it relates to how many times the circuit switches state in a second. The speed at which the current charges or discharges the load is a function of the *RC* time constant of that part of the circuit. In the case of CMOS transistors, the *RC* time constant depends on the sorce and drain capacitance at the driver end and the gate capacitance at the receiver end. For newer processes, with smaller dimensions of source and drain in the transistors, the capacitances will be smaller. With shorter channel length, the resistance will also reduce. Overall, the time constant of the transistor decreases.

A. The inductive drop and large *di/dt*

The decrease in the RC time constant of a circuit will cause the switching current to race quicker, resulting in a large *di/dt*. From one perspective this is good as it will increase the switching speed, which is the ultimate goal. However, the faster switching speed translates into higher harmonics in the signal. In time domain the large *di/dt* translates into large voltage drop across any inductance that may lie in the path of the current,

$$V = L\frac{di}{dt}$$

There will also be resistive drop in any resistance. However, the resistive drop is independent of the rate of change. Overall, the total voltage drop due to the resistive and inductive effect of motherboard, package, socket and silicon [2]can be shown,

$$V_{dr} = \left(L_m\frac{di}{dt} + R_m\right) + \left(L_p\frac{di}{dt} + R_p\right) + \left(L_S\frac{di}{dt} + R_S\right)$$

Expressing the total drop as the sum of all inductive and resistive components, V_{dr} is then equal to

$$V_{dr} = \sum_{n=1}^{3}(L_n\frac{di}{dt} + IR_n)$$

Therefore, the voltage available at the circuit can be written as,

$$V_{eff} = V_{dd} - V_{dr}$$

With a faster switching current, i.e. larger *di/dt*, the voltage drop inductive component will increase, resulting in a reduced power supply to the IC. As the effective power supply voltage (V_{eff}) is shared across a large number of components in the silicon, performance and activities of a lot of transistors will be affected when the voltage drop is large. It is for this reason that V_{dr} must be minimized to improve the performance of a system.

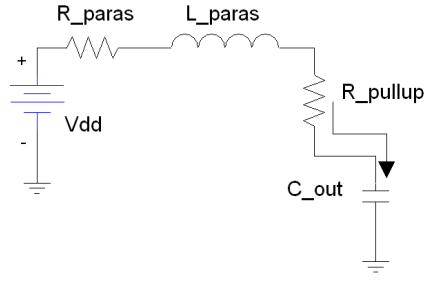


Fig. 3. Charging of a mosfet input gate

B. Improving current supply

The options available to reduce this voltage drop are

- Reduced inductance of the current path (from power supply to circuit).
- Reduced *di/dt* content of the current drawn.
- Find an alternate source for the fast switching current (alternate source is decoupling, Fig. 2).

All power supplies and their path to the circuit are optimized to have a low inductance path – but some inductance cannot be avoided. Reducing *di/dt* is not an option if the intent of a next generation circuit is to provide faster and higher performance. The third option has been addressed to certain extent, where newer technologies are providing better decoupling to products [4]. In this paper we present localized decoupling to circuits with the granularity of the device down to a single CMOS transistor pair.

Depending on the switching action of the circuit, it can either be a charging of the device capacitance (getting power from power supply, V_{dd}) as shown in Fig. 3 or it can be discharging of the device gate to ground. Nevertheless, the discharging of a device, in the case of an inverter, essentially translates to the device at the output of the inverter charging. Therefore, whenever, the clock transitions, we can assume some part of the circuit will be drawing current from the power supply, or the decoupling sources. If there are cases of simultaneous switching of CMOS devices, the current will increase proportionately. From a design perspective, it is therefore desirable to stagger the switching activities of the device to spread the switching actions across a time window instead of having all the devices switch at the same time.

Better system designs target low resistive loss by creating large power and ground planes in motherboard and packages, to reduce inductance and resistance. It includes large number of vias in packages and connectors. A large section of solder ball and pin arrays in a package are devoted to ground and power to reduce the average resistive and inductive impedance. Within integrated circuits, VLSI designs are using low K-dielectric in the upper metal layers to reduce capacitive coupling, thus varying decoupling are achieved. Though they effectively minimize the voltage drop, their response time is slow in reaching the CMOS circuitry which is starved off power. Hence, local decoupling right at the transistor level can

act quickly to nullify the voltage droops. However, they offer small capacitance values in nano-scale dimensions.

III. Incorporating DRAM Capacitor Technology

The DRAM capacitor technology can be incorporated in a fast switching CMOS circuitry to design capacitance, thus reducing power starvation at the CMOS. DRAM capacitor technology is mature in the form of 1T1C (one transistor one capacitor) configuration. The capacitors can be either in the stacked (Fig. 4) or trench (Fig. 5) configuration [5].

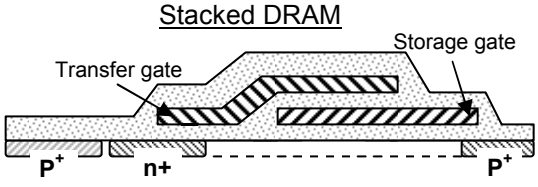


Fig. 4. Stacked DRAM Capacitor Configuration

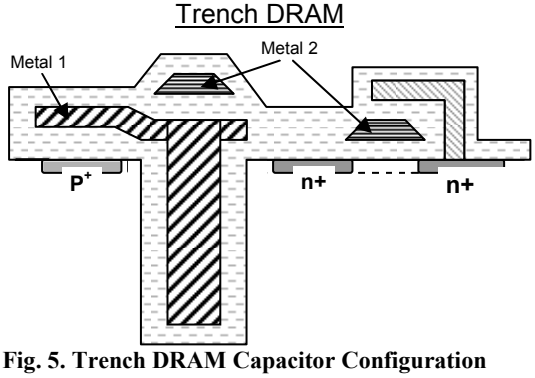


Fig. 5. Trench DRAM Capacitor Configuration

The above shows two different DRAM technologies, each with their advantages and limitations. Compared to a stacked concept, the trench cell offers a large storage capacitance, simpler circuitry and more importantly it consumes the least space. This can be employed between two CMOS, in the STI (silicon trench isolator) between the active and the ground plane to attain maximum output. The trench wall, being an extension of the source in PMOS, is formed after the gate is patterned at low temperature like source formation. A rod-like structure from bias sinks down the trench, see Fig. 6. Rest is filled with dielectric,. New innovations including Checkerboard (CKB), hemispherical silicon grains (HSG) combined with the use of a bottle-shaped trench, increase the surface area of the trench capacitor and thus the capacitance. These capacitors have the ability to store about 10 fF each. With aggressive CVD (chemical vapor deposition) technology drive, this capacitance can be further increased by using higher k (>40) dielectric. Current CVD approach using tantalum oxide provides for a dielectric constant of 25.

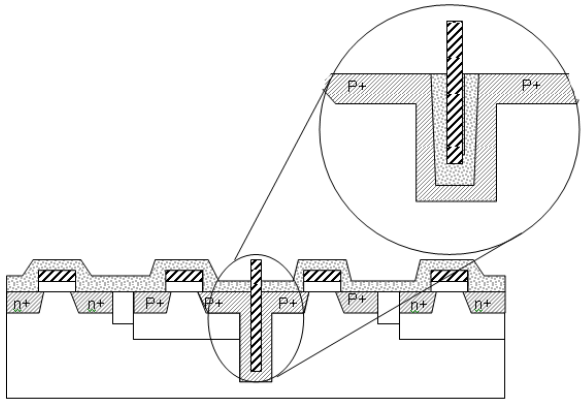


Fig. 6. Trench decoupling capacitor between inverters.

The ITIC DRAM configuration retention time is about a couple of seconds, which means the leakage amount is negligible for use in decoupling technology, particularly systems operating at GHz range. These capacitors are not expected to provide current support for more then a couple of pico seconds as the next layer of decoupling support will kick in to provide the charge support. Also, as the DRAM capacitor pitch is more aggressive then the microprocessor integrated circuit, which makes it easier for DRAM capacitor technology to be incorporated for decoupling usage models (see table 1).

Table 1 MPU and DRAM pitch for next few years, [6]

Year	Metal1 M1 Pitch, nm	MPU physical gate length, nm	DRAM ½ pitch, nm
2008	118	22	57
2009	104	20	50
2010	90	18	45

In a typical CMOS inverter, a PMOS and an NMOS is implanted in a Si Substrate creating an n-tub (n-well) and a p-tub (p-well) inside it. There is also the twin tub technology where two separated tubs appear into a lightly doped substrate to obtain higher channel mobility. However, parasitic BJT affects causes latching effect results in latching base current flows form parasitic PnP to parasitic NPN. Therefore VLSI circuits have widely used deep trench isolation techniques to isolate n-Channel and p-Channel. In this technique oxide is thermally generated on the bottom and the walls of trench which is then refilled with Poly-Silicon or SiO2. The deep trench proposal allows an isolated structure deposited by CVD and eliminates time consumption, hi-temperature and lateral diffusion issues [7]. With the need for higher k dielectric for larger capacitance, the CVD method is preferred. A newer method for deposition which gaining ground is atomic layer deposition, ALD [8]. The trench can also be implemented by Well formation technology, which is accomplished for single well, twin well and retrogrades. For the trench capacitor, the retrograde methodology is preferable as it will control lateral diffusion and have a better definition of the well.

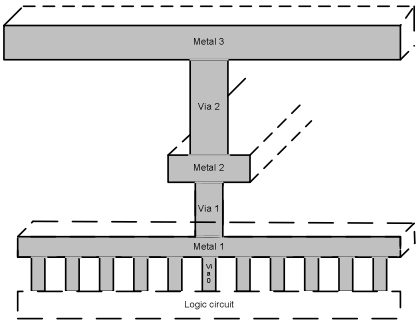


Fig. 7. Three metal layer interconnect assuming orthogonal configuration between layers.

It is critical to minimize the equivalent series resistance and inductance (ESR, ESL) of the DRAM capacitors. Large pitch between metal layers connecting the capacitor to the V_{dd} line will results in a high ESL, while too small a cross section of the trace will result in high ESR. The ESL and ESR of decoupling capacitor show significant degradation with frequency. As the device is operated at higher frequency, the equivalent series resistance of localized decoupling capacitor increases due to skin effect. The equivalent impedance of the capacitor is

$$|Z| = \sqrt{R^2{}_{ESR} + \left(\omega L_{ESL} - \frac{1}{\omega C_c}\right)^2} \text{ where, } \omega = 2\pi f \ .$$

The impedance of the capacitor is minimum at resonance. The operating frequency range of the capacitor should fell near the resonant frequency.

IV. Simulation Data and Analysis

A simulation analysis has been done to demonstrate the advantage of local decoupling technology. The value of the decoupling capacitor is based on the charge stored by a single DRAM capacitor, which is assumed to be 10 fF.

It is also assumed that the capacitor will be shared by about 4 back to back CMOS invertors. The capacitor will be connected to the source of the PMOS device. The RLC parameters of the metal interconnect is calculated using 90nm technology. To keep the analysis at a basic level, only 3 metal layers are assumed (Fig. 7).

At high frequency, skin effect becomes significant. Most of the high frequency currents are crowded within the skin depth of the wire. For systems operating at GHz frequencies, the significant components that define the signal transition edges can be well into the 30-40 GHz. In this paper skin depth is calculated for 40GHz for the realization of high frequency effect.

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} = \sqrt{\frac{1.7\times10^{-8}}{\pi\times40\times10^{9}\times4\pi\times10^{-7}}} = 0.33\mu m$$

Copper is assumed to be the interconnect of choice in different metal layers. Roughness, dishing and trapezoidal effect of the top surface of metal was not considered here. At high frequency two resistance R_{dc} and R_{hf} should be considered [9]. Total dc resistance of a metal line is

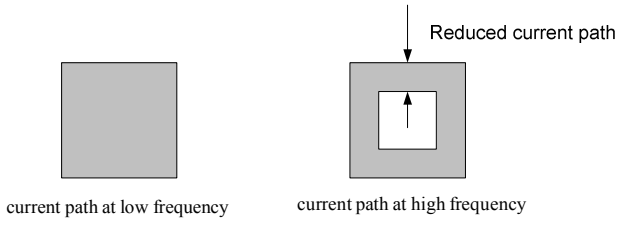


Fig. 8. Skin depth constraint at high frequency

$$R_{dc} = \frac{\rho l}{wt}$$

where l is the length, w is the width and t is the thickness.

High frequency resistance component

$$R_{hf} = \frac{\rho l}{2\delta(w+t)}$$

As the low frequency and high frequency resistive components are orthogonal, the effective high frequency will be [9].

$$R_{ac} = \sqrt{R_{dc}^2 + (\kappa R_{hf})^2}$$

Inductance of metal wires is also considered here using

$$L_{M1} = l\frac{\mu_0}{2\pi}\left(\frac{8h}{w} + \frac{w}{4h}\right) [10]$$

l is the length of the conductor, h is average height from the substrate.

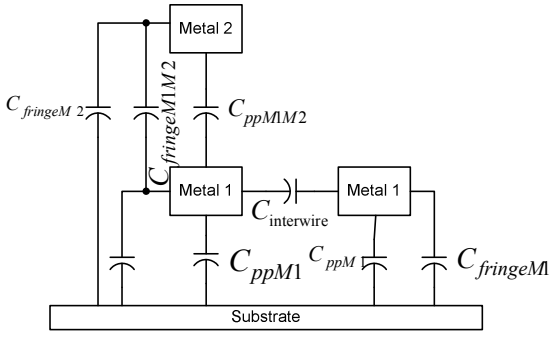


Fig. 9. Capacitive modeling of multi layer interconnect

In this paper only inter wire capacitance and parallel plate capacitance is considered. There is also fringe capacitance which needs to be accounted for in accurate simulations.

An ADS (advanced design system) analysis of a power deliver system was implemented. BSIM4 models were used for NMOS and PMOS transistors. The transistors were taken to be 90nm channel length. Capacitors were attached to the source node of two inverters connected back to back, each driving another inverter, to simulate a capacitive load. The model consisted of a motherboard, package, and a socket. Adequate decoupling was placed on the system, at the motherboard and package. The response time of the system level capacitors were slow enough to not affect the power delivery at the die.

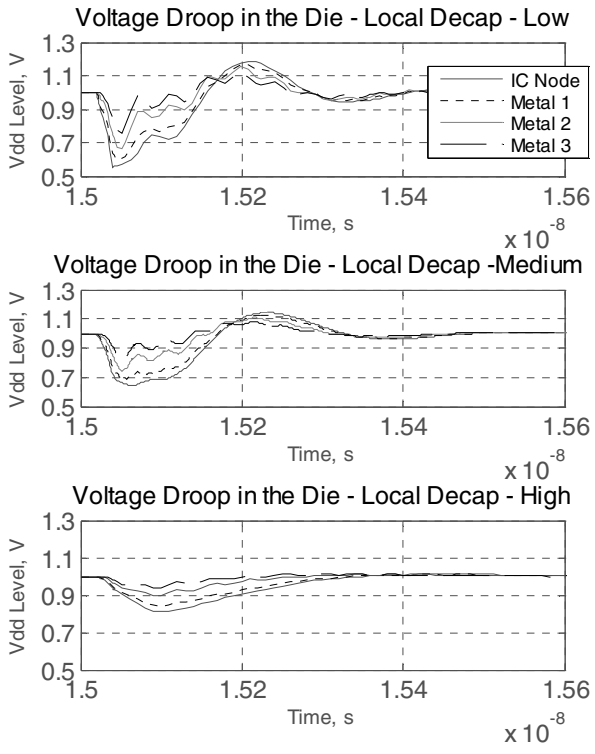


Fig. 10. Power droop analysis of an Integrated Circuit with localized decoupling

In the model three different level of decoupling were attached. The high level was set to 10 fF, medium to 1fF, and low to 0.001fF. The result of the analysis is shown in Fig. 10, which shows droops at different points in the 3 metal layers. As expected, the droop will be worst at the metal 1 layer.

It is seen in Fig. 10 that the voltage droop reach their minimum within 10ps. This level of decoupling cannot be provided by any capacitors located outside the die. Any decoupling placed in white spaces within the die will have to make their way through multiple metal layers which are also constrained by the ESR and ESL of the capacitive path. The analysis above shows an improvement of the voltage level by about 30%. This will have a significant impact on the performance of the device.

V. Conclusion and Discussion

A new decoupling strategy is presented here by bringing together two parallel semiconductor technologies. The capacitor technology is mature for DRAM, while system level decoupling and die white space decoupling is not sufficient to support nano-dimension transistor circuits with 10s pico second transition rates. The trench and stacked capacitor model can be successfully designed right next to transistors without taking more space then the isolation transistors that are part of the design guidelines today. It is demonstrated that the placement of the capacitors to the extent of their values in their current DRAM configuration, provides good amount of decoupling reducing droops in the tens pico-second ranges.

For future work and next step, an aggressive design needs to be implemented in silicon with up to 10 metal layers, where the benefits of the localized decoupling will be ore prominent. A fabrication of the device will provide the critical data which can then be used to use this concept as necessary rule for circuits that support high di/dt.

References

[1] J. Chen and L. He ,"Efficient In-Package Decoupling Capacitor Optimization for I/O Power Integrity", IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, vol. 26, pp. 734 – 738, Apr. 2007.

[2] M. Popovich et al.,"Efficient placement of distributed on-chip decoupling capacitors in nano-scale ICs", ICCAD 2007, IEEE/ACM International Conference on Computer Aided-Design, pp 811-816, Nov. 2007

[3] N. Nanju, T. Budell, C. Chiu, E. Tremble, and I. Wemple, "The effects of on-chip and package decoupling capacitors and an efficient ASIC decoupling methodology", 54the proceedings on Electronic Components and Technology Conference, vol. 1,.pp. 556 – 567, June 2004

[4] M. Xiongfei, K. Arabi, R Saleh, "Novel decoupling capacitor designs for sub-90nm CMOS technology", 7th International Symposium on Quality Electronic Design, ISQED '06. pp 6, March 2006.

[5] T.Y Winarski, "Dielectrics in MOS devices, DRAM capacitors, and inter-metal isolation", IEEE Electrical Insulation Magazine, vol. 17, pp.34 – 47, Nov.-Dec. 2001.

[6] International Technology Roadmap for Semiconductors, 2007 edition, ITRS-Emerging Research Devices and ITRS-Interconnect Edition.

[7] E. Chen, P. Ya-Ling, and W. Tings, "Using effective wet etching technology to improve deep trench shape", IEEE Conference and Workshop on Advanced Semiconductor Manufacturing, 2004. ASMC '04, pp. 244-246. May 2004.

[8] T. Seidel, J. Dalton, Z. Karim, J. Lindner, M Daulesberg, and Z. Wei, "Advances in atomic level deposition technologies", 8th International Conference on Solid-State and Integrated Circuit Technology, ICSICT '06. pp.436-439Oct. 2006.

[9] W Wolf, *Modern VLSI Design System-on-chip Design,* Pearson Education 2005, pp 92-94.

[10] N. Weste and D. Harris, CMOS VLSI Design: A Circuits and Systems Perspective, 3rd ed, Addison Wesley, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Jitter Analysis of a Mixed PLL-DLL Architecture

Md. Sayfullah, Barth Roland, Arpad L. Scholtz
Vienna University of Technology
Institute of Communication and Radio Frequency Engineering
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Qimonda AG, Munich, Germany
Email: md.sayfullah@qimonda.com

***Abstract* — This paper presents the jitter analysis of a mixed mode phase locked loop (PLL) - delay locked loop (DLL) architecture. According to the jitter type, this model can be used as pure PLL or pure DLL or a mixed PLL-DLL. It is observed that mixed mode PLL-DLL architecture can combine the advantage from both PLL and DLL to reduce jitter.**

***Index Terms* — PLL, DLL, mixed PLL-DLL, jitter, jitter transfer function.**

I. INTRODUCTION

ADVANCES in integrated circuit (IC) fabrication technology along with innovative circuit design techniques have led to very high-speed digital systems. As clock frequency is increasing, jitter analysis is becoming paramount important for good PLL/DLL design. A phenomenon known as jitter accumulation makes PLLs more susceptible to power-supply and substrate noises [1], [3], [5]. In cases where a significant amount of noise-generating digital circuitry is present on the same chip, DLLs are preferred because any jitter created by the on-chip noise is completely corrected when a clean reference clock edge arrives at the input of DLL [7], [9], [10]. A clean input reference clock might not be available for some high-speed application in order to reduce the cost. And if the reference clock itself might have significant jitter, the utilization of a DLL does not always guarantee superior jitter performance compared to a PLL.

Phase transfer function of a PLL typically exhibits low-pass filter characteristics where as a DLL exhibits all-pass filter characteristics. Therefore, PLL loop can filter high frequency jitter on the reference clock and on the contrary, DLL does not filter out any noise on reference clock. Moreover phase transfer function of a DLL reveals high-frequency jitter peaking, making it unsuitable for application where there is a significant amount of high-frequency jitter associated with the input reference clock.

Considering all these pros and cons of PLL and DLL, this paper presents a mixed PLL-DLL architecture and its' jitter analysis.

Section II provides a general background on PLL and DLL. Section III gives a detail z-domain model of mixed PLL-DLL architecture and its' simulation result. Section IV concludes with a summary.

II. BACKGROUND

A. PLL

Commonly used PLLs are 2nd order feedback system that generates a clock signal whose output phase is aligned with respect to the phase of an input reference clock. Since phase is the integration of frequency, once the phases are aligned, both phase and frequency are "locked". Block diagram of a commonly-used charge pump based PLL is presented in figure 1, comprised of a voltage-controlled oscillator(VCO), a phase-frequency detector (PFD), a charge pump (CP) and a loop filter (LF).

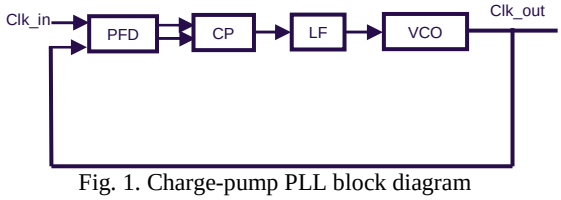


Fig. 1. Charge-pump PLL block diagram

Since PLL is a sampled system, a continuous-time approximation is only valid up to frequencies that are much lower than the reference clock frequency and if the loop bandwidth is much lower than the sampling frequency [11]. Where as z-domain model is more accurate and it also reveals the sampled nature of PLL system. Figure 2 represents z-domain model of the above PLL.

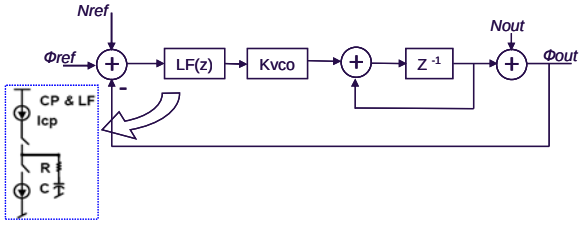


Fig. 2. Z-domain model for PLL

The CP and LF combined transfer function is replaced by its equivalent discrete-time model LF(z), and Kvco is the VCO gain. For a given input reference clock period Tref, and charge-pump current $I_{CP,}$ combined transfer function LF(z) becomes

$$LF(z) = \frac{I_{CP} T_{ref} R(z-\beta)}{2\pi(z-1)} \quad (1)$$

Where, $\beta = \exp(-\frac{T_{ref}}{RC})$

Open-loop transfer function is then

$$LG(z) = \frac{LF(z)K_{vco}}{(z-1)} \quad (2)$$

Input phase transfer function (PTF)

$$\frac{\phi_{out}}{\phi_{ref}}(z) = \frac{LG(z)}{1+LG(z)} = \frac{LF(z)K_{vco}}{z-[1-LF(z)K_{vco}]} \quad (3)$$

Output phase transfer function

$$\frac{\phi_{out}}{N_{out}}(z) = \frac{1}{1+LG(z)} = \frac{z-1}{z-[1-LF(z)K_{vco}]} \quad (4)$$

From figure 2, it is clear that input phase transfer function and input noise transfer function (NTF) is the same. Z-domain PLL model simulation result shows the following input and output noise transfer function.

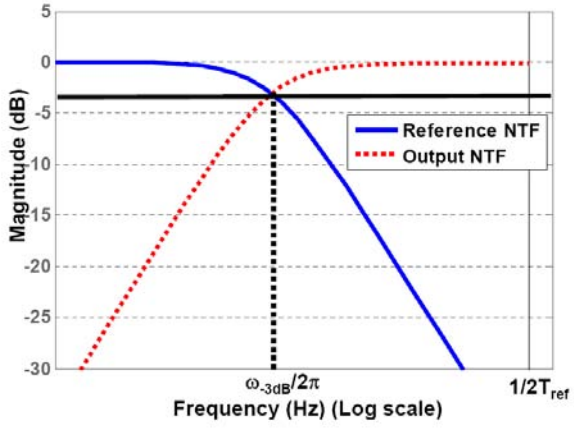


Fig. 3. Reference and output noise transfer function of z-domain PLL model

However z-domain model is only valid up to half of the reference clock frequency due to the sampled nature of the system.

B. DLL

A DLL is also a feedback system and it uses a voltage-controlled delay line (VCDL) instead of VCO to generate an output clock that is a delayed version of the input clock. The delay through the VCDL is a fixed fraction (often 50% or 100%) of the input clock period. While there are several ways to implement a DLL, figure 4 presents a block diagram of a commonly-used charge pump based DLL topology.

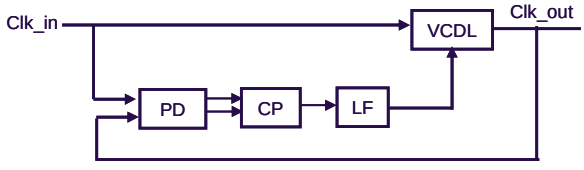


Fig. 4. Charge-pump based DLL block diagram

It is possible to analyze the frequency response of the DLL with a continuous time approximation, where the sampling operation of PD is ignored. However, this approximation fails to capture important phase transfer characteristics that govern DLL operation. Given the inherent discrete time nature of the DLL, a z-domain analysis is more appropriate [11].

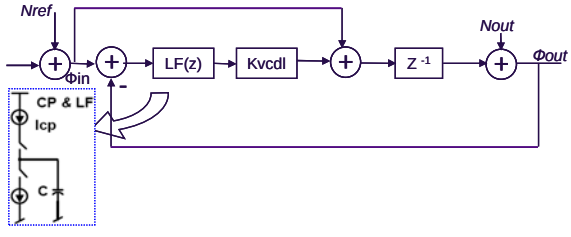


Fig. 5. Z-domain model for charge-pump based DLL

For DLL, transfer function LF(z) becomes

$$LF(z) = \frac{K_{CP} z}{z-1} \quad (5)$$

Where $K_{CP} = \frac{I_{CP} T_{ref}}{2\pi C}$

Based on the above expression for LF, input-to-output phase transfer can then be represented by the following expression:

$$\frac{\phi_{out}}{\phi_{in}}(z) = \frac{(1+K_{VCDL}K_{CP})z-1}{z[z-(1-K_{VCDL}K_{CP})]} \quad (6)$$

Output noise transfer function

$$\frac{\phi_{out}}{N_{out}}(z) = \frac{z-1}{z-(1-K_{VCDL}K_{CP})} \quad (7)$$

Plotting the frequency response of equation (6), figure 6 reveals that the input-to-output phase transfer function has an all pass characteristics, which implies that high frequency jitter on the reference clock passes straight through to the output without being filtering. Actually it is even worse. The high frequency jitter on the reference clock is slightly amplified for frequency higher than

$$\frac{1}{2\pi T_{ref}}\ln\left(\frac{1}{1+K_{VCDL}K_{CP}}\right)$$, which is the frequency of the zero in equation (6). This jitter peaking is inevitable in this type of DLL design, because for positive values of $K_{VCDL}*K_{CP,}$ the frequency of zero, $(1/(1+K_{VCDL}*K_{CP}))$, is always larger than the frequency of the pole, $(1- K_{VCDL}*K_{CP})$, in equation (6).

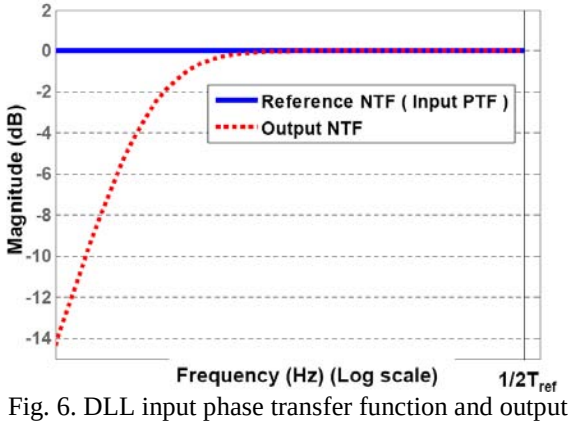


Fig. 6. DLL input phase transfer function and output noise transfer function

III. MIXED PLL-DLL ARCHITECTURE

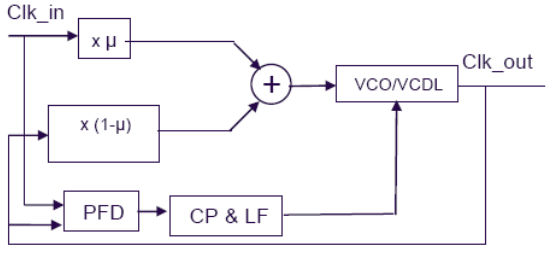


Fig. 7. Mixed PLL-DLL topology. If μ=0, it is a PLL mode, μ =1 means DLL mode and 0<μ<1 is the mixed mode.

This circuit uses a phase mixing interpolator to configure the loop as a PLL, DLL or a mixture of the two. If the on-chip supply is noisy, the high-bandwidth DLL-mode is preferred to avoid supply noise accumulation. On the other hand, if the reference clock is noisy, the low-bandwidth PLL-mode is preferred to take the advantage of low-pass filtering characteristics of the loop. When both noise sources, whose relative amount might not be known *a priori*, are present, a mixed-mode may yield an optimum bandwidth setting to minimize output jitter [6].

Phase mixing interpolator as the first delay element mixes signal energy between output clock and a buffered reference clock based on a 3-bit digital code (μ[2:0]). A mixing weight of 0 corresponds to PLL mode without any reference clock injection (VCO-mode); a mixing weight of 1 corresponds to DLL-mode with full reference clock injection strength (VCDL-mode), and four intermediate mixing weight (0.2, 0.4, 0.6, 0.8) correspond to mixed-mode operations with relative weight strengths between Clk_in and Clk_out (mixed VCO/VCDL-mode).
Z-domain representation of a mixed type PLL-DLL structure is shown in figure 8.

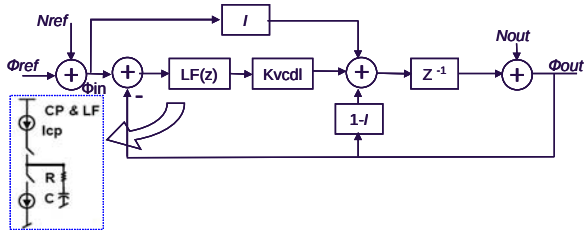


Fig. 8. Z-domain representation of a mixed mode PLL-DLL.

In figure 8, *I* is representing the injection strength of reference clock whose phase is represented by Φ_{in}, and (1-*I*) is the injection strength of Clk_out.

Input phase transfer function

$$\frac{\phi_{out}}{\phi_{in}} = \frac{I + LF(z)K_{VCDL}}{z + LF(z)K_{VCDL} - (1-I)} \quad (8)$$

Where

$$LF(z) = \frac{I_{CP}T_{ref}R(z-\beta)}{2\pi(z-1)}, \beta = \exp(-\frac{T_{ref}}{RC})$$

Output noise transfer function

$$\frac{\phi_{out}}{N_{out}} = \frac{z}{z + LF(z)K_{VCDL} - (1-I)} \quad (9)$$

The simulation results of the PTF and NTF across different mixing weights are shown in the following figures.

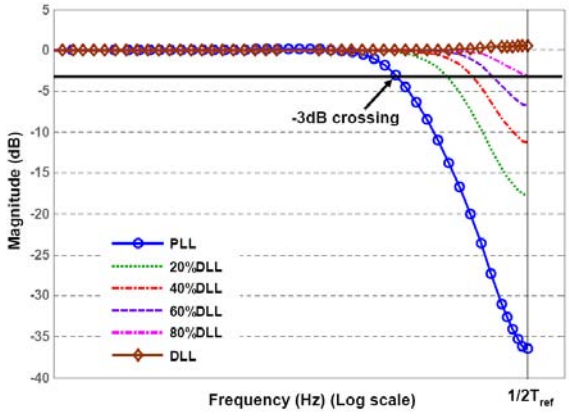


Fig. 9. Input phase transfer function (*Φout/ Φin*) of mixed PLL-DLL

Input phase transfer function plot demonstrate that loop bandwidth of this mixed PLL-DLL can be tuned over a wide range of frequencies, from the lower PLL bandwidth all the way to half the reference clock frequency by changing the relative mixing weight between Clk_in and Clk_out. The jitter peaking in the PLL mode is less than 0.5dB with a large damping ratio. Some jitter peaking is also observed at high frequencies in the DLL mode. This jitter peaking can be reduced by linearly scaling the loop filter resistor with Clk_out injection strength, (1-*I*), such that the zero is removed in DLL mode.

Before we compare the supply noise response of a PLL and a DLL, we must understand how the supply or substrate noise gets into a PLL or DLL. If the supply or substrate reference level changes, a change on the delay of the delay elements in the VCO in a PLL, or the VCDL in a DLL, occurs. A performance measurement called supply sensitivity is usually expressed in a normalized percentage of delay change per percentage of supply change (% delay / % volt). The delay modulation of the delay elements in the VCO/VCDL requires the use of delay element design with good supply noise rejection. A PLL usually has higher supply noise sensitivity than a DLL using the same delay element because a change in supply voltage results in a change in VCO's output frequency, which integrates the phase error within the feedback loop until the loop's correcting action takes effect.

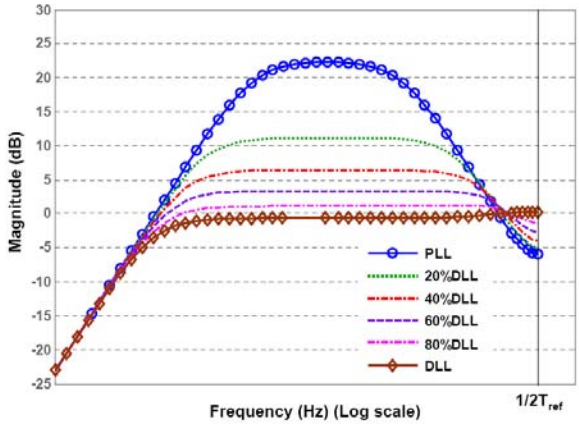


Fig. 10. Supply noise transfer function of mixed PLL-DLL

Since PLL typically has a low bandwidth in order to guarantee stability, the phase error accumulation can result in a large amount output jitter under a noisy supply. On the other hand, a change of a VCDL's supply voltage results in only a delay change occurring once through the delay line. There is no phase error accumulation within the loop due to a fresh reference clock edge that feeds into the VCDL every reference cycle. Therefore a PLL has more supply-induced jitter than a DLL[10].

IV. CONCLUSION

Two of the most significant noise source in today's digital circuitry are –

- Input reference clock noise
- Supply- or substrate-induced noise.

PLL or DLL alone can not attenuate these two noises at the same time. From simulation result, it is shown that a mixed PLL-DLL architecture can attenuate both noises with appropriate mixing weight.

ACKNOWLEDGEMENT

The author wishes to acknowledge the support of all members in Qimonda concept engineering team.

REFERENCES

[1] T. Lee *et al*., "A 2.5-V CMOS delay-locked loop for an 18-Mbit 500-Megabyte/s DRAM," *IEEE J. Solid-State Circuits*, vol. 29, pp. 1491-1496, Dec.1994.

[2] R. Farjad-Rad et al., "A 33-mW 8-Gb/s CMOS clock multiplier and CDR for highly integrated I/Os," in *IEEE Journal of Solid-State Circuits, vol. 39, Sep.* 2004, pp. 1553-1561.

[3] B. Kim, T. Weigandt, and P. Gray,"PLL/DLL system noise analysis for low jitter clock synthesizer design," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, May 1994, pp. 31-38.

[4] M. Meghelli, et al., "A 10Gb/s 5-tap-DFE/4-tap-FFE transceiver in 90nm CMOS," *IEEE International Solid-State Circuits Conference, Digest of Technical Papers*, pp. 80-81, Feb. 2006.

[5] T. Weigandt, B. Kim, and P. Gray, "Analysis of timing jitter in CMOS ring oscillators," in *Proc. Int. Symp. Circuit and Systems*, vol. 4, June 1994, pp. 27-30.

[6] K.-Y. K. Chang, et al., "A 0.4-4Gb/s CMOS quad transceiver cell using on-chip regulated dual-loop PLLs, " *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 747-754, May 2003

[7] G. Chien and P. Gray, "A 900-MHz local oscillator using a DLL-based frequency multiplier technique for PCS applications," in *IEEE Int. Solid-State Circuit Conf. Dig. Tech. Papers*, Feb. 2000, pp.202-203, 458.

[8] B. Kim, T. C. Weigandt, and P. R. Gray,"PLL/DLL system noise analysis for low jitter clock synthesizer design," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, June 1994, pp. 31-34.

[9] R. Farjad-Rad et al., "A 0.2-2-GHz 12-mW multiplying DLL for low-jitter clock synthesis in highly integrated data-communication chips," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2002, pp. 76-77.

[10] G. Chien, "Low-noise local oscillator design techniques using a DLL-based frequency multiplier for wireless applications," Ph.D. dissertation, Univ. of California, Berkeley, 2000.

[11] M.-J. Edward Lee, William J. Dolly, R. Farjad-Rad, "Jitter trasfer characteristics of delay-locked loops-theories and design techniques", in *IEEE journal of Solid-State Circuits, vol. 38, no. 4*, April 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Dual Beam Phased Array Antenna with Wide Scan Angle For Repeater Applications

Ashraf Uz Zaman†, Lars Manholm†† and Anders Derneryd††

† Department of Electrical & Electronic Engineering,
Chittagong University of Engineering & Technology (CUET), Chittagong-4349, Bangladesh
†† Ericsson AB, Ericsson Research, SE-417 56 Göteborg, Sweden
ashraf@cuet.ac.bd; lars.manholm@ericsson.com; anders.derneryd@ericsson.com

Abstract: A dual polarized, dual beam phased array antenna with wide scan angle capabilities was designed and manufactured for repeater application. Two fixed feeding networks were designed and manufactured for the dual beams. With the feed networks, one broadside main beam and one tilted main beam were generated. The antenna achieved good isolation of -35.5dB between the two main beams (broadside and tilted beam) within the frequency band of interest (2.5-2.7 GHz). The return loss for the broadside feed network port was found to be below –7.5dB and return loss for the tilted beam feed network port was less than -15dB. The side lobe levels for the broadside beam and the tilted beam were -12dB and -8dB down, respectively. The cross-polar level for the broadside beam was lower than -20dB at the center frequency both in the elevation plane and in the azimuth plane. Also, for the tilted beam the cross-polar level in elevation plane was measured to be as low as -20dB at the centre frequency

***Index Terms* — Repeater, dual beam antenna, scan angle, port isolation.**

I. INTRODUCTION

Today's cellular communication systems are primarily designed to provide cost effective wide-area coverage for users with moderate bandwidth demands (voice and low data rate). The very high data rates envisioned for next generation wireless systems in reasonably large areas do not appear to be feasible with conventional cellular architecture. The reasons being the proposed spectrum higher than the 2GHz band and more vulnerable radio propagation in higher bands in non-line-of-sight conditions. Another restriction is the lack of bandwidth needed to provide Mbps transmission over the cellular networks. The brute force solution to these concerns is to increase the density of base stations and the operators need to build a dense ''forest'' of base stations to provide wireless broadband over the coverage area. But the increase in base station numbers will result in considerably higher deployment and operation costs, and would be feasible if the number of subscribers is also increased. Also costs for additional infrastructure, and hence cost for additional services to be delivered over the cellular networks cannot be radically higher in comparison to the existing systems such as Wireless LAN at an airport or hotel, or simply a wired high speed connection at office or at home. It is obvious from the above discussion that some fundamental enhancement or upgrading is required for the ambitious data throughput and coverage requirements of future systems, and the integration of multi-hop capability by means of using fixed repeaters or relays into the conventional wireless networks may be one of the cost-effective ways of upgrading existing networks [1-3]. Furthermore, significant researches are performed on future relay assisted MIMO systems and repeater based cellular distributed antenna networks [4-6]. Thus, repeaters or relays may play a significant role in future generation of wireless systems.

In this paper, a dual beam phased array antenna is considered for repeater or relay application. The proposed repeater antenna has two fixed main beams with orthogonal polarization. One of the main beams is a broadside beam directed towards the base station and the other beam is scanned about 40^{o}-50^{o} from broadside and is directed towards the user premises. In order to achieve the required scan range of the second beam, a linear phased array with progressive phase shift concept is considered. The operating frequency for the antenna was selected within the band of 2.5-2.7 GHz. The proposed antenna concept for this application is shown in Figure 1.

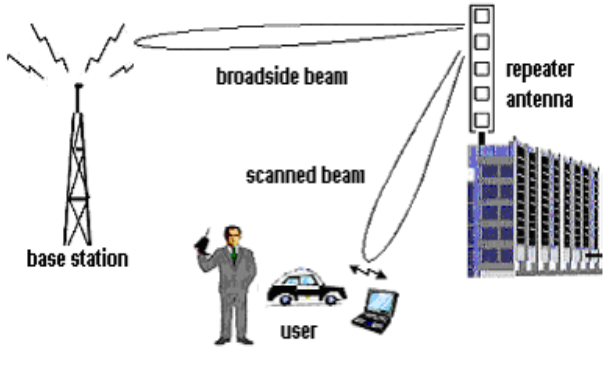


Fig. 1 Dual beam phased array concept for repeater application

II. DESCRIPTION OF THE ANTENNA

A small sized (58cm in length) low cost array antenna is considered for a proof of concept of repeater applications. The antenna was chosen to be a linear array with 8 elements to incorporate a simple feeding network with uniform power division. As the feed networks are composed of power dividers and T-junctions, it is easy to design feed networks having output port numbers equaling a power of 2. The element spacing to avoid grating lobe was chosen to be 0.5 wavelengths.

An improved microstrip inverted stacked patch geometry as presented in [7] was chosen as the single element for this array. After selecting the element type, element spacing and number of elements, the linear array with the feed networks was simulated in ADS. The objective of this simulation was to find the tilt angle for the scanned beam so as to obtain maximum isolation between the feeding ports for the two beams of the array antenna. While simulating the feed networks, ideal power dividers and phase shifters were assumed. The amount of phase shift added between the antenna elements was swept over a range of frequencies to produce different tilt angles, and the isolation between the two polarization ports was analyzed. The ADS simulation results for this case are shown in Figure 2, and the isolation curves track the scanned radiation pattern level at the broadside main direction as expected. The simulation results show that three scan angles, such as 15°, 30° and 48.5° from broadside give highest isolation between the two feed ports. Among these three tilt angles; 48.5° tilt angle seemed most appropriate for the repeater application and thus was chosen for the scanned beam of the proposed antenna.

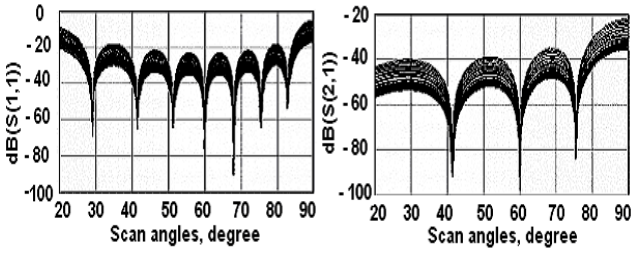


Fig. 2 Simulation results for return loss and isolation at beam ports as a function of scan angle from array axis with frequency as a parameter

The complete linear array was built on (58×10) cm^2 substrates both for the lower and the upper patches. TLC-30 substrate with thickness of 1.5mm was used for the lower patch array, and for the upper patch array FR4 substrate with a thickness of 0.8mm was chosen. The centre to centre distance between the elements was set to be 5.75cm (0.5λ). The array of upper patches was placed 12.1mm above the array of the lower patches with the help of plastic spacers. The manufactured complete array antenna is shown in Figure 3, and only eight out of ten elements were fed from the feed networks with the remaining edge elements terminated in matched loads.

III. DESIGN OF TWO FEED NETWORKS

Two feeding networks were designed, one for the broadside beam and the other one for the 48.5° tilted beam. Simple T-junctions were used as power dividers while transmission lines were used as fixed phase delays. For the broadside beam, the phase difference between adjacent antenna elements is zero degree, and for the 48.5° tilted beam the phase progration between the antenna elements was calculated as [8]: $\beta = kd\cos\theta$

where $k = \frac{2\pi}{\lambda}$, $d = \frac{\lambda}{2}$ and $\theta = \left(90^0 - 48.5^0\right)$.

The phase shift as calculated by the above expression was found to be -135° (or 225°). Thus, the feeding network was designed to have -135° phase shift between adjacent antenna elements. The two manufactured feed networks and the complete array antenna are shown in Figure3.

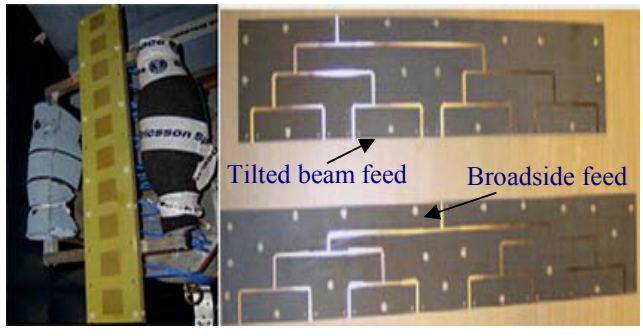


Fig. 3 Complete array antenna at test range and the two feed networks

IV. S- PARAMETER RESULTS

The tilted beam feed network was connected to the horizontal polarization and the broadside feed network was connected to the vertical polarization port of the antenna during measurements. Figure 5(a) and Figure 5(b) present the measured phase progression for the feed networks separately. It was found that the phase progressions for the tilted beam feed network was close to the requirement of -135° and the broadside network was close the required 0° phase progression. Figure 6 presents the measured mutual coupling between antenna elements and it is seen that the mutual coupling between two adjacent elements is less than -11dB, and the mutual coupling between the second adjacent elements is less than –20.5dB within the band of interest. The isolation between the orthogonal ports of two adjacent antenna elements was higher than 25dB. Figure 7(a) shows a block diagram of the complete antenna arrangement, and Figure 7(b) shows measured return loss of the complete array and isolation between the two beam ports.

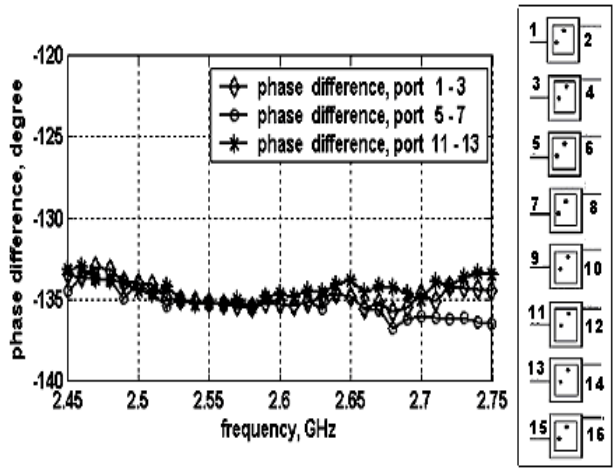


Fig. 5(a) Measured phase progression between adjacent antenna elements of the tilted beam feed network, the results of the remaining ports being similar and are not shown

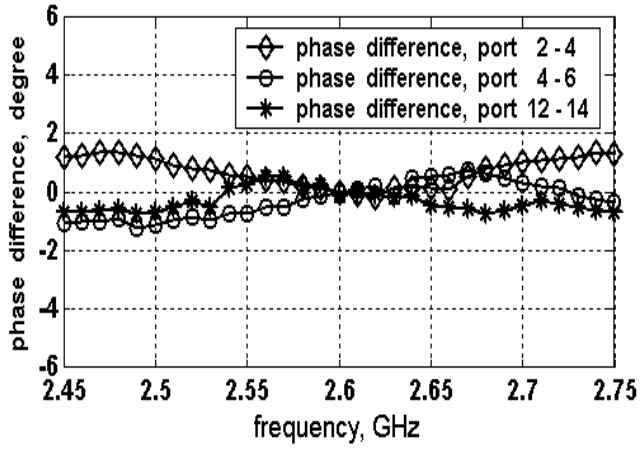


Fig. 5(b) Measured phase progression between adjacent antenna elements of the broadside beam feed network, the results of the remaining ports being similar and are not shown

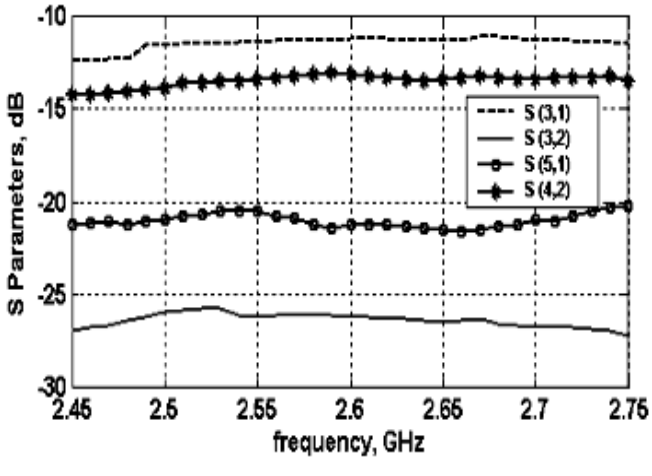


Fig. 6 Measured mutual coupling between antenna elements in the array

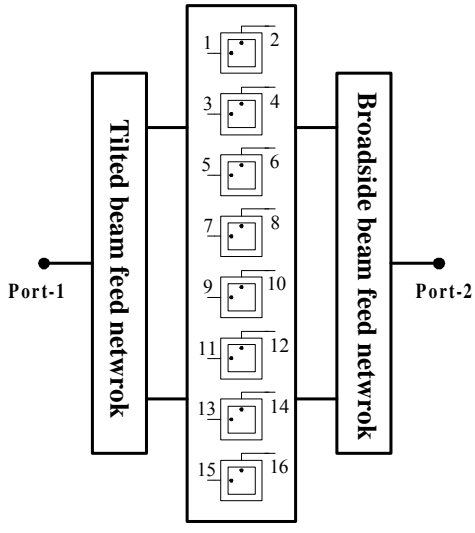


Fig. 7(a) Antenna and feed network arrangement during return loss and beam port isolation measurements

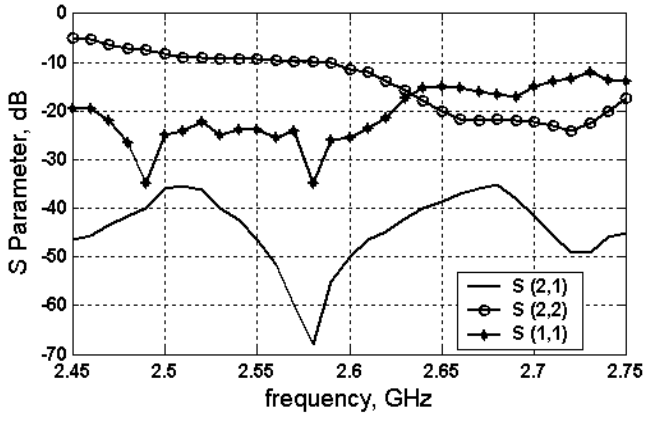


Fig. 7(b) Measured return loss and port isolation for the complete array; feed 1 connected to the horizontally polarized beam and feed 2 to the vertically polarized beam

V. RADIATION PATTERN MEASUREMENTS

Both elevation and azimuth plane radiation patterns were measured for broadside beam while only elevation plane pattern was measured for the tilted beam. Due to the fixture and turntable arrangements, the azimuth pattern of the tilted beam was not possible to measure through the main beam peak. However, the azimuth pattern would be very similar for the two beams. Both co- and cross-polar levels for the two beams are plotted in Figure 8. As observed, the side lobe levels for the broadside beam in the elevation plane is about -12dB down as expected for a uniform amplitude distribution and for the tilted beam, it is -8dB down. For the tilted beam case, the reason for the relatively higher level of side lobes is due to the element factor. The element pattern at a -48.5^{o} tilt angle is almost -3.5dB lower

compared to the broadside direction. In the tilted beam case, the high side lobe level at +48° is roughly the mirror of the main beam (fig. 8(c)). This is due to reflections at the antenna elements and limited isolation at the T-junctions in the feed network.

The cross-polar level for the broadside beam is lower than -20dB at the center frequency of 2.6GHz both in the elevation and the azimuth plane. However, the cross-polar level increases to -13dB at the band edges. The main reasons of high cross-polar level are random probe feed positioning errors in the array as well as small element distance (0.5λ), and high mutual coupling between adjacent array elements. For the tilted beam, the cross-polar level in elevation plane is found to be lower than -18dB at all frequencies within the band of interest.

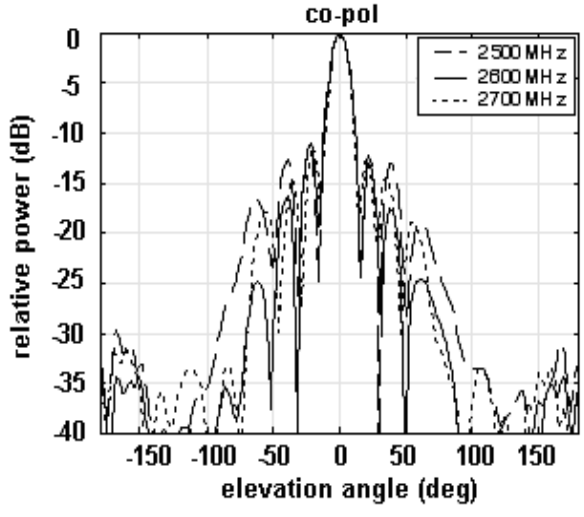


Fig.8(a) Measured elevation plane co-polar radiation pattern for broadside beam

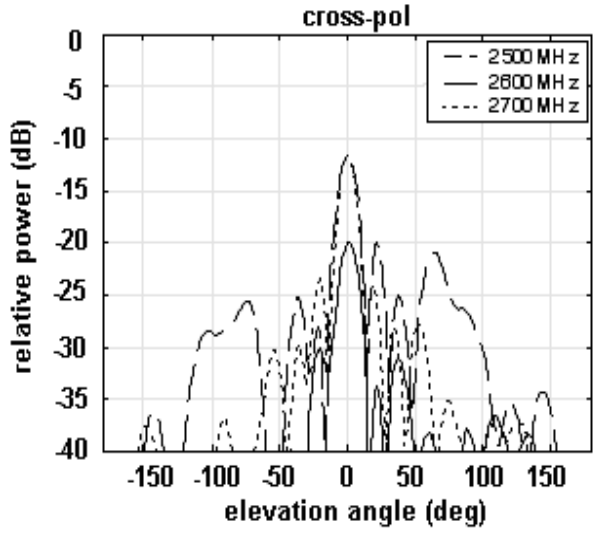


Fig. 8(b) Measured elevation plane cross-polar radiation pattern for broadside beam

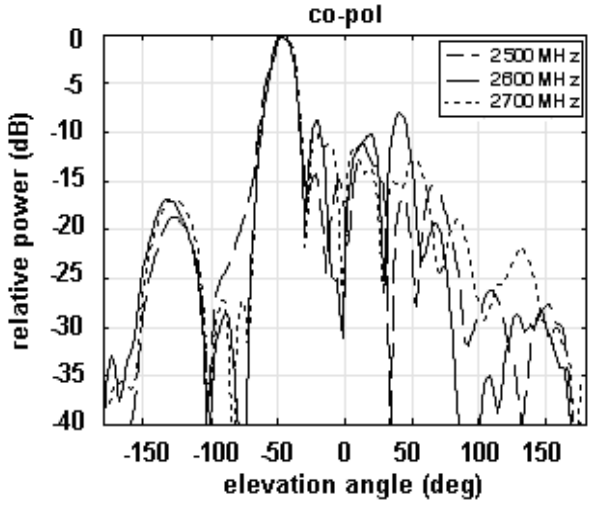


Fig. 8(c) Measured elevation plane co-polar radiation pattern for tilted beam

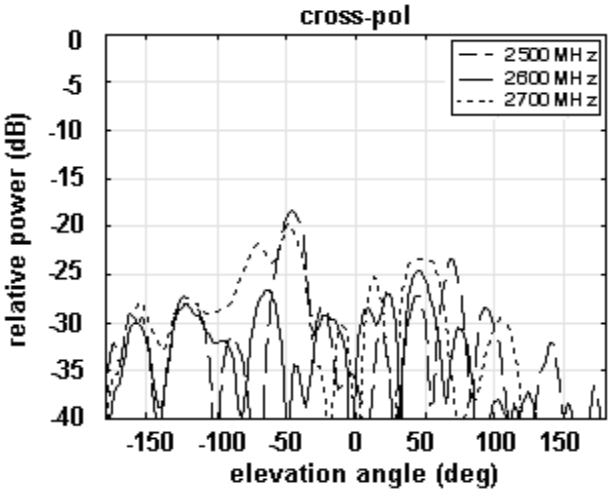


Fig. 8(d) Measured elevation plane cross-polar radiation pattern for tilted beam

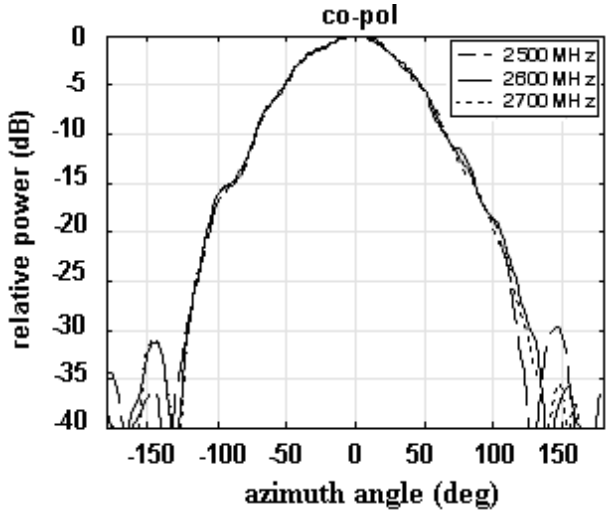


Fig. 8(e) Measured azimuth plane co-polar radiation pattern for broadside beam

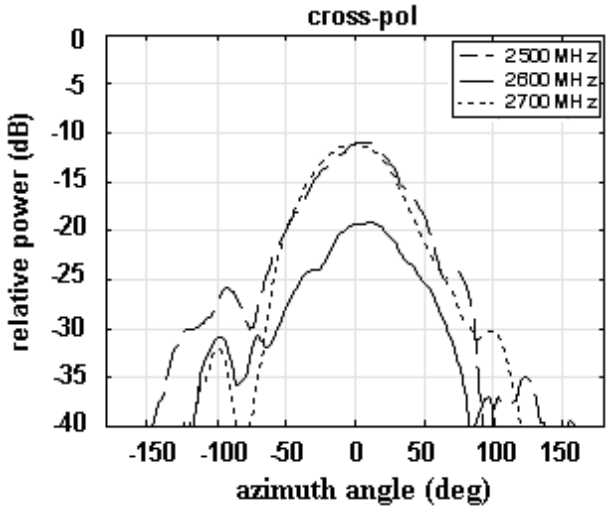


Fig.8(f) Measured azimuth plane cross-polar radiation pattern for broadside beam

VI. MEASURED DIRECTIVITY & GAIN

The directivity for both beams was calculated to be 14.5-15.2 dBi by integrating the patterns within the band of interest. The directivity of a linear array is rather constant as a function of the scan angle. The realized gain of the antenna was 11.8-13.1 dBi and 12.7-13.5 dBi for the broadside and tilted beam cases, respectively. There is a loss of about 1.5dB, which includes losses in feed network, connecting cables, reflection loss, signal power loss in cross-polarization, and misalignment during the measurements. The reduced gain for the broadside beam at the low frequency is mainly due to high reflection losses (fig. 7(b)). A return loss of -8dB gives a reflection loss of 0.75dB, which explains the lower realized gain at 2.5 GHz in the case of the broadside beam. The measured results are summarized in Table 1.

Table 1 Measured directivity, gain and 3-dB beam width

f (GHz)	Beam width (deg) (Broadside, elevation)	Beam width (deg) (Broadside, azimuth)	Beam width (deg) (Tilted-beam, elevation)	Gain (dBi) (Broadside)	Gain (dBi) (Tilted-beam)	Directivity (dBi) (Broadside)	Directivity (dBi) (Tilted beam)
2.5	13.8	78.2	17.9	11.8	13.5	14.5	15.2
2.6	13.5	78.0	16.7	13.1	13.4	15.1	14.9
2.7	12.3	77.6	17.6	13.0	12.7	15.0	14.8

VII. CONCLUSION

Design of a dual beam phased array antenna with wide scan angle has been demonstrated. The manufactured antenna performances agree with the theoretical and simulated results. The manufactured antenna with added power amplifiers can easily be used for future investigations and field measurements in repeater applications.

ACKNOWLEDGEMENT

The authors wish to thank Saab Space, Sweden, for the use of the indoor antenna test range.

REFERENCES

[1] M. O. Hasna, and M. S. Alouini.: 'A Performance Study of Dual-Hop Transmissions with Fixed Gain Relays', IEEE Trans. on Wireless Communications, Vol. 3, No. 6, November 2004, pp. 1963–1968.

[2] M. Rahman, and P. Ernström.: 'Repeaters for Hot-Spot Capacity in DS-CDMA Networks', IEEE Trans. on Vehicular Technology, Vol. 53, No. 3, May 2004, pp. 626–633.

[3] J. N. Laneman, and G. W. Ornell.: 'Energy-Efficient Antenna Sharing and Relaying for Wireless Networks', IEEE Proc. on Wireless Communication Networking Conf., Chicago, September 2000, pp. 7-12.

[4] B. Rankov, and A. Wittneben.: 'On the Capacity of Relay-Assisted Wireless MIMO Channel Channels', IEEE Proc. on Workshop on Signal Processing Advances in Wireless Communications, Lisbon, Portugal, July 2004, pp. 323-327.

[5] K. J. Kerpez.: 'A Radio Access System with Distributed Antennas', IEEE Trans. on Vehicular Technology, Vol. 45, No. 2, May 1996, pp. 265-275.

[6] A. Obaid, and H. Yanikomeroglu.: 'Reverse-link Power Control in CDMA Distributed Antenna System', IEEE Proc. on Wireless Communications and Networking Conf., Chicago, Sept. 2000, pp. 608–612.

[7] A. Uz Zaman, L. Manholm,and A. Derneryd.: 'Dual polarized Microstrip Patch Antenna With High Port Isolation', Electronics Letters, Vol. 43, No. 10, May 2007, pp 551-552.

[8] C. A. Balanis.: 'Antenna Theory Analysis and Design', Second Edition, John Wiley & Sons, USA, 1997, ISBN 0-471 59268-4.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Circularly Polarized Compact Passive RFID Tag Antenna

Hidayath Mirza, Mohd Imran Ahmed[†] and Mohammad Fazleh Elahi[††]*
* Kyung Hee University, Suwon, South Korea, e-mail:hidayathmirza@gmail.com
[†]University of Applied sciences in Lübeck , Lübeck,Germany
[† †]IT Division, Telecom Malaysia (Bangladesh) Ltd,e-mail:fazleh_elahi_71@yahoo.com

Abstract—This paper presents a compact (35x36.5 mm) RFID tag antenna which has large RCS patterns and easy conjugate impedance matching property by use of an inductively-coupled feeding. Its simulated maximum (match state) and minimum (short state) RCS are -15.36dBm^2 and -22.0 dBm^2, with a difference of 6.69 dB between the two states. We have also calculated the detection distance of the tag for different values of reader antenna gain.

Keywords—Radar Cross Section (RCS), Inductively-Coupled and RFID chip, RFID chip, Chip sensitivity.

I. Introduction

Radio Frequency identification (RFID) has gained much interest because of its numerous applications especially in tantalizing benefits for supply chain management, inventory control, and etc. [1]. A passive back-scattered RFID system operates in the following way: An interrogator or a reader transmits a modulated signal with periods of unmodulated carrier, which is received by the tag antenna. The Radio Frequency (RF) voltage developed at the antenna is converted in direct current (D.C.). This voltage provides power to the application specific integrated circuit (ASIC) chip [2].

The impedance of the antenna varies between two states i.e. short and match. And the difference between short RCS and match RCS of the tag has a significant impact in backscattering the information to the reader as shown in Fig.1. In general, tags are dipole type, so the problem we face is nulls because it is not isotropic. If the tag lies in the null direction with respect to the reader, sometimes it is not possible for tag to get enough signals and it may cease functioning and stop responding to reader. We usually use linear polarization, to make the antenna size small and it is known that tag and reader antennas are fixed. The best example is Toll Plaza on Expressway. Mismatch problems arise due to random and arbitrary postures of the tag. In this situation, we have to use circular-polarization (CP). This type of polarization is used in departmental stores such as Wall-mart because, we are not sure of random posture tag antenna at the time of tagging [3]. Even though we use CP reader, it cannot completely avoid the null problems of dipole-type tag antennas [4, 5].

Size and impedance matching are some of the prominent factors that need to be considered carefully while designing a tag antenna. Size of the tag is a prevalent problem since; the size of ASIC is in millimeters, the world is growing smaller due to the advent of nanotechnology. However, it is difficult to reduce the size because the size of the antenna is a specific fraction of the wavelength [6]. In fact, if we reduce the size, then the gain of the tag will decrease and this will directly decrease the RCS and hence the read range will also affected. A tag should posses the following characteristics: large read range, inexpensive and as small as possible so that it can be attached to any product. Many parameters of the antenna, such as the amount of power absorbed by the chip, power transmission co-efficient and read-range are directly determined by the degree of impedance mismatch between chip and antenna [7].

In this paper, we present a compact inductively-coupled tag antenna with large radar cross section (RCS) patterns, which can overcome the problem of nulls (in two planes). By simply adjusting the position of the feeding element, desired impedance for any chip can be achieved. Rest of the paper is organized as follows: Section 2 discusses the theory and design aspects of the proposed design. Section 3 contains simulation results. Finally, section 4 concludes the paper and highlights some future work.

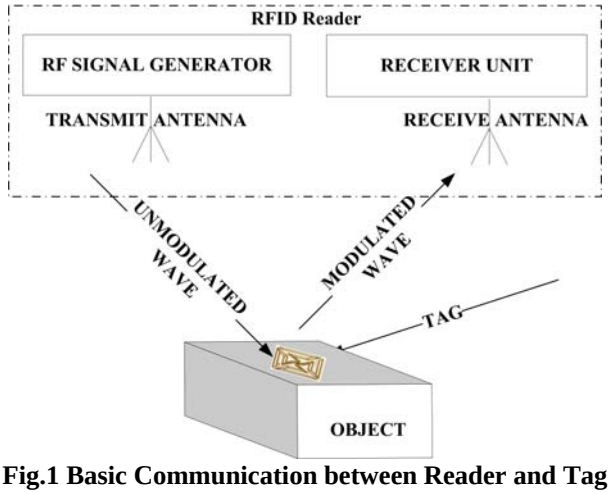


Fig.1 Basic Communication between Reader and Tag Antenna

II. Theory & Design

The tag antenna discussed in this paper is meandered line antenna (MLA) with inductive coupling. It is easily fabricated with a copper layer of 0.018 mm and PET substrate of thickness 0.018 mm with relative permittivity of 3.2. The designed antenna consists of a feeding loop (because chip is attached here) and a radiating body. The impedance of the antenna depends upon the distance between the feeding loop and the radiating body. It also depends on the size of the feeding loop.

The salient features of this design are as follows:

1. Compact 35x36.5mm,
2. Near-isotropic RCS patterns only in two planes,
3. Large RCS, and
4. Large difference between short and match state 6.69 dB.

The overall size of the proposed tag antenna is relatively small at 911MHz when compared with that of most commercial RFID tags. It is well known that maximum power transfer will occur when the source impedance is equal to the conjugate of the load impedance [7]. So, we need a chip which is capacitive coupled as antenna is inductive.

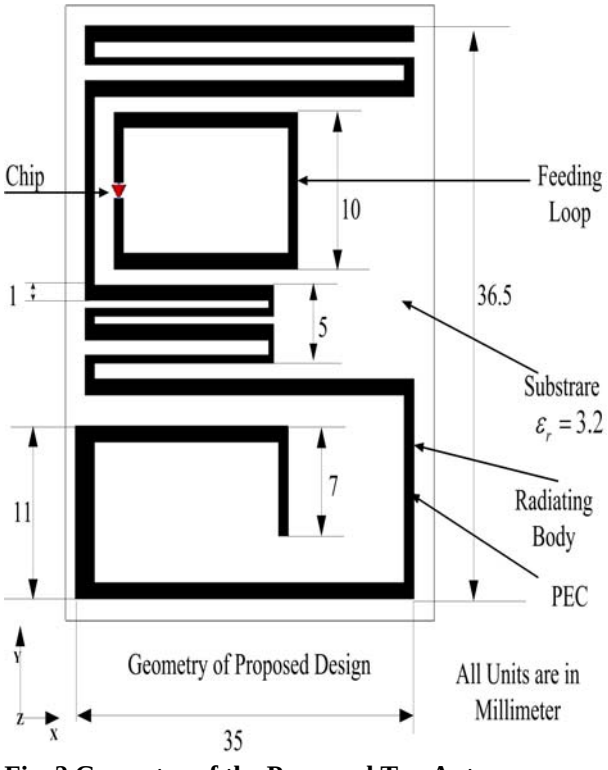


Fig. 2 Geometry of the Proposed Tag Antenna

III. Simulation

Based on a high quality factor "*Q*", typical chip impedance is 1.258-j194.2Ω. The designed tag antenna impedance is Z_a=1.258+j194.2 Ω at 911MHz for the frequency range of 0.5-1.5GHz.

The quality factor "*Q*" for a combined system of tag and chip is roughly the ratio of chip reactance and resistance $Q = \frac{X}{R}$ (in our case 154.37). The Q factor of the proposed tag antenna is thus about 154.37. Most of the commercial chips being developed these days have high *Q* factor to lower the required threshold of open-circuit voltage of the tag. If more bandwidth is required, then the *Q* factor must be lowered by reducing the ratio of chip reactance and resistance at an initial design stage. A trade-off is usually necessary between the required *Q* factor for the chip and RCS bandwidth [8].

A. Radar Cross Section (RCS)

The power received by the reader antenna from the tag antenna is directly proportional to gain of the reader antenna, power density and RCS of the tag antenna [9]. Moreover, RCS is totally depended on tag's gain as seen in equation (1).

$$\sigma_{tag} \approx \frac{\lambda^2 R_a G_{tag}^2}{\pi \left| Z_a + Z_c \right|} \tag{1}$$

Where λ is the wavelength, R_a is the real part of tag antenna impedance Z_a, Z_c is the chip impedance and G_{tag} is the gain of the tag antenna.

In simulation, we used circularly polarized wave (CP). CP wave is generated to hit the tag in –z direction. After hitting the tag, it scatters in all directions. As shown in Fig 3, with this RCS is calculated.

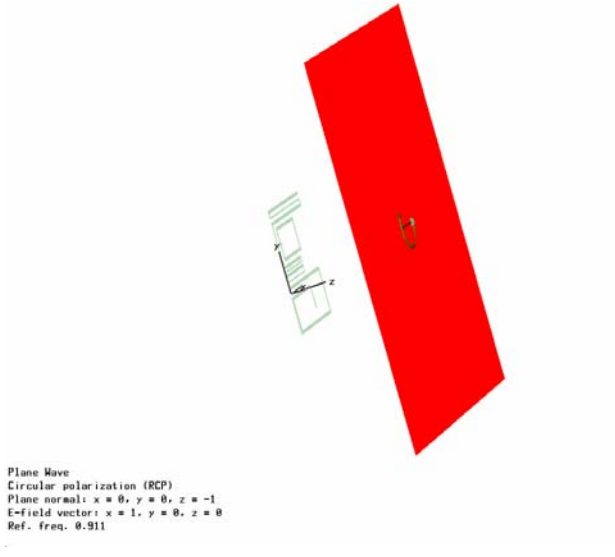


Fig.3 CP Wave Incident on Tag Antenna.

Simulated RCS are obtained for every 30 degrees for two cases of the chip loads: Z_c=-j194.2 Ω (short) and *Zc*=1.258-j194.2 Ω (match), and plotted as a function of incident angles. We usually check RCS in two different states a) short and b) matched state. These two distinct RCS values are received by the reader antenna [9].

$$\Delta\sigma = \left| \sigma_{short} - \sigma_{match} \right| \tag{2}$$

Where σ_{short} the RCS is in short case and σ_{match} in matched case (short case RCS is usually bigger than matched).

Larger the difference between short and match state; better the tag antenna, these values are used as modulation states; assumed to be 1 and other 0 in case of amplitude shift keying (ASK). Tag antenna can communicate with reader antenna in ones and zeros.

For short case, Z_L=-j 194.2 Ω (short) at 911MHz, a equivalent chip capacitor of 0.9 pF was mounted as shown in Fig. 2. The simulated RCS in x-y and x-z planes are shown in Fig. 4(a) and (b) respectively. It clearly shows that both the patterns are near-isotropic in x-y and x-z planes.

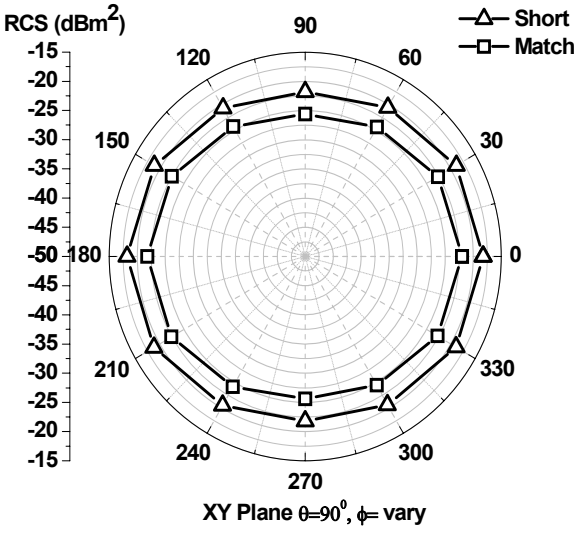


Fig 4(a)

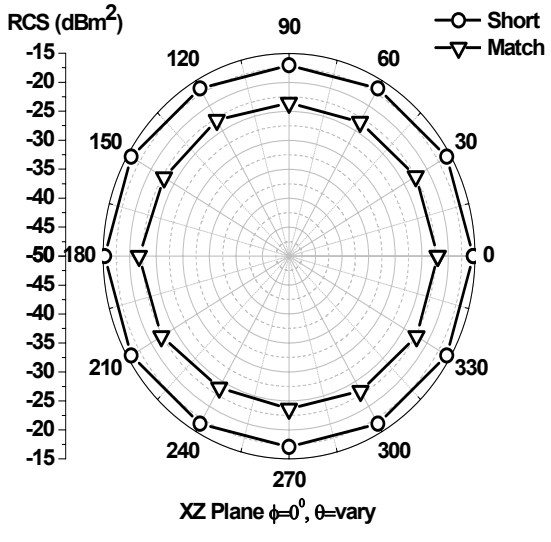


Fig.4 (b)

B. Detection Distance:

There are two ways for calculating detection distance:

1. Detection distance based on reader sensitivity,
2. Detection distance based on chip sensitivity.

It is found that detection distance based on chip sensitivity is very smaller as compared with reader sensitivity. The reader sensitivity is usually constant whereas the chip sensitivity depends on a manufacturer. As the chip sensitivity depends on company, so we have calculated detection distance based on the reader sensitivity. If the chip is more sensitive, then it means it requires very less power level to activate. Hence, the cost of the chip depends upon the degree of sensitivity. The typical chip sensitivity ranges from -8 to -20 dBm.

In Fig. 5, we have plotted the detection distance limited by reader sensitivity based on values shown in Table 1 for the tag antenna by using equation as shown below:

$$R_{\max} = \frac{\lambda}{4\pi}\sqrt{\frac{P_{in}\ p\ G_{reader}G_{tag,normalised}}{P_{l,\min}}} \qquad (3)$$

where, $R_{\max}$ is the maximum detection distance, λ is the wavelength, P_{in} is the input power, p is the polarization factor, G_{reader} is the gain of a reader antenna , $G_{tag,normalised}$ is the normalized gain of the tag antenna and $P_{l,\min}$ is the sensitivity of the chip.

Normalized gain can be defined as shown in the equation below.

$$\text{Normalized Gain} = \frac{\text{Gain of the tag antenna}}{\text{Gain of the half wave dipole (1.64)}} \qquad (4)$$

Table 1 Values assigned for detection distance

Parameter	Assigned Value
λ	0.329 meter
P_{in}	1W
p	1
G_{reader}	4W
$G_{tag,normalised}$	0 to 1
$P_{l,\min}$	can chose any value between -8dBm to -15 dBm

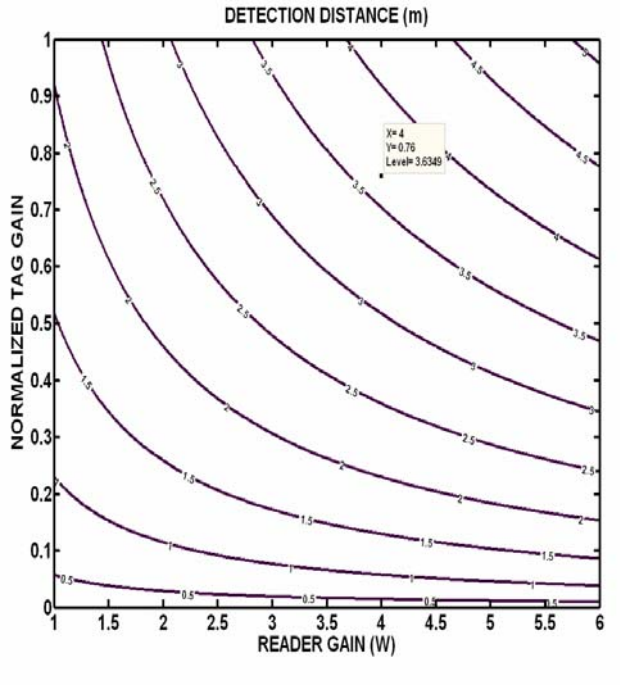


Fig. 5 Detection Distance Contours Based on Normalized Tag Gain and Reader Sensitivity

Different countries have different reader antenna gain which is assigned by Federal Communications Commission. The advantage of the above graph is that if we know the normalized gain of the given tag and reader's gain (country specific), then we can easily point out the detection distance.

For the proposed tag, normalized gain is 0.76 and detection distance is 3.6349 meters, when the reader gain is 4 Watts.

IV. Conclusion

A compact (0.12x0.14λ) RFID tag with inductively-coupled feed and near-isotropic RCS patterns has been designed and simulated. Future work on this RFID tag design includes further optimization of the overall RCS patterns, such that it becomes isotropic and to achieve commercial chip impedances. It can be done by optimizing distance between feeding loop and radiating body. The work presented in this paper will be a useful contribution towards the RFID industry in terms of small size with large RCS.

References

[1] Finkenzeller, K, *RFID Handbook,* 2nd ed., John Wiley & Sons, England, 2003.

[2] Rao, K. V. S., and Nikitin, P. V, "Antenna design for UHF RFID tags: a review and a practical application", *IEEE Trans. Antennas and Propag,* 53, (12), pp. 3870-3876, 2005.

[3] Yuri Tikhov, "Antenna design for UHF RFID tag: A Review and a Practical Application", *IEEE Trans. Antennas and Propag,* Vol. 54.No.6, June 2006.

[4] Penttila, K., Keskilammi, M., Sydanheimo, I., and Kivikoski, M, "Radar cross-section analysis for passive RFID systems", IEE Proc. Microw. Antennas Propag, 153, (1), pp. 103-109, 2006.

[5] Kim, N., Kwon, H., Lee, J., Lee, B, "Performance analysis of RFID tag antenna at UHF (911MHz) band", IEEE Antennas and Propag. Symposium, pp. 3275-3278, 2006.

[6] Hong-kyung Ryu and Jong-Myung Woo, "Small-Sized square loop antenna using Meander Line for RFID tag applications", Antenna and Propagation Society, 2007

[7] K. V. S. Rao, Pavel V. Nikitin, Sander F. Lam, "Impedance Matching Concepts in RFID Transponder Design", Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05), pp. 39-42, 2005.

[8] Jong-Wook Lee; Hongil Kwon; Bomson Lee, "Design Consideration of UHF RFID Tag for Increased Reading Range," Microwave Symposium Digest, 2006. IEEE MTT-S International , vol., no., pp.1588-1591, June 2006.

[9] C.-H. Loo, K. Elmahgoub, F. Yang, A. Z. Elsherbeni, D. Kajfez, A. A. Kishk, T. Elsherbeni, L. Ukkonen, L. Sydanheimo, M. Kivikoski, S. Merilampi, and P. Ruuskanen, "Chip impedance matching for uhf rfid tag antenna design," Progress In Electromagnetics Research, PIER 81, 359-370, 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A UHF-RFID Tag Antenna for Commercial Applications

Hidayath Mirza, Mohammad Fazleh Elahi***
*School of Electronics & Information, Kyung Hee University, South Korea
Email:hidayathmirza@gmail.com
**System Analyst, Telecom Malaysia (Bangladesh) Ltd.
Email:fazleh_elahi_71@yahoo.com

Abstract - An RFID tag consists of tag antenna and a microchip. The terminal impedance of the chip varies in two states. It also affects the RCS and read range of the tag antenna. This paper presents a circularly polarized compact UHF RFID tag antenna with impedance of commercially available chip. The tag antenna's RCS patterns and read range are also analyzed in different planes.

Key words: commercial chip, RCS, Detection distance.

I. Introduction

RFID systems are operated at widely differing frequencies, ranging from 135 kHz- 5.8 GHz. (long wave to microwave region). An RFID system consists of a *tag* and a *reader*. The tag is data carrying device, usually located or attached to the object, which has to be identified or interrogated. It consists of a tag antenna and a microchip; also known as application specific integrated circuit (ASIC) [1]. A reader is a data capturing device, which reads/writes by transmitting a combination of modulated and un-modulated carrier to the tag. RF voltage developed on antenna terminals is used for biasing the ASIC. In return, the tag back-scatters information is written in the chip. The performance of the tag is usually evaluated by the detection distance and the detection distance is usually determined by measurement as a tag is moved away from a reader antenna [2, 3].

A good tag antenna structure must be conjugate-matched to specific chip impedance with simple adjustment of its dimensions and its impedance bandwidth has to be broad enough. Due to size and cost factors, external matching networks should be avoided. Hence, complex conjugate impedance matching without extra circuits needs to be given preponderant importance [3-4]. Moreover, if the antenna impedance is not exactly complex conjugate to the chip impedance then losses will occur. In some situation, it is hard for the reader to detect the tag.

Size and impedance matching are among some of the prominent factors to be considered carefully while designing tag antenna. Size is prevalent problem since the size of ASIC is in millimeters, as the world is growing smaller due to the advent of nanotechnology. However, it is difficult to reduce the size because the size of the antenna is a specific fraction of the wavelength [5]. In fact, if we reduce the size, gain of the tag will decrease and this will directly decrease the RCS and hence the read range is also affected. Therefore, tag should posses following characteristics: size should be low profile, inexpensive and as small as possible so that it can be attached to any product. Many parameters of the antenna, such as the amount of power absorbed by the chip, power transmission co-efficient and read-range directly determined by the degree of impedance mismatch between chip and antenna [6].In this paper, a novel compact inductively-coupled tag antenna, which has large RCS patterns is proposed. It can be simply printed on a cheap PET substrate having relative permittivity of 3.2. In this paper, we propose a compact (36x29.9mm) meander line, inductively coupled RFID tag antenna. The proposed tag is compact and has relatively large radar cross section (RCS) patterns compared to its size. It has a typical impedance of 56.25+j156.1Ω at 911MHz, which is equal to *Alien* commercial chip impedance. We have also interpreted the detection distance with respect to tag's gain for different angles in all the three planes.

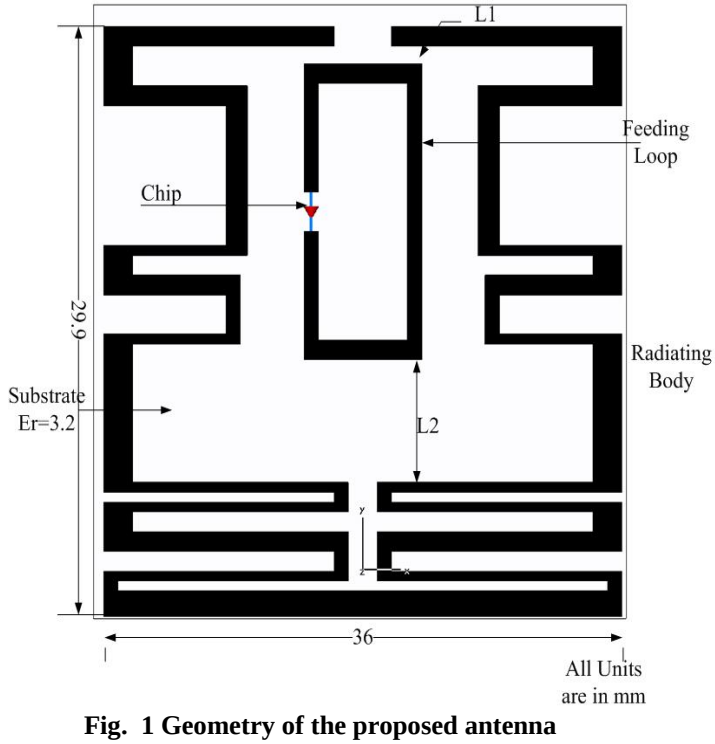


Fig. 1 Geometry of the proposed antenna

II. Tag Antenna Design

The tag antenna discussed in this paper is inductively coupled and fabricated with a copper layer of 0.018mm, PET substrate of thickness 0.018 mm with relative permittivity of 3.2. The salient features of this design are as follows: compact 36x30.6mm, near-isotropic directivity, simple in design, symmetric in shape and having a gain of 0.9769 dB, wide variety of tag chips can be used because of large variation in real and imaginary part within the band of interest (for Korea RFID center frequency is 0.911GHz). By adjusting the separations, L1 and L2 as shown in Fig. 1, the required antenna impedance for a specific chip can be achieved with ease

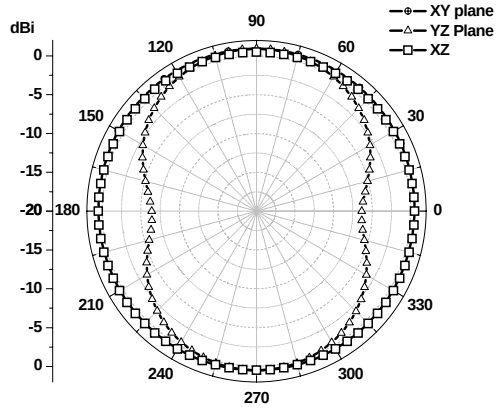


Fig. 2 Antenna's directivity patterns in xy, yz, and xz plane

Fig. 2 shows the simulated directivity patterns in all three planes. We can clearly observe that xy and xz planes are near-isotropic where as yz plane is not near- isotropic. This is because of flow of current in opposite direction. Hence it will cancel out.

III. Simulation

The impedance of designed tag antenna is 56.25+j156.1Ω at 911MHz. The locus of the antenna for the frequency range of 0.5 to 1.5 GHz is shown in Fig 3. It can be clearly observed that the locus circles around $Z_A = 56.25 + j156.1\Omega$ near 911 MHz. The range of chip resistance could be found from manufacturer's data sheet is 2 ~ 12Ω and that of a chip reactance is -150 to -100Ω [7].But these days new RFID chip values ranges from 2-55Ω and chip reactance -100 to 200Ω. In our case, we have connected commercially available Alien chip. Its impedance is 55-j155Ω. So, we have designed antenna impedance near about it with complex conjugate value.

Fig.4 Exploits, tag antenna's impedance (resistance and reactance) as a function of frequency. It can be found that, the structure can achieve resistance from zero to fifty five ohms and reactance from five to two hundred and ten ohms. This means a wide range of real and imaginary part only for desired frequency. Therefore, without changing the structure of the antenna, commercial chip impedance value or desired value of chip impedance for specific application can be achieved by simply varying the distance between feeding loop and radiation body.

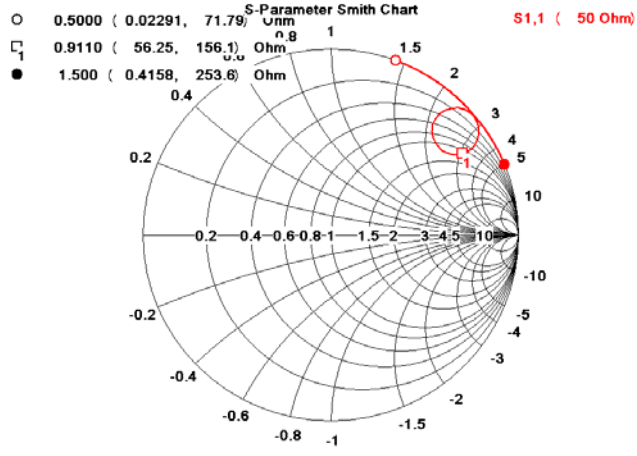


Fig. 3 Loci of the tag impedance

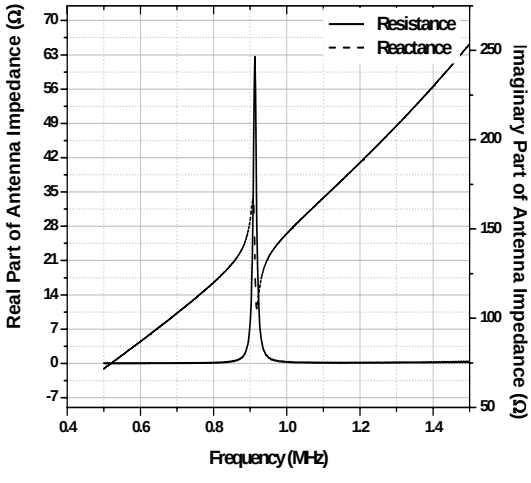


Fig. 4 Antenna impedance

IV. RCS

In simulation, we have created a tag and reader atmosphere. When a tag is hit by a plane wave from the reader, how it will respond. We used circularly polarized wave (CP). CP wave is generated to hit the tag. After it hits the tag, it scatters in all directions. The circularly polarized reader antenna is used to overcome the problem of tag orientation. It distributes UHF energy in a radially symmetrically patterns. However, for this we have to sacrifice 3dB power loss due to the mismatch in polarization [8]. We have calculated RCS at every thirty degree for short and match case. Fig. 6 (a), 6(b) depicts CP RCS in xy plane, yz plane, respectively. We found that xy and xz plane RCS patterns are near-isotropic.

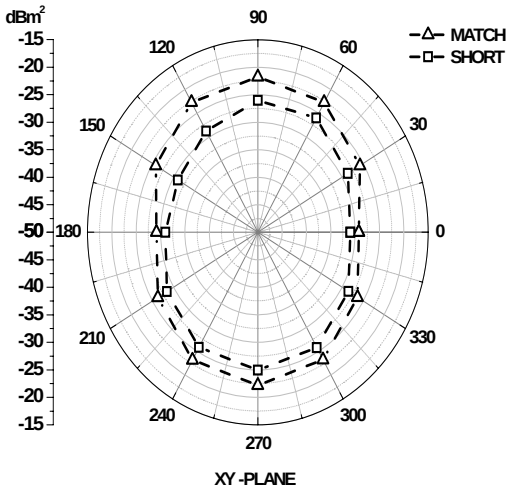


Fig. 6(a) RCS patterns at 911MHz in XY plane

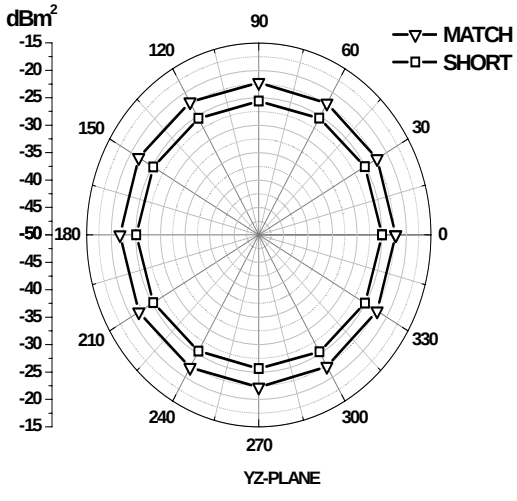


Fig. 6(b) RCS patterns at 911MHz YZ planes

V. Detection Distance

There are two ways for calculating detection distance:

1. Detection distance based on reader sensitivity.
2. Detection distance based on chip sensitivity.

It's known that detection distance based on chip sensitivity is very smaller compared with reader sensitivity. Reader sensitivity is usually constant where as chip sensitivity depends on manufacturer. If the chip is more sensitive means it requires very less power level to activate. Hence, the cost of the chip depends upon the degree of sensitivity. Typical chip sensitivity ranges from -8 to -20 dBm.

In Fig. 7, 8 and 9 we have plotted the detection distance based on tag's gain at different angles in yz, xy and xz planes, respectively for chip sensitivity of -10 dBm and -15 dBm and other parameter values are listed in Table 1.

The detection distance is given by:

$$R_{\max} = \frac{\lambda}{4\pi}\sqrt{\frac{\left(1-\left|\Gamma_{tag}\right|^2\right)P_{in}\ p\ G_{reader}G_{tag}}{P_{l,\min}}} \quad (1)$$

where, $\left(1-\left|\Gamma_{tag}\right|^2\right)$ is the matching factor antenna and chip, $R_{\max}$ is the maximum detection distance, λ is the wavelength, P_{in} is the input power, p is the polarization factor, G_{reader} is the gain of a reader antenna , G_{tag} is the gain of the tag antenna and $P_{l,\min}$ is the sensitivity of the chip, which is usually represented in dBm.

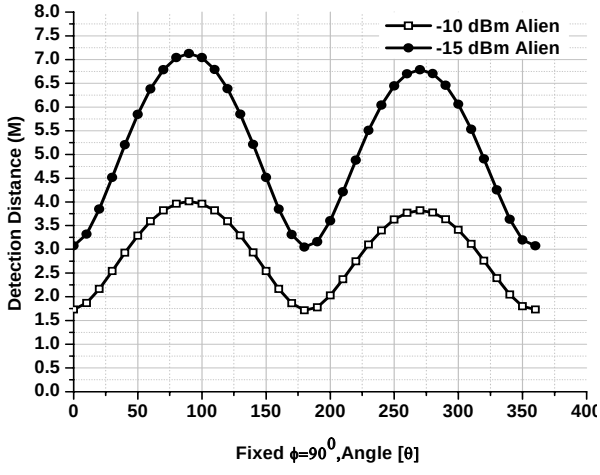


Fig 7 Detection distance of YZ plane

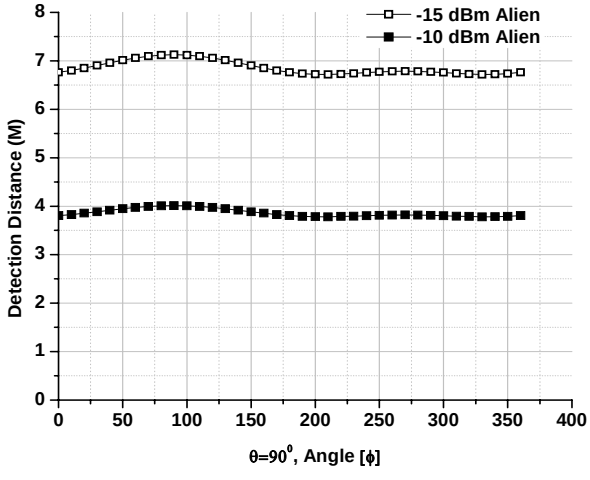


Fig 8 Detection Distance of XY plane

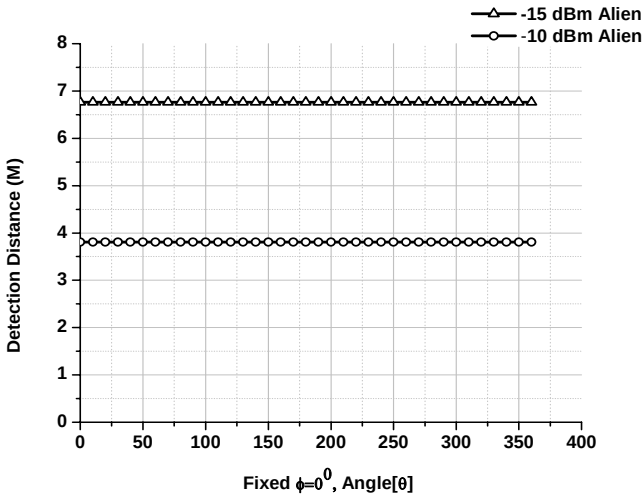


Fig 9 Detection distance of XZ plane

Table 1 Values assigned for detection distance

Parameter	Assigned Value
λ	0.329 meter
P_{in}	1W
p	0.5
G_{reader}	4W
G_{tag}	Specific for given angle
$P_{l,\min}$	-10dBm & -15 dBm

VI. Conclusion

This paper has presented a small passive UHF-RFID tag with simple, symmetric and low profile design which is suitable for tagging objects. The impedance of the tag antenna matches with commercially available RFID chips. By changing the L1 and L2 (antenna dimensions) desired chip impedance can be achieved with ease. The antenna has been simulated using CST Microwave studio™ 5.1

References

[1] K.Finkenzeller, RFID Handbook: Radio-Frequency Identification Fundamentals and Applications, 2nd ed: Wiley, 2004.

[2] H. Kwon, B. Lee, 'Evaluation of RFID tag antenna performance using radar cross sections', Proc. EUMC, pp. 491-493, 2005.

[3] K.V.S. Rao, P.V. Nikitin, and S.F. Lam, "Antenna design for UHF RFID Tags: A review and a practical application", IEEE Trans. Antennas Propa., Vol53, No.12, pp. 3870-3876, Dec 2005.

[4] H. Kwon, B. Lee, "Meander line RFID tag at UHF band evaluated with radar cross sections", Volume: 4, page(s): 4 pp., APMC, 2005.

[5] Hong-kyung Ryu and Jong-Myung Woo "Small-Sized square loop antenna using Meander Line for RFID tag applications", Antenna and Propagation Society, 2007

[6] K. V. S. Rao, Pavel V. Nikitin, Sander F. Lam, "Impedance Matching Concepts in RFID Transponder Design," autoid, Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05), 2005, pp. 39-42

[7] Hyungmin Jang, Bomson Lee "UHF- Band Inductively-Coupled RFID Antenna with Near- Isotropic Radar Cross Section Patterns", pages 1209- 1212 IEEE AP-S 2007.

[8] Yuri Tikhov, Comments and Replies on "Antenna design for UHF RFID Tags: A Review and a Practical Application", IEEE Trans. Antennas Prop., Vol. 54, No.6, pp. 3870-3876, June 2006

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Calibrated_Time-Frequency_MUSIC Method for Direction of Arrival Measurement of Pilot Signals

A. K. M. Baki[a!], K. Hashimoto[!!], N. Shinohara[!!], T. Mitani[!!], M. Matsumoto[!!!] And H. Matsumoto[!!]

!) School of Engineering & Computer Science (SECS), Independent University, Bangladesh
House # 10, Road # 10, Baridhara, Dhaka-1212, Bangladesh

!!) Research Institute for Sustainable Humanosphere (RISH), Kyoto University, Uji, Kyoto 611-0011, Japan
!!!) Furuno Co. Ltd., Japan
a) E-mail: baki@secs.iub.edu.bd

Abstract- The accuracy of the microwave (MW) power beam steering is one of the most critical goals for Solar Power Satellite (SPS). Retrodirective beam control system is the proposed method of beam steering technique for SPS. Detection of precise Direction of Arrival (DOA) of the pilot signal in SPS system is an integral part of the retro directive beam steering. High precision DOA estimation methods like MUltiple SIgnal Classification (MUSIC) and Time-Frequency_MUSIC (TF_MUSIC) based methods require calibration of antenna gain and phase errors prior to the DOA estimation. TF_MUSIC method of DOA detection can increase the effective SNR but its performance also degrades if antenna gain and phase errors exist. The performance of conventional MUSIC was tested by an experiment. Antenna gain and phase errors were calibrated by using the MUSIC method and at the same time the DOA was estimated by calibrating the gain and phase errors and the method is named as Calibrated_MUSIC (C_MUSIC). C_MUSIC performed better than MUSIC. The performance of TF_MUSIC was also tested by the same antenna calibration data and it was found that this Calibrated_TF_MUSIC (CTF_MUSIC) performed better than the MUSIC and C_MUSIC.

I. Introduction

The preferred method of beam control system is the retrodirective beam steering technique [1-4] to achieve a high precision beam control for Solar Power Satellite (SPS) because this can not be achieved by mechanical attitude control. A software retrodirective [5] system can be implemented in SPS if the Direction of Arrival (DOA) of the pilot signal can be estimated precisely. A pilot signal is sent to the power transmitting antenna from the rectenna site in a retrodirective SPS beam control system. The word 'rectenna' is derived from rectifying antenna and it is the receiving antenna for the MW power. The SPS will be located at the Geostationary Earth Orbit (GEO) and the rectenna will be located on the Earth surface. The phase of the incoming pilot signal is detected at the transmitting side and a power beam is formed accordingly. The source of Microwave (MW) power is the DC power from solar cell. The power beam is then directed towards the center of the rectenna. The research and development goals on retrodirective system mainly focused on communication applications [6]. There has been no research of precise and high speed beam control in SPS system to date. Spread Spectrum (SS) modulation can be used for the DOA detection of pilot signal and it has some interesting characteristics [7].

MUltiple SIgnal Classification (MUSIC) method can be applied for DOA detection of pilot signal in SPS. But this method has some demerits of not recognizing the signal in low SNR environment and for closely spaced signals. The DOA error can be decreased by increasing the SNR. The Time-Frequency_MUSIC (TF_MUSIC) can decrease the DOA error because it has the property of increasing the effective SNR. MUSIC and TF_MUSIC based DOA detection requires accurate knowledge of gain and phase errors of antennas and related circuits. All eigenstructure based method will also fail to detect the DOA of incoming signals accurately if antenna gain and phase errors are not taken into consideration. An eigenstructure-based method for simultaneously calibrating the unknown gain and phase parameters and the estimation of the DOA of incoming signals are reported in [8, 9]. We performed an experiment on simultaneously calibrating the antenna gain and phase errors and detecting the DOA of two pilot signals by using the methods discussed in [9] and mentioned as Calibrated_MUSIC (C_MUSIC) in this paper. The performances of MUSIC and C_MUSIC are discussed. The performance of TF_MUSIC was also tested by the calibration data. The TF_MUSIC with antenna gain and phase errors is named as Calibrated_TF_MUSIC (CTF_MUSIC). The performance of CTF_MUSIC was better than that of MUSIC and C_MUSIC.

II. DOA Estimation By Using Time Frequency-MUSIC

The time-frequency distributions (TFD) [10-13] are mainly used in the areas of non stationary signal environments, speech, biomedicine, the automotive

industry, and machine monitoring etc. [14]. The authors in [14] have shown that the TF-MUSIC outperforms conventional MUSIC in low SNR and closely spaced sources. The DOA estimation of chirp signal in low SNR environment by using TF-MUSIC is presented in [15]. One dimensional signal in the time-domain is mapped into two-dimensional signals in the TF domain in time frequency distribution (TFD). The correlation matrices of conventional MUSIC are replaced by the Spatial Time Frequency Distribution (STFD) matrices in TF-MUSIC. Only the time-frequency points in the auto term regions of STFD are considered for STFD matrix construction. The auto term region refers to the time-frequency points along the true instantaneous frequency. The cross term is the effect of two or more signals. The effect of cross term on auto term is neglected in this study because the signals are assumed stationary signals and they don't overlap. Their time frequency signatures are very distinct and the signals can be easily separated out. The subspaces obtained from the STFD are robust to both noise and angular separation of the signals incident on array. The reason of the robustness is the spreading of noise power while concentrating the source energy in the time-frequency domain [14]. This property of TFD increases the effective SNR. The increased SNR decreases the DOA error. For a data snapshot vector **x** (t) the discrete form of Wigner-Ville Distribution (WVD) is given by,

$$W_{xx}(t,f)=\sum_{\tau=-L/2}^{L/2} x(t+\tau)x^{H}(t-\tau)e^{-j2\pi f\tau/L} \quad (1)$$

Here H denotes the conjugate transpose and L is the sample length and is power of 2.

The spatial time-frequency matrix can be expressed by the following structure,

$$E[W_{xx}(t,f)]=AW_{ss}(t,f)A^{H}+E[W_{nn}(t,f)] = AW_{ss}(t,f)A^{H}+\sigma I \quad (2)$$

Where σ is the noise power at each antenna sensor and I denotes identity matrix.

Source TFD matrix can be expressed as,

$$W_{ss}(t,f)=\sum_{\tau=-L/2}^{L/2} s(t+\tau)s^{H}(t-\tau)e^{-j2\pi f\tau/L} \quad (3)$$

Where *s(t,f)* denote source signals.

The dimension of source TFD is $n\times n$ and is of the form

$$W_{ss}(t,f)=\begin{bmatrix} W_{s1s1}(t,f) & o & .. & .. & o \\ o & W_{s2s2}(t,f) & .. & .. & o \\ .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. \\ o & o & .. & .. & W_{snsn}(t,f) \end{bmatrix} \quad (4)$$

Here n is the number of sources and o denotes a negligible small value.

Now it is possible to write the matrix (4) in the following way:

$$W_{ss}(t,f)=\begin{bmatrix} G_1(t,f)S_1^2 & o & .. & .. & o \\ o & G_2(t,f)S_2^2 & .. & .. & o \\ .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. \\ o & o & .. & .. & G_n(t,f)S_n^2 \end{bmatrix} \quad (5)$$

Where $G_n(t,f)$ is the amplification factor of S_n^2 and S_n^2 is the diagonal elements of source correlation matrix of the conventional MUSIC. $G_n(t,f)S_n^2$ are the diagonal elements of the source TFD matrix.

In conventional MUSIC the correlation matrix of the array signals **x** is an $m\times m$ matrix and can be expressed as:

$$R_{xx}=E[x(t)x(t)^{H}]=AR_{ss}A^{H}+\sigma I \quad (6)$$

Here m is the number of sensors or antenna elements.

Equation (2) is similar to equation (6). In equation (2) the correlation matrices are replaced by the WVD matrices.

In TFD matrix the $SNR_{tmu}=\frac{G_n(t,f)S_n^2}{\sigma}$

But in conventional MUSIC the $SNR_{mu}=\frac{S_n^2}{\sigma}$.

So improvement of SNR in TFD based system is $G_n(t,f)$ times that of MUSIC system. This improved SNR is the reason that TF-MUSIC outperforms conventional MUSIC.

The signal TFD matrix can be written as:

$$W_{xx}=USV^{H}=[E_s E_{\bar{n}}]S[E_s E_{\bar{n}}]^{H} \quad (7)$$

Here U is an $m\times n$ unitary matrix and V is an $n\times n$ unitary matrix. E_s and $E_{\bar{n}}$ are eigen vectors of signal sub-space and noise sub-space respectively and S is the diagonal matrix of the signal eigen values and noise eigen values respectively and the values are in decreasing order. DOA are estimated by determining n values of θ for which the following Spectrum is maximized:

$$P_{tmu}(\theta)=\frac{1}{a^{H}(\theta)E_{\bar{n}}E_{\bar{n}}^{H}a(\theta)} \quad (8)$$

where a(θ) is the steering vector corresponding to θ. Fig.1 shows the simulation results of two signals coming from $\pm 0.001^{\circ}$ apart from the array broad side of 3 elements array of spacing of 0.75λ for 5.8 GHz signal frequency. The SNR of the signals was 1 dB. The DOA error was zero. Total number of snap-shots was 256. The DOA can be estimated with even much lower SNR environment and increasing the number of samples. Normal MUSIC could not estimate the DOA of the signals in this situation.

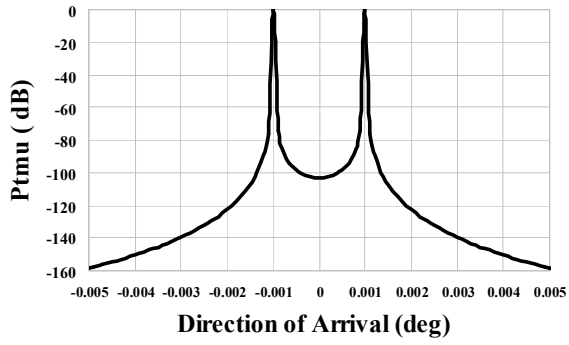


Fig.1 Time-Frequency _MUSIC(TF_MUSIC) spectrum of two signals coming from $\pm 0.001^{\circ}$ apart from the array broad side

III. Calibration of Sensor Gain and Phase Uncertainties

The sensor's output with gain and phase uncertainties can be expressed as [8]:

$$x(t) = GA(\theta)\, s(t) + n(t) \qquad (9)$$

where, $G = diag\{g_1, g_2, \ldots\ldots\ldots\ldots g_m\}$

g_m are gain and phases of antenna elements and are complex numbers and $g_1 = 1$.

The correlation matrix of the array signals **x** can be expressed as:

$$R_{xx} = E[x(t)x(t)^H] = GAR_{ss}A^H G^H + \sigma I \qquad (10)$$

The minimization of following cost function is needed to obtain G [8-9]:

$$\tilde{J} = \sum_{s=1}^{m} \left\| U_p{}^H A'(\theta_s) g \right\|^2 \qquad (11)$$

Where the matrix U_p has its columns *m-n* noise subspace eigenvectors u_p.

$p = m\text{-}n+1, \ldots\ldots\ldots m.$

$\|.\|$ stands for the Frobenius matrix norm,

$$A'(\theta_s) = diag\{a(\theta_s)\} \text{ and}$$

The vector $\tilde{g} = [1, g_2, \ldots\ldots\ldots\ldots g_m]^T$

$A'(\theta_s)$ is a diagonal matrix and θ_s are the sources' DOA.

The detailed of estimation of $\tilde{g}$, $\tilde{G}$ from $\tilde{J}$ is explained in [9].

$$\tilde{G} = diag[\tilde{g}]$$

Once $\tilde{G}$ is estimated the DOA from *n* signals can be estimated from the following formula:

$$P_{cmu}(\theta) = \frac{1}{a^H(\theta)\tilde{G}^H E_{\bar{n}} E_{\bar{n}}^H \tilde{G} a(\theta)} \qquad (12)$$

where, $E_{\bar{n}} = [e_{n+1}\ e_{n+2} \ldots\ldots\ldots\ e_m]$ = noise sub-space eigen vectors.

The maxima of the spectrum produce a sharp peak and point the DOAs of the incident signals because they are reciprocal of the nulls.

Directions of Arrivals of two signals were measured in the laboratory by using equation (12) and the method is mentioned as C_MUSIC in this paper. Later the performance of TF_MUSIC was tested by using the calibration data $\tilde{G}$ and it was found that the TF_MUSIC performed better than C_MUSIC when sensor gain and phase errors were considered. The TF_MUSIC is named as CTF_MUSIC after considering the sensor gain and phase uncertainties. The equations of TF_MUSIC (equations 2 and 8) were modified in the following ways for CTF_MUSIC:

$$E[W_{xx}(t,f)] = \tilde{G}AW_{ss}(t,f)A^H + \sigma I \qquad (13)$$

DOA for CTF_MUSIC can be estimated by determining *n* values of θ for which the following Spectrum is maximized:

$$P_{ctmu}(\theta) = \frac{1}{a^H(\theta)\tilde{G}^H E_{\bar{n}\_tmu} E_{\bar{n}\_tmu}^H \tilde{G} a(\theta)} \qquad (14)$$

Total number of antenna elements was 6 and number of incoming pilot signals was 2 during the experiment. The frequency of both the signals was 1 GHz and element spacing was half wavelength. The schematic diagram of the experimental setup is shown in Fig.2. Transmitting antenna for signal-1 (Transmitting Antenna I) was fixed at the zero degree direction from the broad side of the receiving antenna array. Transmitting antenna for signal-2 (Transmitting Antenna II) was movable. DOA of the incoming signals were measured by positioning Transmitting antenna II at four different positions. The directions of arrivals of two signals for different positions of signal-2 were measured by using MUSIC, C_MUSIC [9] and CTF_MUSIC. The actual DOA was also simulated by using TF_MUSIC and by considering 0 dB SNR.

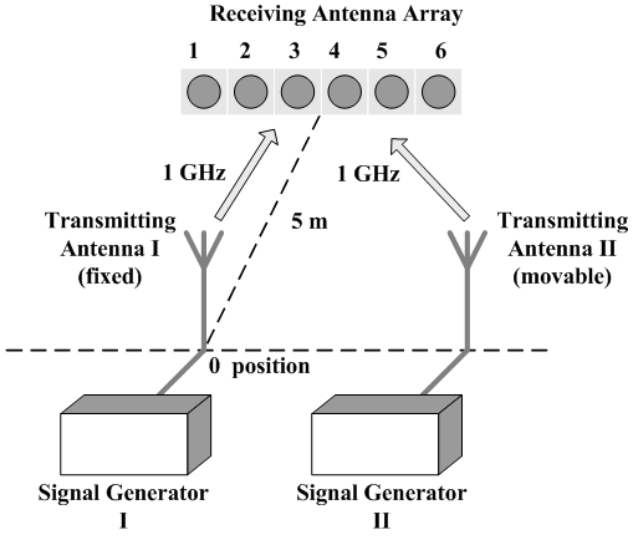


Fig.2 Schematic of experiments on Direction of Arrival (DOA) of two signals

Fig.3 shows the spatial spectrum of MUSIC, C_MUSIC, CTF_MUSIC and TF_MUSIC (simulation) for signal-1 and signal-2. Actual DOA of the signals were 0° and -16.7°. It can be seen from the figure that it is difficult to distinguish the peaks of the spatial spectrum in conventional MUSIC. CTF_MUSIC performed better than C_MUSIC and MUSIC after sensor gain and phase estimation. Sensor gain and phase were calibrated by using MUSIC method.

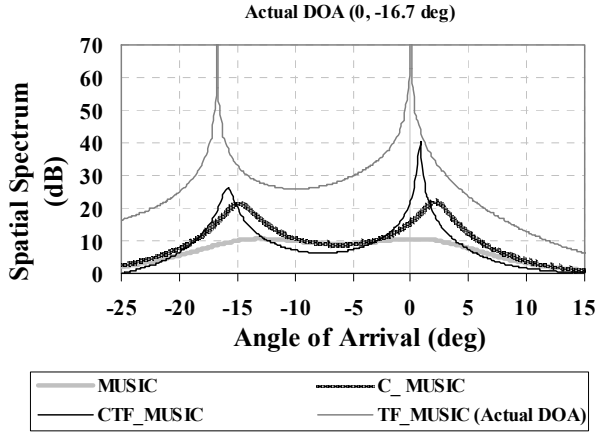


Fig.3 Spatial spectrum of MUSIC, C_MUSIC, and CTF_MUSIC of measured incoming signals and TF_MUSIC of actual (simulated) incoming signals (without any error). CTF_MUSIC performs better than MUSIC and C_MUSIC in real situations

DOA errors for signal-1 and signal-2 for different direction of arrivals by using C_MUSIC and CTF_MUSIC are shown in Table 1 and in Fig 4 and Fig 5. The DOA error for CTF_MUSIC was less than that of the C_MUSIC in most of the cases.

IV. Conclusion

Retrodirective phased-array antenna system is the preferred method of microwave (MW) beam steering for Solar Power Satellite (SPS). However, large-scale retrodirective power transmission has not yet been proven and needs further development [16]. Precise estimation of the Direction of Arrival (DOA) of pilot signal is needed for accurate MW beam steering. The Calibrated_TF_MUSIC is studied for precise DOA estimation of pilot signal. It is shown that the Calibrated_Time-Frequency_MUSIC (CTF_MUSIC) outperforms the Calibrated_MUSIC (C_ MUSIC) when antenna gain and phase errors are considered. TF-MUSIC and CTF_MUSIC work well particularly in low SNR environment.

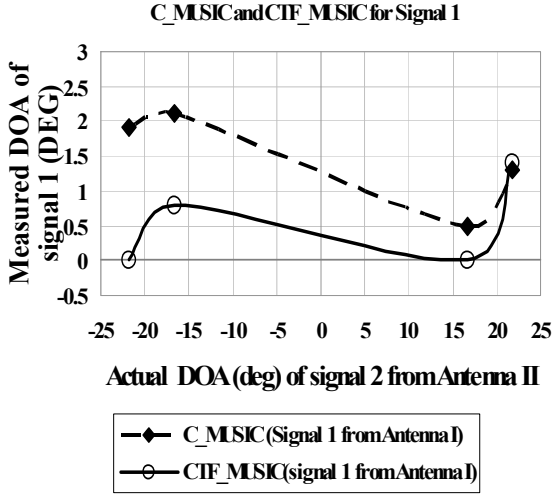


Fig.4 Direction of Arrival (DOA) errors for signal 1 for different direction of arrivals of signal 2 for C_MUSIC and CTF_MUSIC

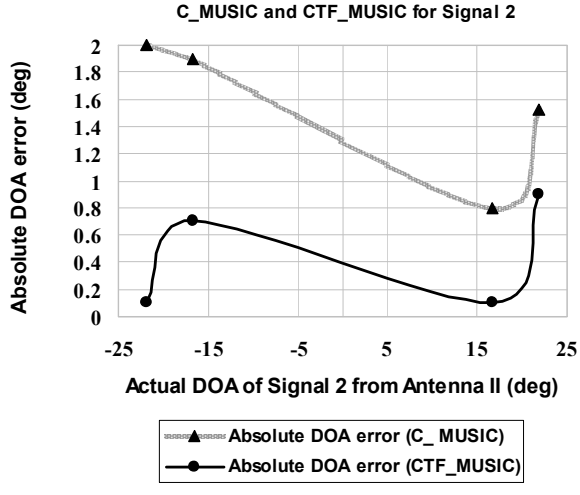


Fig.5 Direction of Arrival (DOA) errors for signal 2 for different direction of arrivals of signal 2 for C_MUSIC and CTF_MUSIC

Table 1 Direction of Arrival (DOA) errors for C_MUSIC and CTF_MUSIC

Actual DOA (deg)		Absolute DOA error (deg) For C_MUSIC		Absolute DOA error (deg) For CTF_MUSIC	
Signal-1	Signal-2	Signal-1	Signal-2	Signal-1	Signal-2
0	-21.8	1.91	2	0	0.1
0	-16.7	2.1	1.9	0.8	0.7
0	16.7	0.5	0.8	0	0.1
0	21.8	1.31	1.52	1.4	0.9

References

[1] Ralph C. Chernoff, "Large Active Retrodirective Arrays for Space Applications", IEEE Trans on Antenna & Porp. Vol. AP-27, no. 4, July 1979, p.489-496.

[2] James O. McSpadden,John C. Mankins'" Space Solar Power Programs and Microwave Wireless Power Transmission Technology", IEEE Trans. Microwave magazine for Microwave & Wireless Engineer.,vol 3,2002, pp.46-57.

[3] M.I.Skolnik, D.D.King,"Self-phasing array antennas," IEEE Trans. Antenna & Propagat.,vol. AP-12,no. 3, pp.142-149, 1964.

[4] D.L.Margerum,"Self phased arrays," in Microwave Scanning Antennas, vol. III, Array Systems, Academic Press, 1966.

[5] M. Omiya and K. Itoh, "A Fundamental System Model of the Solar Power Satellite, SPS2000", Proc. ISAP1996,2, pp. 417-420, 1996.

[6] K.Leong and T. Itoh, "Retrodirective Active Antennas". Japan-United States Joint Workshop on Space Solar Power System (JUSPS'03) proceedings, pp.216-218, 2003.

[7] K. Hashimoto, T. Tsutsumi, H. Matsumoto and N.Shinohara, "Space Solar Power System Beam Control with Spread-Spectrum Pilot Signals", Radio Science Bulletin, No. 311, dec. 2004, pp. 31-37.

[8] A. J. Weiss and B. Friedlander, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties", Journal Circuits, Systems, and Signal Processing vol. 9 no. 3 , 1990, pp. 271-300.

[9] M. Zhang and Z Zhu, "A Method for Direction Finding Under Sensor Gain and Phase Uncertainties," IEEE Trans. on Antenna and Propag. Vol.43, no.8 April 1995, pp. 880-883.

[10] L. Cohen, "Time-frequency signal analysis-A review", IEEE proceedings, vol. 77, July 1989, pp.941-981.

[11] B. Boashash, "Time-frequency signal analysis", in Advances in Spectrum Analysis and Array Processing, S. Haykin, Ed. Englewood Cliffs, NJ:Prentice-Hall, 1990, vol.I.

[12] L. Cohen, Time-Frequency Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[13] S. Qian and D. Chen Joint Time-Frequency Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[14] Y. Zhang, W. Mu, and M. G. Amin, "Subspace Analysis of Spatial Time-Frequency Distribution Matrices", IEEE Trans. On Signal Processing, vol. 49, no. 4, April 2001, p.747-759.

[15] Adel Belouchrani, " Time-Frequency MUSIC " , IEEE Signal Processing Letters, vol. 6, no. 5, May 1999, p. 109-110.

[16] H. Matsumoto," Research on Solar Power Satellites and Microwave Power transmission in Japan", IEEE Microwave magazine, vol 3, 2002, pp.36-45.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

New and Improved Method of Beam Forming with Reduced Side Lobe Levels for Microwave Power Transmission

A. K. M. Baki[a!], Kozo Hashimoto[!!], Naoki Shinohara[!!], Tomohiko Mitani[!!] and Hiroshi Matsumoto[!!]

!) School of Engineering & Computer Science (SECS), Independent University, Bangladesh
House # 10, Road # 10, Baridhara, Dhaka-1212, Bangladesh

!!) Research Institute for Sustainable Humanosphere (RISH), Kyoto University, Uji, Kyoto 611-0011, Japan
a) E-mail: baki@secs.iub.edu.bd

Abstract- Solar Power Satellite (SPS) is one option for meeting the huge future energy demand as the Earth will require sustainable electricity sources equivalent to 3 to 5 times the commercial power presently produced by 2050. One of the critical goals of SPS is to maintain highest Beam Efficiency (BE) because the microwaves (MW) from SPS will be converted to utility power unlike the MW from communication satellites. Therefore the SPS will need a highly efficient microwave power transmission (MPT) system. Another aspect of SPS is to maintain Side Lobe Levels (SLL) as minimum as possible to reduce interference to other communication systems. The SPS costs can be reduced if MPT efficiency is increased. Edge tapering of a phased array antenna, that requires a complicated system, is one way to decrease SLL and increase BE. It is also possible to achieve minimum SLL with randomly spaced element position but it does not guarantee higher BE and the determination of random element position is also a difficult task. A new method of edge tapering system from the concepts of Isosceles Trapezoidal Distribution (ITD) of array excitation and unequal element position of antenna array was investigated both through the simulation and the experimentation. The method is named as 'ITD edge tapering with Unequal element spacing (ITDU)'. ITDU method was compared with Gaussian edge tapering and simple ITD edge tapering. The lowest Maximum Side Lobe Level (MSLL) and the highest BE was achieved in newly derived ITDU technique.

I. Introduction

To supply electric power through Microwave Power Transmission (MPT) technology from the Solar Power Satellite (SPS) is one option for meeting the increased energy demand of this century. Peter Glaser conceived the first idea of SPS in 1968 [1] and it paved the way of DOE/NASA study of MPT from a satellite. Solar panels of the SPS would be placed in Geostationary Earth Orbit (GEO) at a distance of 36000 km from the Earth's surface.

Global warming is another key problem in the 21st century. Electricity generated from any renewable energy source, such as hydro, wind, biomass, geothermal and solar, is considered "green" because of the negligible impact on greenhouse gas emissions [2].

Different approaches of MPT have been studied. The two indices for the evaluation of the Microwave (MW) beam are the Beam Efficiency (BE)/Beam Collection Efficiency (BCE) and Maximum Side Lobe Level (MSLL). BE is the ratio of energy flow within the main beam to the whole transmitted power and BCE of the MPT systems is the ratio of energy flow that is intercepted by the rectenna to the whole transmitted power. The word "rectenna" is derived from the words "rectifying antenna". Rectenna is the receiving antenna for the MW power from the SPS. The received MW power on the rectenna will be converted to the utility power. BE for two dimensional beam pattern quantifies the solid angle extent of the main beam relative to that of the entire pattern and can be expressed as [3]:

$$BE_{2D} = \iint_{main\_beam} \left|P(\theta,\phi)\right|^2 d\Omega \Big/ \iint_{4\pi} \left|P(\theta,\phi)\right|^2 d\Omega \quad (1)$$

Here,
$P(\theta,\phi)$ is the radiated electric field.
BE for one dimensional case can be expressed as:

$$BE_{1D} = \int_{\theta_m} \left|P(\theta)\right|^2 d\theta \Big/ \int_{\theta_w} \left|P(\theta)\right|^2 d\theta \quad (2)$$

θ_m is the angle sector due to one dimensional main beam and θ_w is the angle sector of $\pm 90^\circ$.
$P(\theta)$ is the one dimensional radiated electric field.

BCE for two dimensional array and rectangular/square rectenna can be expressed as[4]:

$$BCE_{2D} = \int_{\theta_{ry}} \int_{\theta_{rx}} \left|P(\theta_x,\theta_y)\right|^2 d\theta_x d\theta_y \Big/ \int_{\theta_{ty}} \int_{\theta_{tx}} \left|P(\theta_x,\theta_y)\right|^2 d\theta_x d\theta_y \quad (3)$$

978-1-4244-2014-8/08/$25.00 (c)2008 IEEE

Here,

θ_{tx} ; θ_{ty} ; are ±90 degree angle sector.

θ_{rx} ; angle sector due to x dimension of rectenna.

θ_{ry} ; angle sector due to y dimension of rectenna.

$P(\theta_x, \theta_y)$ is the energy of the radiated electric field.

BCE for one dimensional case can be expressed as:

$$BCE = \int_{\theta_r} |P(\theta)|^2 d\theta \Big/ \int_{\theta_w} |P(\theta)|^2 d\theta \qquad (4)$$

θ_r is the angle sector due to one dimensional rectenna and θ_w is the angle sector $\pm 90^\circ$.

$P(\theta)$ is the energy of the one dimensional radiated electric field.

It is necessary to suppress Grating Lobes (GL) and Side Lobe Levels (SLL) for higher BCE and to avoid interference to other communication systems. When GL appear and SLL increase, the transmitted power is absorbed into these lobes and it causes the reduction of BE/BCE and increase of interference. Phased array antenna system has been proposed in SPS to maintain higher BCE. If all antennas are uniformly excited then the main beam will carry only a part of the total energy due to higher SLL. The targeted BCE for SPS is 90% to reduce the cost. It is possible to increase BCE and reduce SLL if edge tapering system can be adopted in MPT antennas.

Different power distribution systems for MPT are discussed in [4-6] and the references there in. Edge tapering for SPS like Gaussian, Chebyshev, and Taylor distribution has some complexity of different power output levels at different antennas and problem of heat radiation. It is also possible to reduce SLL with statistically thinned array but it does not guarantee higher BCE.

Isosceles Trapezoidal Distribution (ITD) edge tapered antenna, which is a new concept, is experimented for the first time for SPS. ITD is better than full edge tapering and uniform amplitude distribution. It was found that the highest BCE and lowest SLL are possible to achieve in ITD edge tapering [4]. Different amplitude distribution systems like uniform, Gaussian, Dolph-Chebyshev and the newly derived ITD method have been compared. The SLL reduction in ITD is even higher than those of other kinds of edge tapering. Only a small number of antennas from each side of the phased array antenna are tapered in this method. ITD edge tapering is almost uniform so it is technically better. The power density at the center of the array of the ITD system can be made lower than that of the Gaussian or similar kinds of distributions for the same power transmission. Therefore thermal behavior at the center of the ITD edge tapered phased array antenna is better than that of the Gaussian and other kinds of edge tapering.

ITD performance is further improved from the perspective of both MSLL and BE/BCE by using combined equal and unequal spacing of the antenna elements. This later method is named as 'ITD edge tapering with Unequal element spacing (ITDU)'. Amplitude distribution of ITDU is same as that of ITD and is also technically better than that of Gaussian distribution. The unequal element spacing for ITDU is derived from the ITD concept and by using a mathematical function. There was no GL in the observation range of $\pm 90^\circ$ due to unequal spacing with ITD. The calculation process of unequal element spacing for ITDU is much easier than that of thinned array or random array and it will be technically easier to be built. Experimental results on ITD and ITDU are also presented.

II. Concept of Isosceles Trapezoidal Distribution (ITD) Edge Tapering

The concept of ITD edge tapering for one dimensional (1D) array is shown in Fig. 1. Amplitude of only a few edge transmitting antennas/units are tapered in this method. Amplitudes of remaining most antennas/units are uniform. In recent SPS design, a concept of "unit" is adopted. Each unit consists of several phased array antenna elements. For example 25 elements or 10 elements can be considered as a "unit". Array Factor (AF) with ITD is:

$$AF = \sum_{n=1}^{N_t} A_{tn} e^{jn\psi_1} + \sum_{m=1}^{N} e^{j(m+N_t)\psi} + \sum_{n=1}^{N_t} A_{tn} e^{j(n+N_t+N)\psi_2} \qquad (5)$$

where, N = Number of SPS units with uniform amplitudes,

N_t = Number of units tapered from each side,

$\psi = \beta D_u (\sin\theta - \sin\theta_0)$.

$\psi_1 = -\beta D_u (\sin\theta - \sin\theta_0)$.

$\psi_2 = [(N-1)D_u + nD_u] (\sin\theta - \sin\theta_0)\beta$.

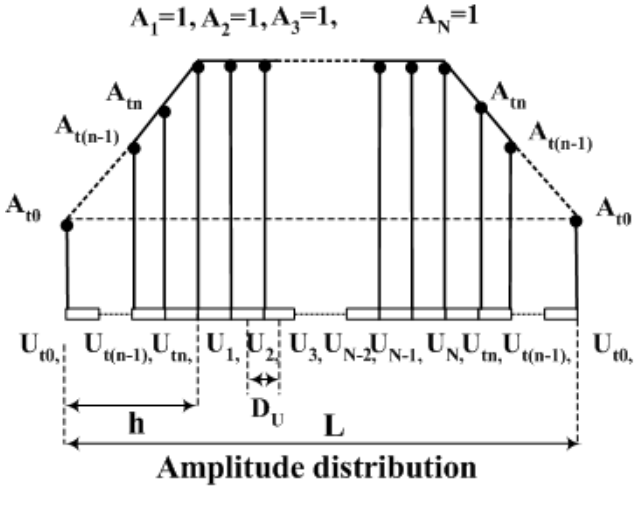


Fig. 1 Isosceles Trapezoidal Distribution (ITD) type edge tapering for Solar Power Satellite (SPS) transmitting antennas.

D_u = inter-unit spacing (m).

N_T = Total number of antenna units= N+ $2N_t$.

$\beta = 2\pi/\lambda$ = phase constant.

The amplitude distribution of the ITD tapering is:

$$A_{tn} = (1 - A_{t0})n/N_t + A_{t0} \qquad (6)$$

A_{t0} is the amplitude of each of the end units.

θ_0 = Direction of beam maximum along the broad side.

n = 0, 1, 2…… N_t , $\qquad 0 \le A_{tn} \le 1$

In Figure 1,
h = Length of the antennas tapered from each side.
L= Total length of the array.
$A_1=A_2=A_3=$ …… $= A_N=1$ (normalized maximum power).

m = 1, 2, 3 ….N.

.

III. Concept of Unequal Element Spacing for 'ITD edge tapering with Unequal element spacing (ITDU)'

If the equal element spacing for ITD is d then the unequal element spacing for ITDU will be d×*M(x)*. Here the multiplying factor *M(x)* for the determination of element spacing is to be expressed as equation (7):

$$M(x) = Sinc[A(x)] \Big/ \min\langle Sinc[A(x)]\rangle = Sinc[A(x)]/0.8415 \qquad (7)$$

Here *A(x)* is the amplitude distribution of an isosceles trapezoid and not the power distribution of the antennas as is considered in section II. The variable *x* is the position of antenna elements. If the total number of elements is (2P+1) then the range of *x* is from 1 to 2P.

IV. Simulation of Different Edge Tapering Systems

MW Radiation patterns, MSLL and BE for different tapering levels, inter-element spacing, beam steering angles and for Uniform, Gaussian, ITD and ITDU power distribution systems were compared and summarized in Tables 1-3. The BE is calculated instead of BCE. If BE is highest then BCE will also be highest when the size of the rectenna is equal to or greater than the size of the main beam on rectenna. The Normalized radiation patterns for phased array antenna of -10 dB ITD (53909 elements) edge tapering and -10 dB ITDU (50001 elements) are shown in Fig. 2. The number of antenna elements in two cases is different to keep the same array antenna length.
An example of power distributions of total 109 watts power for different power distribution is shown in Fig. 3. The number of antennas in Uniform, Gaussian and ITD edge tapering is chosen 109 and that in ITDU is chosen 101 to keep the same array length as mentioned before. In Gaussian distribution the power distribution at the centre of the transmitter is the highest which is a demerit because of thermal problem and other technical problems.

The MSLL for ITDU was found to be the lowest in all cases when it was compared with Gaussian, ITD and uniform amplitude distribution [5]. The MSLL of ITDU is much lower than that of ITD. Moreover the BE was found to be the highest in newly derived ITDU. For example the MSLL for -10 dB ITDU was -26 dB which was the lowest and BE was 98.97 % (for 101 elements) which was the highest when compared with those of equally tapered Gaussian and ITD.

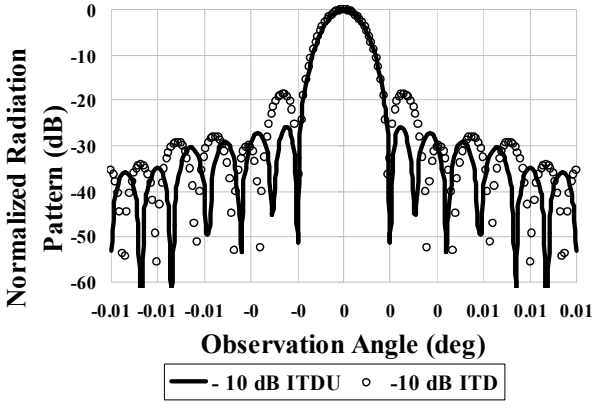


Fig. 2 Normalized radiation pattern for phased array antenna of -10 dB Isosceles Trapezoidal Distribution (ITD) (53909 elements) edge tapering and -10 dB ITD edge tapering with Unequal element spacing (ITDU) (50001 elements). 15000 elements from each were tapered in ITD and ITDU and 20000 elements from each side were of unequal spacing in ITDU

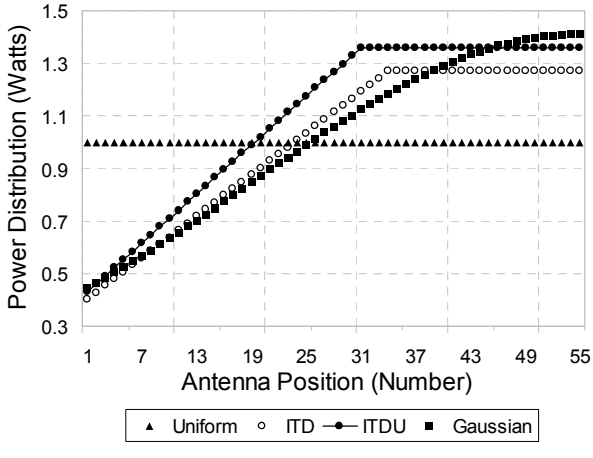


Fig. 3 Uniform amplitude distribution, -10 dB ITD (33 elements tapered), -10 dB ITDU (30 elements tapered and 40 elements are of unequal spacing) and -10 dB Gaussian power distribution for total 109 watts power in each case (The number of elements in ITDU is 101 and the number of elements in each of the other cases is 109). Power distributions for half of the arrays are shown in the figure. Power distributions for rest half of the arrays are symmetrical.

Table 1 Beam Efficiency (BE) and Maximum Side Lobe Level (MSLL) for uniform distribution, -10 dB Gaussian, -10 dB Isosceles Trapezoidal Distribution (ITD) and -10 dB ITD edge tapering with Unequal element spacing (ITDU) for 0 degree beam steering angle from broad side

Type of distribution	Total No. of elements	No. of Elements tapered	No. of elements of unequal spacing	BE (%)	MSLL (dB)
Uniform	109/-	0	0	89.54/-	-13/-
Gaussian	109/ 53909	all	0	98.33/ 99.96	-22.46/ -22.5
ITD	109/ 53909	33/ 16172	0	98.30/ 98.31	-18.7/ -18
ITDU	101/ 50001	30/ 15000	40/ 20000	**98.97/ 99.999**	**-26/ -26**

Table 2 Beam Efficiency (BE) and Maximum Side Lobe Level (MSLL) for uniform distribution, -10 dB Gaussian, -10 dB Isosceles Trapezoidal Distribution (ITD) and -10 dB ITD edge tapering with Unequal element spacing (ITDU) for 10 degree beam steering angle from broad side

Type of distribution	Total No. of elements	No. of Elements tapered	No. of elements of unequal spacing	BE (%)	MSLL (dB)
Uniform	109	0	0	89.8	-13
Gaussian	109	all	0	98.6	-22
ITD	109	33	0	98.2	-19
ITDU	101	30	40	**99.06**	**-26**

Table 3 Beam Efficiency (BE) and Maximum Side Lobe Level (MSLL) for uniform distribution, -20 dB Isosceles Trapezoidal Distribution (ITD) and -20 dB ITD edge tapering with Unequal element spacing (ITDU) for 0 degree beam steering angle from broad side

Type of distribution	Total No. of elements	No. of Elements tapered	No. of elements of unequal spacing	BE (%)	MSLL (dB)
ITD	109/ 53909	33/ 16172	0	98.76/ 98.84	-18.8/ -18.8
ITDU	101/ 50001	30/ 15000	37/ 18500	**99.69/ 99.84**	**-32/ -32**

V. Laboratory Experimentation with ITD and ITDU

Laboratory experiments on uniform, ITD and ITDU edge tapering system were done by using the SPORTS (**S**olar **PO**wer **R**adio **T**ransmission **S**ystem) 5.8 GHz "beam forming subsystem" in the **M**icrowave **E**nergy **T**ransmission **LAB**oratory (METLAB) of the Kyoto University. METLAB is an anechoic chamber for MPT experiment. The experimental set up is shown in Fig. 4. 11 one dimensional phased array antenna elements were used during the experiments. Here only the experimental result on ITDU is shown. Fig. 5 shows the measured and simulated radiation patterns of ITDU. There was a difference between measured and simulated radiation patterns and the reasons are explained elaborately in [5]. From this experimental and simulated data it was confirmed that ITDU is the best when it is compared with ITD and Gaussian edge tapering. The experimentations on ITD and ITDU were done for the first time and it was done no where before.

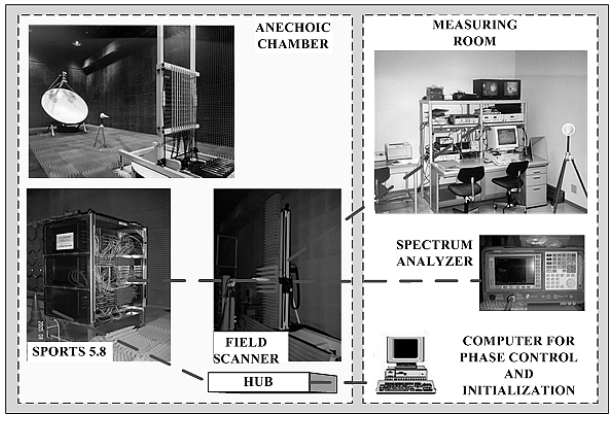


Fig. 4: The experimental set up of Solar POwer Radio Transmission System (SPORTS) 5.8.

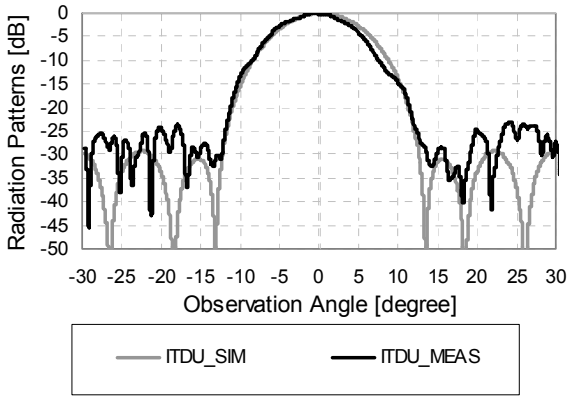


Fig. 5 Measured and simulated radiation patterns of ITD edge tapering with Unequal element spacing (ITDU) with 11 phased array antenna elements.

VI. Simulation with 2D Array

The 2D radiation pattern for -20 dB ITDU (101×101 elements) is also simulated and is shown in Fig.6. The number of elements tapered was 30 from each side of the X and Y coordinates and for both ITD and ITDU. The number of elements of unequal spacing was 37 from each side of the X and Y coordinates of ITDU. Equally well performance, as described in Table 3, was also achieved in -20 dB ITDU tapering.

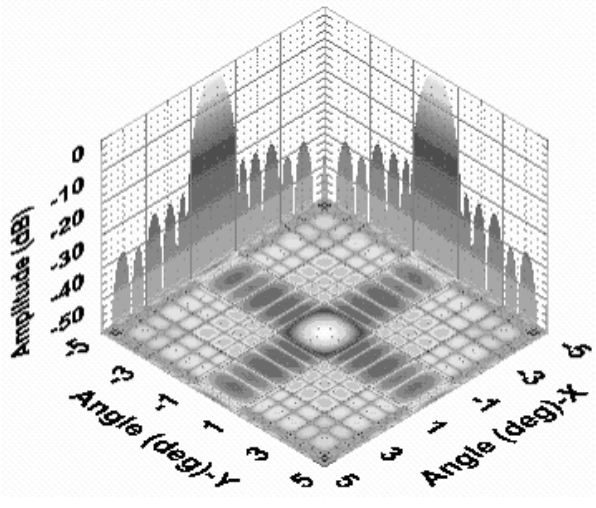


Fig. 6 Three dimensional projected beam pattern for -20 dB ITDU of two dimensional arrays in the XY, XZ and YZ planes. Projected pattern in the XY plane is the contour plot of the radiation pattern.

VII. Conclusion

Microwave power distribution systems of different kinds for Solar Power Satellite (SPS) phased array transmitting antennas are presented and compared. Edge tapering is essential for higher Beam Efficiency (BE) in Microwave Power Transmission (MPT) though it has thermal and technical complexities. Isosceles Trapezoidal Distribution (ITD) edge tapering system which is a new concept is presented. It is possible to maintain higher BCE but higher number of antennas is needed to be tapered in ITD to achieve lower Maximum Side Lobe Level (MSLL). ITD is still better than Gaussian distribution considering BCE, power distribution and other SLL. ITD edge tapering with Unequal element spacing (ITDU) concept which is an improvement of ITD edge tapering is also presented. It would be possible to maintain highest BE and lowest MSLL by incorporating ITDU edge tapering of phased array antenna in MPT system. The exposure level to humans and all other living animals/things outside the rectenna area is less due to ITDU because of its minimum SLL when it is compared to that of Gaussian distribution or ITD edge tapering. The better performance of ITDU over Gaussian and ITD was confirmed through experimentation. Finally it can be concluded the best form of power distribution for SPS MPT system could be considered to be the ITDU.

References

[1] P. E. Glaser, "Power from the sun; its future," Science, vol. 162, pp. 857-861, Nov. 22, 1968.

[2] Saifur Rahman, “Green Power: What Is It and Where Can We Find It? “ IEEE Power & Energy magazine, pp.30-37, January/February 2003.

[3] Warren L. Stutzman and Gary A. Thiele, “Antenna Theory and Design”, 2nd edition, John Wiley & Sons, Inc. pp.296.

[4] A. K. M. Baki, N. Shinohara, H. Matsumoto, K. Hashimoto, and T. Mitani, “Study of Isosceles Trapezoidal edge tapered phased array antenna for Solar Power Station/Satellite”, IEICE TRANS. COMMUN., VOL.E90-B, NO.4, pp 968-977, APRIL 2007.

[5] A. K. M. Baki, N. Shinohara, H. Matsumoto, K. Hashimoto, and T. Mitani, “Isosceles-Trapezoidal-Distribution Edge Tapered Array Antenna with Unequal Element Spacing for Solar Power Station/Satellite”, IEICE TRANS. COMMUN., VOL.E91-B, NO.2, pp 527-535, FEBRUARY 2008.

[6] N. Shinohara, B. Shishkov, H. Matsumoto, K. Hashimoto, and A. K. M. Baki, “New Stochastic Algorithm for Optimization of Both Side Lobes and Grating Lobes in Large Antenna Arrays for MPT”, IEICE TRANS. COMMUN., VOL.E91-B, NO.1, pp 286-296, JANUARY 2008.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Analysis of Wire Antennas by Solving Pocklington's Integral Equation Using Wavelets

Bindubritta Acharjee and *Md. Abdul Matin, Member IEEE*

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka – 1000.

E-mail: bindubritta@yahoo.com; amatin@eee.buet.ac.bd

Abstract – Solving Pocklington's integral equation for the thin and straight wire antennas by moment method, the current distributions of a half-wave and a full- wave dipole have been deduced using semi-orthogonal wavelet bases that are constructed from the second-order cardinal B-spline. Results agree well with King's Three-term theory and the measured values of Mack. The input impedance and radiation pattern of these antennas are calculated with the corresponding current distribution. The method is then extended to antenna array analysis. Current distributions on each antenna of a three-element Yagi-Uda array have been derived and compared with those based on King's Three-term theory. With the current distribution, the input impedance and radiation pattern of the array have also been obtained within reduced computing time.

I. Introduction

Analysis of electromagnetic radiation, scattering from material bodies and wave propagation in a dispersive medium were formulated by James Clerk Maxwell [1] in 1864, which provide the foundation of classical electromagnetism. In 1887 Hertz [2] experimentally verified the wave phenomena consequent to Maxwell's equations. In 1897 Pocklington [3] extended the insights of Lorentz [4] and Hertz [2] by deducing an integral equation for the current along a cylindrical conductor. The equation is well known as Pocklington's integral equation and it forms the basis of linear wire antenna analysis. But due to the highly singular kernel, no attempt was made till 1937 for analytical solution of the equation to determine the actual current distribution. Instead, a convenient sinusoidal distribution was assumed for the half-wave dipole by Carter [5] in 1932 and for an antenna with arbitrary length by Brown [6] in 1937. However, leading to infinite impedance for an one-wavelength antenna, approximate solutions of the integral equation have been resolved by L. V. King [7] in 1937, by Hallén [8] in 1938, by King [9] in 1965 and continuing to the present. Among the various analytical methods some permit successive improvement; e.g. iteration method applied by King and Middleton [10], Fourier series expansion of the current distribution by Duncan and Hinchey [11], moment method for electrically short antenna by Harrington [12], generalized ray method for high-frequency analysis by Burkholder [13] and so on. In the moment method using conventional bases the resultant impedance matrix become dense. Hence the inversion and final solution of the system is very time consuming. To overcome the huge memory requirement and computation time, wavelet bases was first proposed by Beylkin, Coifman and Rokhlin [14]. However, using orthogonal wavelet bases the matrix is dense yet. Thus Chui and Quak [15] used wavelets on a bounded interval; Nevels, Goswami and Tehrani [16] used semi-orthogonal spline wavelets; Tretiakov and Pan [17] used discrete wavelet packet to solve Pocklington's equation by moment method. In this paper, we solve the Pocklington's integral equation of a wire antenna by using semi-orthogonal wavelet bases in the moment method. For half-wave and full-wave dipole antennas, we compare the results with those obtained by King's [9] three-term theory and Mack's [18] experiment. We extend the wavelet based moment method to analyze a three-element Yagi-Uda array. Calculation of input impedance and radiation pattern in each case is carried out and concluded with a discussion about the accuracy and efficiency of the analysis.

II. General Formulation for the Wire Antenna Analysis

A. Current Distribution of Linear Antenna

For a z-directed thin cylindrical wire antenna, as shown in Fig. 1 of length l and radius a, with a current distribution I(z) along its length; the Pocklington's integral equation is written as –

$$\frac{j\eta\lambda}{8\pi^2}\int_{-l/2}^{l/2} I_z(z')\left(\frac{\partial^2}{\partial z^2}+k^2\right)\frac{e^{-jk\sqrt{(z-z')^2+a^2}}}{\sqrt{(z-z')^2+a^2}}dz' = E_z^{in}(z)$$

where, $\eta = \sqrt{(\mu_0/\varepsilon_0)}$, $k = 2\pi/\lambda$. μ_0 is the permeability, ε_0 is the permittivity of space and λ is the wavelength. $E^{in}(z)$ is the incident field, which induces the current in the antenna. For a transmitting antenna, $E^{in}(z) = V_0/\Delta z$, where V_0 is the applied voltage within a short gap Δz of antenna length and for a receiving antenna with a uniform plane wave E_0 incident at a polar angle θ, $E^{in}(z) = E_0 \sin\theta\, e^{jkz\cos\theta}$; where the propagation vector **k** is co-planner with the antenna axis. The magnetic vector potential of the antenna will be z-directed as –

$$\overline{A}(\hat{r}) = \hat{z}\frac{\mu}{4\pi}\int_{-l/2}^{l/2} I_z(z')\frac{e^{-jk\sqrt{(z-z')^2+a^2}}}{\sqrt{(z-z')^2+a^2}}dz'$$

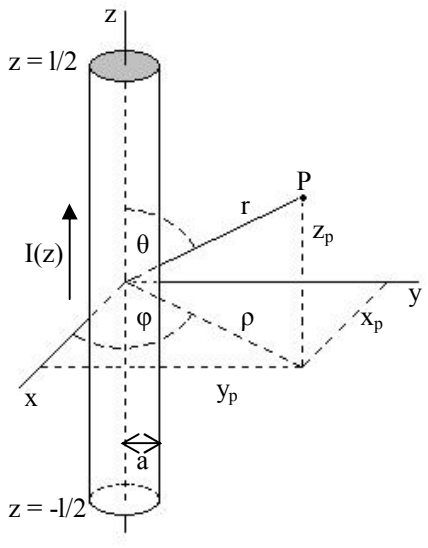


Fig. 1 Thin-wire model of cylindrical antenna

Now, to expand the unknown current function $I_Z(z')$, semi-orthogonal wavelet bases constructed from the second-order cardinal B-spline, can be used.

The scaling functions, at the lowest scale, are given by –

$$\phi_{2,q} = \begin{cases} z_2 - q; & z_2 \in [q, q+1] \\ 2-(z_2-q); & z_2 \in [q+1, q+2] \end{cases}$$

where, q = 0, 1 and 2.

And the wavelets are given by –

$$\psi_{p,q} = \frac{1}{6}\begin{cases} z_p - q; & z_p \in [q, q+0.5] \\ 4-7(z_p-q); & z_p \in [q+0.5, q+1] \\ -19+16(z_p-q); & z_p \in [q+1, q+1.5] \\ 29-16(z_p-q); & z_p \in [q+1.5, q+2] \\ -17+7(z_p-q); & z_p \in [q+2, q+2.5] \\ 3-(z_p-q); & z_p \in [q+2.5, q+3] \end{cases}$$

where, p is the respective scale (≥ 2) and q = 0, …, 2^p -3.

The actual coordinate position z is related to z_p according to –

$$z_p = 2^p \left[\frac{z' + l/2}{l} \right]$$

Hence, the current function can be expressed as –

$$I_z(z') = \sum_{p=2}^{p_u} \sum_{q=0}^{2^p-3} d_{p,q}\ \psi_{p,q}(z') + \sum_{q=0}^{2} c_{2,q}\ \phi_{2,q}(z')$$

Substituting the current expression in Pocklington's integral equation we obtain –

$$\sum_{p=2}^{p_u} \sum_{q=0}^{2^p-3} d_{p,q} \int_{-l/2}^{l/2} Z(z,z')\ \psi_{p,q}(z')\ dz'$$

$$+ \sum_{q=0}^{2} c_{2,q} \int_{-l/2}^{l/2} Z(z,z')\ \phi_{2,q}(z')\ dz' = E_z^{in}(z)$$

where,

$$Z(z,z') = \frac{j\eta\lambda}{8\pi^2}\left(\frac{\partial^2}{\partial z^2} + k^2 \right) \frac{e^{-jk\sqrt{(z-z')^2+a^2}}}{\sqrt{(z-z')^2+a^2}}$$

Applying Galerkin method we have –

$$\sum_{p=2}^{p_u} \sum_{q=0}^{2^p-3} d_{p,q} \int_{-l/2}^{l/2} \int_{-l/2}^{l/2} Z(z,z')\ \psi_{p,q}(z')\ \psi_{i,m}(z)\ dz'dz$$

$$+ \sum_{q=0}^{2} c_{2,q} \int_{-l/2}^{l/2} \int_{-l/2}^{l/2} Z(z,z')\ \phi_{2,q}(z')\ \psi_{i,m}(z)\ dz'dz$$

$$= \int_{-l/2}^{l/2} E_z^{in}(z).\psi_{i,m}(z)\ dz$$

where, i ε (2, i_u), m ε (0, 2^{iu} -3); and

$$\sum_{p=2}^{p_u} \sum_{q=0}^{2^p-3} d_{p,q} \int_{-l/2}^{l/2} \int_{-l/2}^{l/2} Z(z,z')\ \psi_{p,q}(z')\ \phi_{2,m}(z)\ dz'dz$$

$$+ \sum_{q=0}^{2} c_{2,q} \int_{-l/2}^{l/2} \int_{-l/2}^{l/2} Z(z,z')\ \phi_{2,q}(z')\ \phi_{2,m}(z)\ dz'dz$$

$$= \int_{-l/2}^{l/2} E_z^{in}(z)\ \phi_{2,m}(z)\ dz$$

where, m ε (0, 2^{iu} -3).

These equations can be written in a compact matrix form as –

$$\begin{pmatrix} Zg_{0,0} & Zg_{0,1} & \cdots & Zg_{0,n} & \cdots & Zg_{0,2^{p_u+1}+1} \\ Zg_{1,0} & Zg_{1,1} & \cdots & Zg_{1,n} & \cdots & Zg_{1,2^{p_u+1}+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Zg_{n,0} & Zg_{n,1} & \cdots & Zg_{n,n} & \cdots & Zg_{n,2^{p_u+1}+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Zg_{2^{p_u+1}+1,0} & Zg_{2^{p_u+1}+1,1} & \cdots & Zg_{2^{p_u+1}+1,n} & \cdots & Zg_{2^{p_u+1}+1,2^{p_u+1}+1} \end{pmatrix} \bullet \begin{pmatrix} d_{2,0} \\ d_{2,1} \\ \cdots \\ d_{p,q} \\ \cdots \\ c_{2,2^{p_u+1}+1} \end{pmatrix} = \begin{pmatrix} Eg_0 \\ Eg_1 \\ \cdots \\ Eg_n \\ \cdots \\ Eg_{2^{p_u+1}+1} \end{pmatrix}$$

Solving the matrix the unknown current coefficients can be calculated to deduce the current distribution $I_Z(z')$.

For an arbitrary array of parallel linear antennas the current distribution on each antenna will be effected by mutual couplings between the antennas.

Let us consider a Yagi-Uda array of three z-directed parallel dipoles with centers at locations (0, 0, 0), (0, d_{12}, 0) and (0, d_{13}, 0) as shown in Fig. 2. The second or active dipole is center-driven by a voltage generator V_2 and the other two dipoles, i.e. reflector and director are parasitic. Let $I_1(z)$, $I_2(z)$ and $I_3(z)$ are the currents induced on the dipoles; l_1, l_2 and l_3 are the antenna lengths; a_1, a_2, a_3 are their radii respectively. Therefore, the magnetic vector potential of the antenna array with respect to an observation point (x, y, z) will be –

$$\overline{A}(\hat{r}) = \hat{z}\frac{\mu}{4\pi}\sum_{n=1}^{n=3}\left(\int_{-l_n/2}^{l_n/2} I_{z_n}(z') \frac{e^{-jk\sqrt{(z-z')^2+(x-x_n)^2+(y-y_n)^2}}}{\sqrt{(z-z')^2+(x-x_n)^2+(y-y_n)^2}} dz' \right)$$

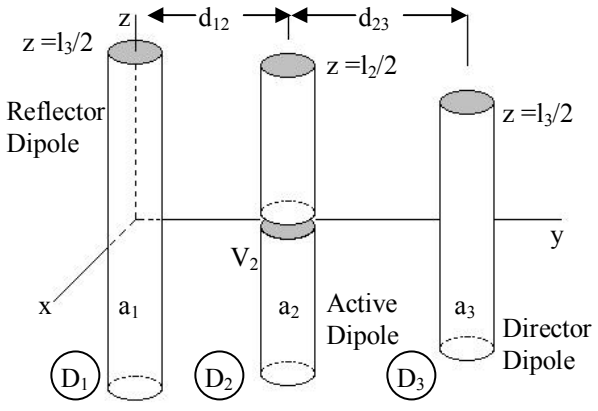


Fig. 2 Three-element Yagi-Uda array

The Pocklington's integral equation for the m-th antenna can be generalized into –

$$\frac{j\eta\lambda}{8\pi^2}\left(\frac{\partial^2}{\partial z^2}+k^2\right)\sum_{n=1}^{n=2}\left(\int_{-l_n/2}^{l_n/2} I_{z_n}(z')\frac{e^{-jk\sqrt{(z-z')^2+d_{mn}{}^2}}}{\sqrt{(z-z')^2+d_{mn}{}^2}}dz'\right)=E_{z_n}^{in}(z)$$

where, $d_{mn} = y_m - y_n$; when $m \neq n$
$= a_n$; when $m = n$
and $E^{in}_{zn}(z) = 0$; for n = 1, 3.
$= V_2 / \Delta z$; for n = 2

The unknown current function $I_{Zn}(z')$ of each antenna can be expanded using semi-orthogonal wavelets. Applying Galerkin method in each equation and solving the impedance matrix, the current distribution on each antenna can be derived.

B. Input Impedance of Linear Antenna

The input impedance of a linear wire antenna can be calculated by evaluating the near-field on the surface of the antenna. Hence, the impedance is –

$$Z = -\frac{1}{|I(0)|^2}\int_{-l/2}^{l/2} E(z)\ I(z)\ dz$$

C. Far-field Radiation Pattern of Antenna

At far-field condition ($r \gg 2l^2/\lambda$) the magnetic vector potential can be expressed as –

$$\overline{A}(\hat{r}) = \hat{z}\frac{\mu\, e^{-jkr}}{4\pi\, r}\overline{F}_z(\theta,\phi)$$

where, $\mathbf{F}_z(\theta, \varphi)$ is known as the radiation vector.

For a single linear antenna –

$$|F_z(\theta)| = \int_{-l/2}^{l/2} I(z')\ e^{jk\ \cos\theta\ z'}\ dz'$$

For an array of linear antennas –

$$|F_z(\theta,\phi)| = \sum_{n=1}^{n=3} e^{jk\ \sin\theta\ \sin\phi\ y_n}\left(\int_{-l_n/2}^{l_n/2} I_n(z')\ e^{jk\ \cos\theta\ z'}dz'\right)$$

In terms of radiation vector, the radiation intensity of the antenna becomes –

$$U(\theta,\phi) = \frac{\eta\, k^2}{32\pi^2}|F_z(\theta,\phi)|^2$$

and the normalized power gain is –

$$g(\theta,\phi) = \frac{U(\theta,\phi)}{U_{\max}}$$

III. Simulations and Results

In this paper we have considered a half-wave dipole antenna, a full-wave dipole antenna and a three-element Yagi-Uda array for simulation of the wire antenna analysis using semi-orthogonal wavelets. At first we have calculated the current distribution from the Pocklington's integral equation by moment method. Then we have deduced the necessary antenna parameters e.g. the input impedance and the radiation pattern with the current distribution.

A. Current Distribution

A1. Current Distribution of a Half-wave Dipole

A half-wave transmitting dipole antenna of radius 7.022 millimeter, operating at 300 MHz, has been considered. In Fig. 3 the analyzed normalized components of currents are shown in solid lines, calculated values from Three-term theory are shown in dash-dotted lines and the measured values of Mack are shown in dotted line. The analysis has been executed at scale = 3.

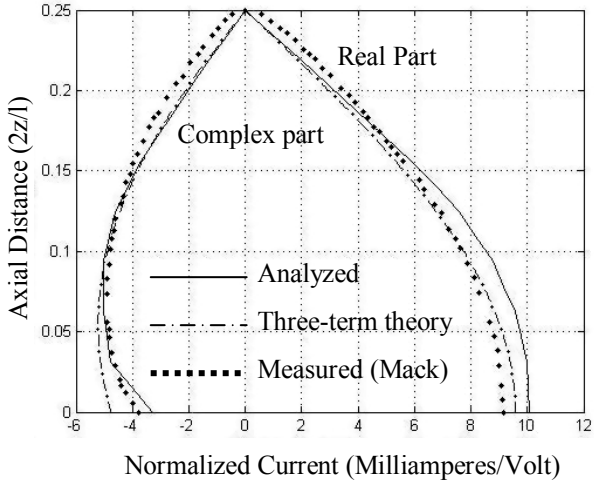


Fig. 3 Current distribution along a half-wave dipole

A2. Current Distribution of a Full-wave Dipole

Similarly, a full-wave transmitting dipole antenna of radius 7.022 millimeter, operating at 300 MHz, has been considered. In Fig. 4 the analyzed normalized components of currents are shown in solid lines, calculated values from Three-term theory are shown in dash-dotted lines and measured values of Mack are shown in dotted line. The analysis has been executed at scale = 4.

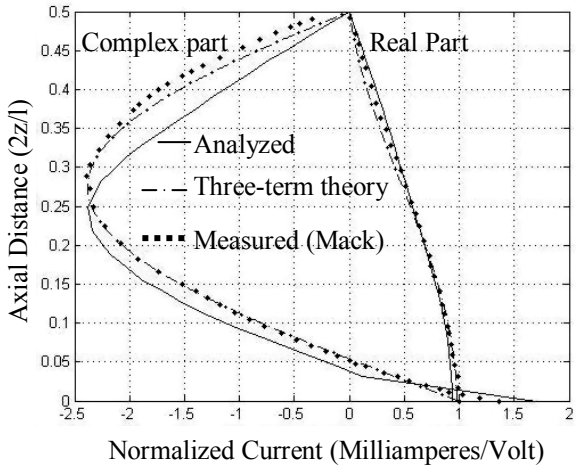


Fig. 4 Current distribution along a full-wave dipole

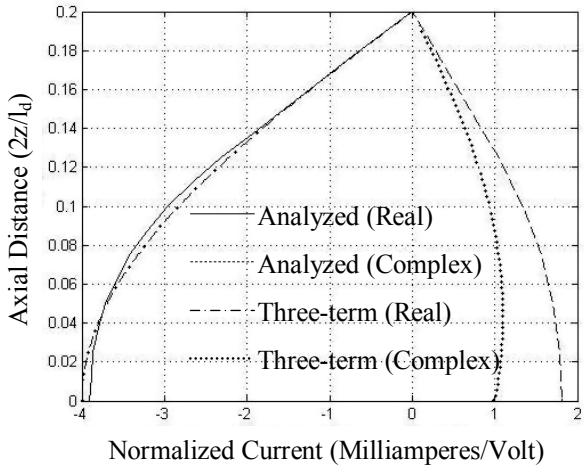


Fig. 7 Current in director

A3. Current Distributions of a three-element Yagi-Uda array

Current distributions of a three-element Yagi-Uda array are respectively shown in Fig. 5, Fig. 6 and Fig. 7.

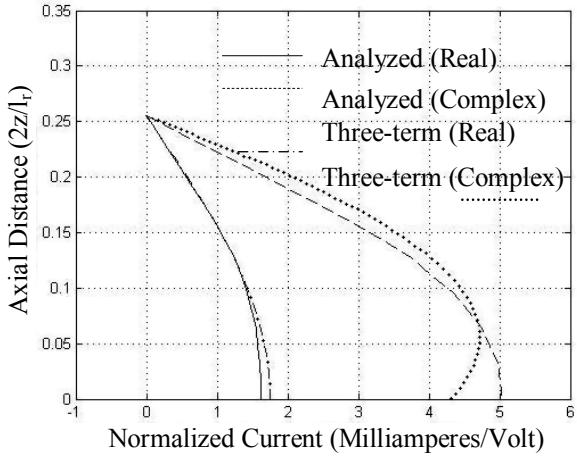


Fig. 5 Current in reflector dipole

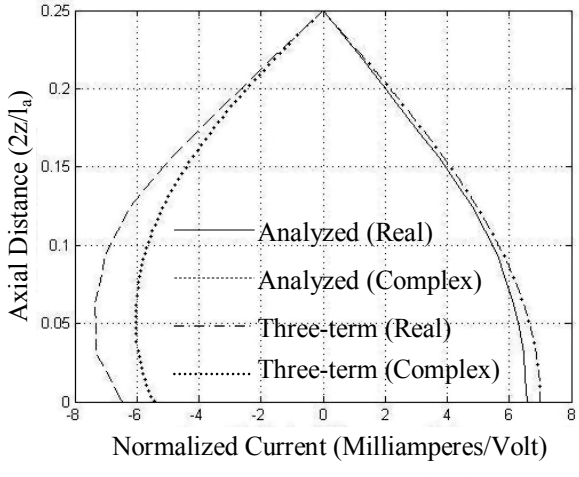


Fig. 6 Current in active dipole

The array has a reflector, an active dipole and one director of length, $l_r = 0.51\lambda$, $l_a = 0.5\lambda$ and $l_d = 0.4\lambda$ respectively and radius of each element is 0.00337λ. The spacing between reflector and active dipole is 0.25λ and between director and active dipole is 0.3λ.

B. Input Impedances of antennas

The calculated input impedance of the half-wave dipole is (90.0667 + j 29.0237) Ω; where according to King's Three-term theory it is (83.3333 + j 41.6667) Ω and as per Mack's measurement it is (94.6746 + j 39.4477) Ω.

For full-wave dipole, the calculated input impedance is (251.8 – j 449.99) Ω; where according to King's Three-term theory and Mack's measurement it is (506.04 – j 512.26) Ω and (337.84 – j 472.97) Ω respectively.

The input impedance of the 3-element Yagi-Uda array is found (77.2734 + j 76.1564) Ω; when as per King's Three-term theory it is (88.3281 + j 69.4) Ω.

C. Far-field Radiation Pattern

The azimuthal far-field radiation patterns of different antennas are shown in Fig. 8.

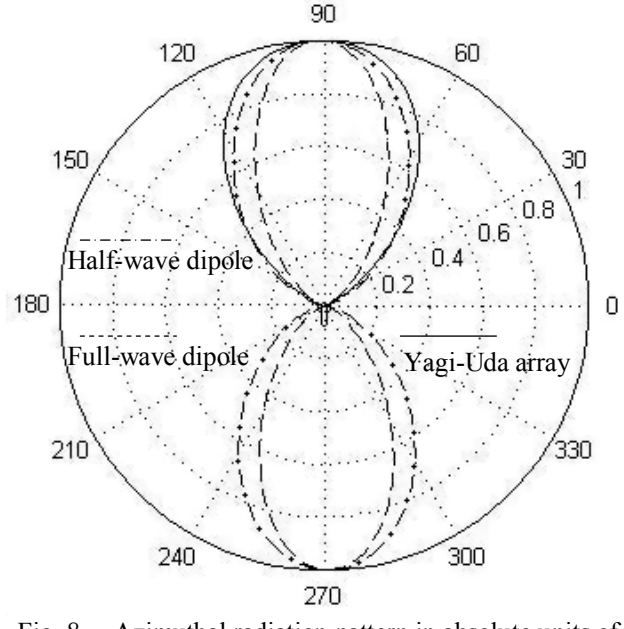


Fig. 8 Azimuthal radiation pattern in absolute units of the half-wave, full-wave and the Yagi-Uda array

The solid line, dash-dot line and dotted line represent the azimuthal radiation pattern for Yagi-Uda array, half-wave and full-wave dipole antenna respectively.

IV. Conclusion

The current distributions for a half-wave, a full-wave dipole antenna and a three-element Yagi-Uda antenna array are obtained by solving Pocklington's integral equation using wavelet bases in moment method. The determined current distribution function is verified with theoretical and experimental data and found that the accuracy is quite satisfactory. Using the current distribution, characteristics of wire antennas as input impedance, radiation pattern are analyzed and found physically significant. Moreover, the computation time is greatly reduced by the method.

References

[1] J. C. Maxwell, "A dynamical theory of the electromagnetic field," in Proc. Royal Soc. (London), vol. 13, pp. 531, 1864.

[2] H. Hertz, "Ueber sehr schnelle electrische Schwingungen," Wied. Ann., vol. 31, pp. 421, 1887.

[3] H.C. Pocklington, "Electrical oscillations in wires," in Proc. Camb. Phil. Soc. 9, pp. 324-333, 25 October 1897.

[4] L. Lorenz, "On the identity of the vibrations of light with electrical currents", Phil. Mag. 34, pp. 287-301, June 1867.

[5] P. S. Carter, "Circuit relations in radiating systems and applications to antenna problems," in Proc. IRE, vol. 20, pp. 1004-1041, June 1932.

[6] G. H. Brown, "Directional antennas," in Proc. IRE, vol. 25, pp. 79-145, January 1937.

[7] L.V. King, "On the radiation field of a perfectly conducting base-insulated cylindrical antenna over a perfectly conducting plane earth and the calculation of radiation resistance and reactance," Phil. Trans. Roy. Soc. (London), ser. A, vol. 236, pp. 381-422, 2 November 1937.

[8] E. Hallén, "Theoretical investigations into the transmitting and receiving antennae," Nova Acta Regiae Sco. Sci. Upsaliensis, Ser. 4, vol. 2, p. 1, November 1938.

[9] Ronold W. P. King and Tai Tsun Wu, "Currents, charges and near fields of cylindrical antennas," Radio Science Journal of Research NBS/UNSC-URSI, Vol. 69D, No. 3, pp. 429-446, March 1965.

[10] R. King and D. Middleton, "The cylindrical antenna: current and impedance," Quart. Appl. Math, vol. 3, pp. 302-335, 1946.

[11] R. H. Duncan and F. Hinchey, "Cylindrical antenna theory," J. Res. NBS, vol. 64D, pp. 569-584, September-October 1960.

[12] R. F. Harrington, "Matrix methods for field problems," in Proc. IEEE, Vol. 55, pp. 136-149, February 1967.

[13] R. J. Burkholder, "High-frequency asymptotic methods for analyzing the EM scattering by open-ended waveguide cavities," Ph.D dissertation, The Ohio State University, Columbus, OH, 1989.

[14] G. Beylkin, R. Coifman and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I," Comm. Pure Appl. Math, vol. 44, pp. 141-183, 1991.

[15] C. K. Chui and E. Quak, "Wavelets on a bounded interval," in Num. Math. Approx. Theory, D. Braess and L. L. Schumaker, Eds. Basel: Birkhauser Verlag, Vol. 9, pp. 53-75, 1992.

[16] Robert D. Nevels, Jaideva C. Goswami & Hooman Tehrani, "Semi-orthogonal versus orthogonal wavelet basis sets for solving integral equations," IEEE Transactions on Antennas and Propagation, Vol. 45, No. 9, pp. 1332-1339, September 1997.

[17] Y. Tretiakov and G. Pan, "Malvar wavelet based Pocklington equation solutions to thin-wire antennas and scatterers," Progress In Electromagnetics Research, PIER 47, pp. 123-133, 2004.

[18] R. B. Mack, "A study of circular arrays," Cruft Lab., Harvard University, Cambridge, Mass., Tech. Reports. 381-386, May 1963.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Stacked Multiple Slot Microstrip Patch Antenna for Wireless Communication System

Mohammad Tariqul Islam[1], Norbahiah Misran[1,2], Mohammed Nazmus Shakib[2], and Baharudin Yatim[1]

[1]Institute of Space Science (ANGKASA), [2]Dept. of Electrical, Electronic & Systems Engineering, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

Email: titareq@yahoo.com, engmdns@yahoo.com

Abstract - A novel multiple slot microstrip patch antenna for wireless communication is proposed. This paper presents stacked probe fed inverted multiple slot microstrip patch antenna. The composite effect of integrating these techniques and by introducing the novel multiple shaped patch, offer a low profile, high gain, and compact antenna element. Simulated results for main parameters such as return loss, impedance bandwidth, radiation patterns and gains are also discussed herein. The study showed maximum achievable gain of about 11.44 dBi with simplicity in designing and feeding, can well meet for wireless communication system especially for base station.

Keywords- Slotted antenna, microstrip patch antenna, probe fed.

I. Introduction

The current demand of wireless communication systems and their miniaturization, antenna design becomes more challenging in present days. Microstrip patch antennas have several well-known advantages, such as low profile, low cost, light weight, ease of fabrication and conformity [1]. However, the microstrip antenna inherently has a low gain and a narrow bandwidth. To overcome its inherent limitation of narrow impedance bandwidth and low gain, many techniques have been suggested e.g., for probe fed stacked antenna, microstrip patch antennas on electrically thick substrate, slotted patch antenna and stacked shorted patches have been proposed and investigated [2]. In general, the impedance bandwidth of a patch antenna is proportional to the antenna volume, measured in wavelengths. However, by using two stacked patches with the walls at the edges between the two patches, one can obtain enhanced impedance band width. There has recently been considerable interest in the two layer probe fed patch antenna consisting of a driven patch in the bottom and a parasitic patch [3], [4]. By stacking a parasitic patch on a Microstrip patch antenna, the antenna with high gain or wide bandwidth can be realized [5]. These characteristics of stacked microstrip antenna depend on the distance between a fed patch and a parasitic patch. When the distance is about 0.1λ (wavelength), the stacked Microstrip antenna has a wide bandwidth [5].

Recently, aperture-coupled fed stacked patch antenna [6] have been investigated and bandwidths up to 69% have been reported, however, the major drawbacks are the level of back radiation due to the use of a resonant aperture and the surface wave excitation. Other feeding techniques such as the use of L-shaped or F-shaped probes have also been proposed yielding to wide impedance bandwidths [7], [8], at the expense of increased complexity of the design and fabrication, especially of the probe. In [7], an L-probe fed stacked U-slot patch antenna was proposed with a bandwidth up to 44.4% being achieved. V-slotted rectangular microstrip antenna with a stacked patch has been shown able to achieve bandwidths as high as 47% [9].

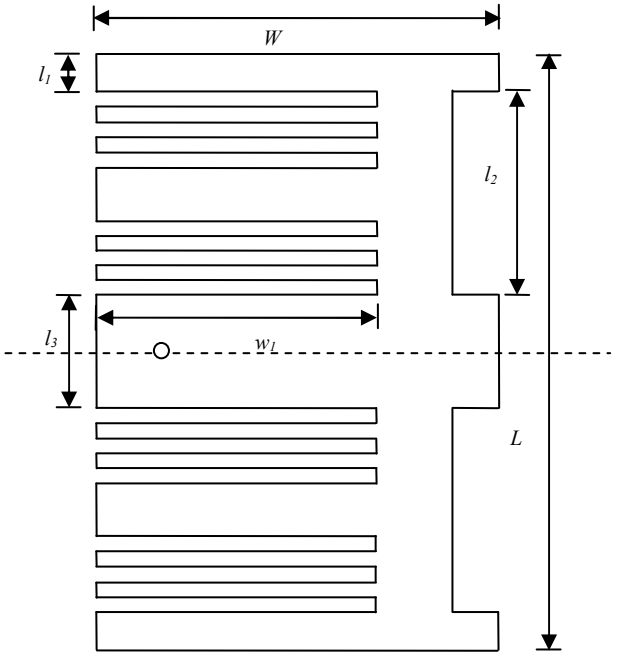
In this paper, a novel multiple slotted stacked patch antenna is investigated for enhancing the impedance bandwidth and gain. The design employs contemporary techniques namely, the probe feeding, inverted patch, and stacked multiple slotted patch techniques to meet the design requirement.

II. Antenna Layout and Structure

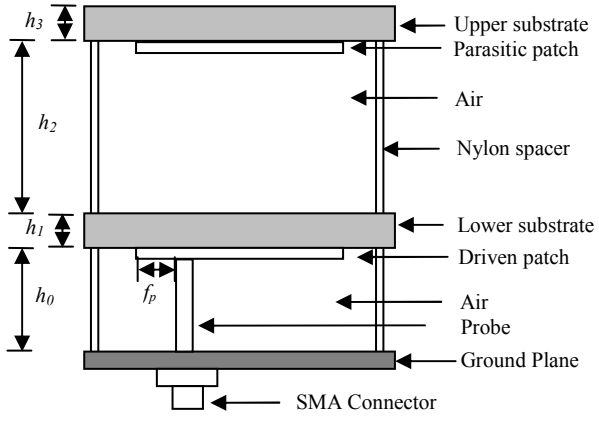
The geometry of the proposed multiple slotted stacked patch antenna structure is depicted in Fig. 1. The antenna is composed of two stacked inverted patches, two layers of air, and a vertical probe connected to the driven patch. The driven patch, with width W and length L is supported by a low dielectric substrate with dielectric permittivity ε_1 and thickness h_1. The parasitic patch with the same width and length as driven patch is stacked at the height h_2 above the lower substrate and supported by lower dielectric substrate with permittivity ε_1 and thickness h_3. Two air-filled layers are used in between the lower substrate and ground plane with permittivity ε_0 and thickness h_0 and in between two patches with permittivity ε_0 and thickness h_2 respectively.

The proposed two stacked patches are on two different radiating elements. These patches are multiple slotted and embedded in parallel on the radiating edge of the patch symmetrically with respect to the centerline (x-axis) of the patch. The multiple slots on the patch are shown in

Fig. 1(a), where, *l and w* are the length and width of the slots. The driven patch is fed by a direct connected probe along the centerline (x-axis) at a distance f_P from the edge of the patch as shown in Fig. 1(b). Table 1 shows the optimized design parameters obtained for the proposed multiple slotted patch antenna. A Rogers RT 5880 Duroid™ dielectric substrate with dielectric permittivity, ε_1 of 2.2 and thickness of 1.5748mm has been used at the lower and upper patch respectively in this research. An Aluminum plate with dimensions of $1.34\,\lambda_0 \times 1.21\,\lambda_0$ (where λ_0 is the guided wavelength of the center operation frequency) and thickness of 1 mm is used as the ground plane.



(a)



(b)

Fig. 1 The geometry of the proposed multiple slotted patch antenna. (a) Top view. (b) Side view.

The use of probe feeding technique, multiple slotted patch with thick air-filled substrates provide the bandwidth enhancement and the use of parallel slots reduce the cross polarization. The gain increases by stacking the parasitic patch. The higher gain is obtained when the height of the parasitic patch is about a half wavelength. By increasing the thickness of parasitic substrate, the gain decreases and the size of parasitic element to obtain the maximum gain becomes large.

Table 1 Proposed patch antenna design parameter.

Parameters	Value [mm]
W	53
w_1	37
L	79
l_1	5.0
l_2	27
l_3	15
h_0	12.5
h_1	1.5748
h_2	73.5
h_3	1.5748
f_P	8.5

III. Results

The resonant properties of the proposed antenna have been predicted and optimized using a frequency domain three- dimensional full wave electromagnetic field solver (Ansoft HFSS).Fig. 2 shows the simulated result of the return loss of the proposed multiple slotted antenna. The two closely excited resonant frequencies at 1.85 GHz and at 2.1 GHz as shown in the figure gives the measure of the wideband characteristic of the patch antenna. The simulated impedance bandwidth (VSWR≤2) is 19.8% from 1.82 GHz to 2.22 GHz is achieved at 10dB return loss.

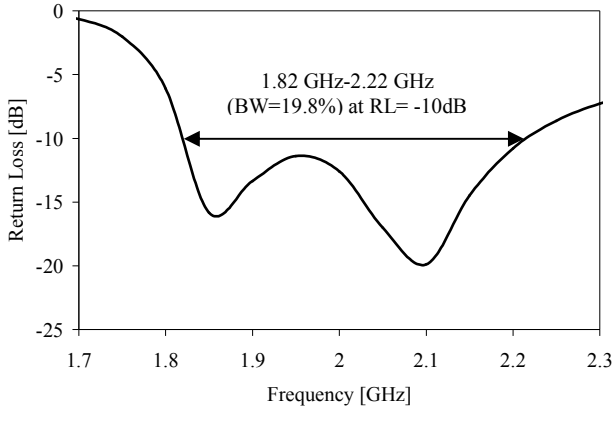


Fig. 2 Simulated return loss of the proposed multiple slotted patch antenna.

The simulated radiation pattern is at the second resonant frequency in the *xz*-plane and *yz*-plane are plotted in Fig. 3. For the sake of brevity, only simulated radiation pattern for second resonance frequency is given in this paper. As shown in Fig. 3, the designed antenna displays good broadside radiation patterns in the *xz*-plane and *yz*-plane at 2.1 GHz (second resonance). It can be seen that 3-dB beamwidth of 44º and 38º for *xz*-plane and *yz*-plane respectively at 2.1 GHz. The cross polarization pattern is lower than about -40dB in *xz*-plane. The proposed multiple slotted patch antenna exhibits better cross polarization than the design reported in [10]. Notable, the radiation characteristics of the proposed multiple slotted antenna is better to those of the conventional patch antenna.

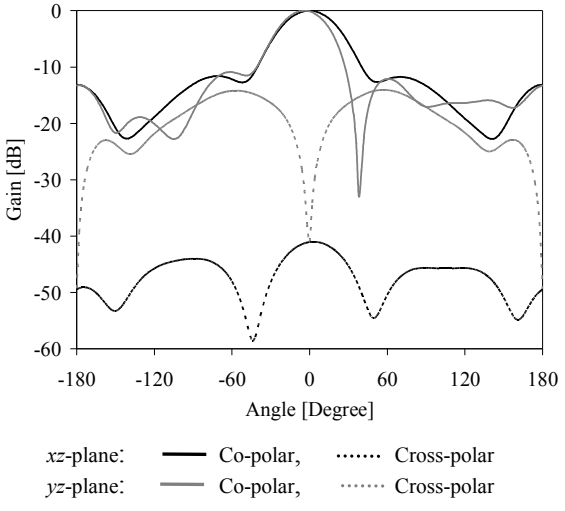


Fig. 3 Normalized radiation pattern of proposed multiple slotted patch antenna at resonance frequency 2.1 GHz for *xz*-plane

The simulated gain of the proposed multiple slotted patch antenna at various frequencies is shown in Fig. 4. As shown in the figure, the maximum achievable gain is 11.42 dBi at the frequency of 1.92GHz and the gain variation is 1.6 dBi at the operating frequency.

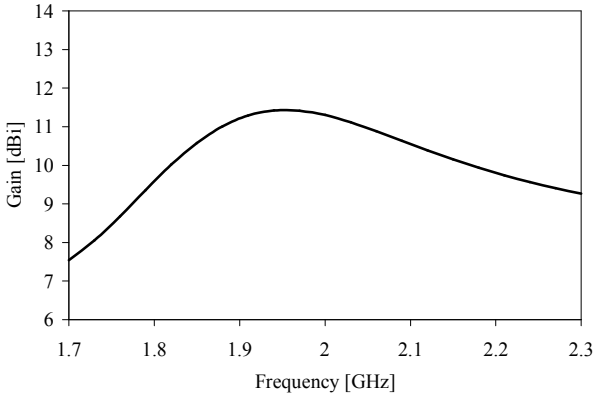


Fig. 4 Simulated gain of proposed multiple slotted patch antennas at different frequencies.

IV. Conclusion and Discussion

A multiple slotted stacked microstrip patch antenna has been designed for wireless communication system. A novel technique for enhancing gain of microstrip patch antenna is successfully designed in this research. Main parameters such as return loss, impedance bandwidth, gains and far field patterns have been studied. The results indicate that the designed antenna shows satisfactory characteristics of high gain, which well meet base station application.

Simulation results of a wideband microstrip patch antenna covering 1.82 to 2.22 GHz frequency have been presented. Techniques for microstrip broadbanding, gain enhancement, and cross polarization reduction are applied with significant improvement in the design by employing proposed multiple slotted patch shaped design, inverted patch, and probe feeding.

The proposed multiple slotted microstrip patch antenna achieves a fractional bandwidth of 19.8% (1.82 to 2.22 GHz) at 10 dB return loss. The maximum achievable gain of the antenna is 11.42 dBi. The proposed driven and parasitic patches have compact dimension of $0.532\,\lambda_0 \times 0.357\,\lambda_0$. The design has demonstrated that stacked patch with multiple slots and probe fed can be used to form an antenna with the broad bandwidth of 19.8%, furthermore due to its high gain and broad bandwidth more applications can be anticipated.

Acknowledgement

The authors would like to thank the MOSTI Secretariat, Ministry of Science, Technology and Innovation of Malaysia, e- Science fund: 01-01-02-SF0376, and Institute for Space Science UKM for sponsoring this work.

References

[1] S. L. S. Yang, A. A. Kishk, and K. F. Lee, "Frequency reconfigurable U-slot microstrip patch antenna," IEEE Antennas and Wireless Propagation Letters, vol. 7, pp. 127-129, 2008.

[2] D. M. Pozar and D. H. Schaubert, *Microstrip antennas, the analysis and design of microstrip antennas and arrays*. New York: IEEE press, 1995.

[3] K. L. Wong and W. H. Hsu, "A broadband rectangular patch antenna with a pair of wide slits," IEEE Transactions on Antennas and Propagation, vol. 49, pp. 1345-1347, September 2001.

[4] J. Y. Sze and K. L. Wong, "Slotted rectangular microstrip antenna for bandwidth enhancement," IEEE Transactions on Antennas and Propagation, vol. 48, pp. 1149-1152, August 2000.

[5] S. Egashira and E. Nishiyama, "Stacked microstrip antenna with wide bandwidth and high gain," IEEE Transactions on Antennas and Propagation, vol. 44, pp. 1533–1534, November 1996.

[6] S. D. Targonski, R. B. Waterhouse, and D. M. Pozar, "Design of wide-band aperture stacked patch microstrip antennas," IEEE Transactions on Antennas and

Propagation, vol. 46, no. 9, pp. 1245-1251, September 1998.
[7] B. L. Ooi and C. L. Lee, "Broadband air-filled stacked U-slot patch antenna," Electronics Letters, vol. 35, no. 7, pp. 515-517, April 1999.
[8] B. L. Ooi, C. L. Lee, P. S. Kooi, and S. T. Chew, "A novel F-probe fed broadband patch antenna," in IEEE Antennas and Propagation Society International Symposium, vol. 4, pp. 474-477, July 2001.
[9] G.Z. Rafi, and L. Shafai, "V-slotted rectangular microstrip antenna with a stacked patch," in IEEE International Symposium on Antennas and Propagation Society, vol. 2, pp. 264-267, June 2003.
[10] M. Tariqul Islam, N. Misran, and K. J. Ng, "A 4×1 L-probe fed Inverted Hybrid E-H Microstrip Patch Antenna Array for 3G Application," American Journal of Applied Sciences, vol. 4, no.11, pp. 897-901, 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Wave-Particle Interaction in an Unstable Plasma – Four-Particle Approach

Md. Abdul Matin, *Member IEEE*, Imtiaz Ahmed, *Member IEEE* and Rummana Matin*
Department of Electrical & Electronic Engineering,
Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
* Department of Physics,
Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
Email: amatin@eee.buet.ac.bd, imtiaz123b@gmail.com, rummanamatin@yahoo.com

Abstract - Different steps have been taken to define the wave-particle interaction in an unstable plasma over the last decades. In this paper, a four particle approach has been proposed for explaining the wave particle interaction in an unstable plasma. High speed particles cannot maintain their linear translational motion and when all the considerations come into act, then four types of particles play vital role in the propagation of instability. This paper describes the analytical formulation through necessary equations considering all the possible cases. The physical interpretation along with the energy calculation help to analyze the analytical expressions quite effectively.

I. Introduction

Wave particle interaction in an unstable plasma is of recent interest to ascertain the particle acceleration as well as the possible modes of wave propagation [1]-[2]. Different issues have come into consideration according to different types of research in wave particle interaction over the last few years [1]. Some researchers discussed the analogies between two classical models of the single-pass free electron laser dynamics and of the beam-wave plasma instability where some other inventors investigated saturation mechanisms of the explosive instability (EI) in a beam-plasma system [2]-[3].

The previous paper of the author [4] explains the formulation of electromagnetic pulses (EMP) by high speed particles. In another paper [5], the authors explore the fact that Lorentz's field accelerates and retards the particles in such a manner that energy bicones are formed around linear straight antennas and if the antenna arms are unequal the bicones are not symmetrical. In all the cases wave particle interactions in plasma media give rise to several categories of particles such that some categories of particles have speed slightly faster than that of light and the other categories have their speed slightly slower than that of light. They combinedly play vital role in the instability of plasma particle. In this paper, we applied a four particle approach to explain the plasma instability by ignoring the collision effect and by assuming that a high speed particle cannot maintain linear translational motion. Rather it is dispersed into four types of particles characterized by different types of velocities all having a natural tendency to trace back to its initial position.

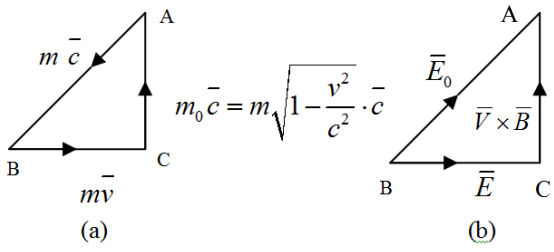
Although an equilibrium situation can be expected by momentum distribution of the particles in a right angled triangle, the physical situation is different. There may be acute angled triangle and obtuse angled triangle of momentum distribution. When all these possibilities are taken into account altogether four types of particles play vital role in the propagation of instability.

The rest of the paper is organized as follows: In Section II, the wave-particle interaction has been modeled. The physical interpretation for the wave-particle interaction has been given in Section-III. Section IV energy calculation and finally Section V finishes the paper by giving conclusion.

II. Modeling of wave particle interaction

A. Wave-Particle Interaction Associated with $\overline{V}\times\overline{B}$ perpendicular to $\overline{E}$

A particle having mass m and speed v approaching the speed of light c cannot have linear translational motion. It retraces its path, changing momentum from $m\overline{v}$ to $m_0\overline{c}$ and then to $m\overline{c}$ in a right- angled triangle shown in Fig.1. m_0 is the rest mass of the particle. The point B is assumed to be the starting point of the particle, where it returns after tracing the paths BC , CA and AB and conserves its momentum in accordance with AB+BC+CA= 0. Such behavior is common in electromagnetic pulses (EMP) in a saw-tooth wave associated with Lorentz field $\overline{E}_0 = \overline{E} + \overline{v}\times\overline{B}$ depicted in Fig.1(b). $\overline{E}_0$ is the applied electric field balanced by $\overline{E}$, the Ohmic part or loss part of the electric field and the motional electric field or the Hall field $\overline{V}\times\overline{B}$



(a) (b)

Fig.1(a). Momentum distribution of a relativistic particle moving in a right-angled - triangle.(b) The corresponding Lorentz field. $\bar{E}_0 = \bar{E} + \bar{v}\times\bar{B}$

Energy corresponding to the momentum $m\bar{c}$ of AB branch $= \frac{1}{2}mc^2$ (1)

where m is the relativistic mass

$$m = \frac{m_0 c}{\sqrt{c^2 - v^2}} \quad (2)$$

Energy of the AB branch is considered potential , because the applied electric field $\bar{E}_0$ acts along BA . According to. Kirchhoff's voltage law (KVL) the applied electric energy is in the form of a voltage or potential given by

$$V_0 = -\int \bar{E}_0 . d\bar{l} \quad (3)$$

which drops off as

$$V_0 = V_R + jV_L \quad (4)$$

where V_R represents a resistive drop or Ohmic drop due to $\bar{E}$ drifting through the conducting medium and V_L represents a reactive drop arising from the Hall e.m.f. $\bar{V}\times\bar{B}$ along CA, $j = \sqrt{-1}$ is electrical engineering j indicating phase difference of 90^0 between V_R and V_L. V_0 is assumed to maintain a constant potential energy within $\frac{1}{2}mc^2$ and to deliver another $\frac{1}{2}mc^2$ into the circuit, so that the total energy of the circuit remains within $E=mc^2$. The following calculations make it clear.

Energy corresponding to the momentum $m_0\bar{c}$ of CA branch

$$= \frac{1}{2}m(c^2 - v^2) \quad (5)$$

According to KVL it corresponds to an inductive or capacitive storage of energy.

Energy of BC or $m\bar{v}$ branch $= \frac{1}{2}mv^2$ (6)

This energy is considered kinetic, because it refers to Ohmic loss or dissipation.

Summarizing the above results we get

total potential energy (P.E).of AB branch = $\frac{1}{2}mc^2$ (7)

Total kinetic energy (K. E). of BC and CA branches

$$= \frac{1}{2}m(c^2 - v^2) + \frac{1}{2}mv^2 = \frac{1}{2}mc^2 \quad (8)$$

Summing up the total P.E. and the total K.E., the total energy is given by

$$E=mc^2 \quad (9)$$

This is Einstein's energy relation, confirmed by our recent study[. Now let us see whether some different phenomena arise with cases where $\bar{V}\times\bar{B}$ does not maintain purely inductive, or capacitive storage in the CA branch. In such cases $\angle$C is either acute or obtuse. These two cases are treated separately in the next sections.

B. Wave-particle Interaction Associated with $\bar{V}\times\bar{B}$ not perpendicular to $\bar{E}$

When the applied field $\bar{E}_0$ is such that $\bar{E}$ is not perpendicular to the Hall e.m.f. $\bar{V}\times\bar{B}$, then it can be reasonably assumed that the rest mass m_0 traces the path CA at a velocity $\bar{c}$ not perpendicular to BC. In such cases the distribution of momentum of a relativistic mass m having its value different from that in eqn (2) may be assumed as shown in Fig.2.

Two cases of vector-analysis may arise: case-1 that corresponds to the momentum distribution on an acute-angled triangle and case-2 corresponding to the momentum distribution on an obtuse-angled triangle. shown in Fig.2. EMP of such triangular shape is common and time-harmonic analysis of periodic pulses are available in a large number of texts, see for example [6]. However, our approach is to study such pulses by relativistic approach. in a region where motion of a charged particle is governed by Lorentz field.

B. 1. Case-1 Analysis for $\angle$C an Acute Angle

As in Fig.2(a) $\angle$C is an acute angle,. Hence Here net electric field along BD = E – vBcosC

AD is perpendicular from A on BC

$$\cos C = \frac{AC^2 + BC^2 - AB^2}{2.BC.CA} \quad (10)$$

Net momentum along BD = $mv-m_0c$

$$= \frac{m_0^2c^2 + m^2v^2 - m^2c^2}{2.mv.m_0c} \quad (11)$$

Electric field along DA=vBsinC

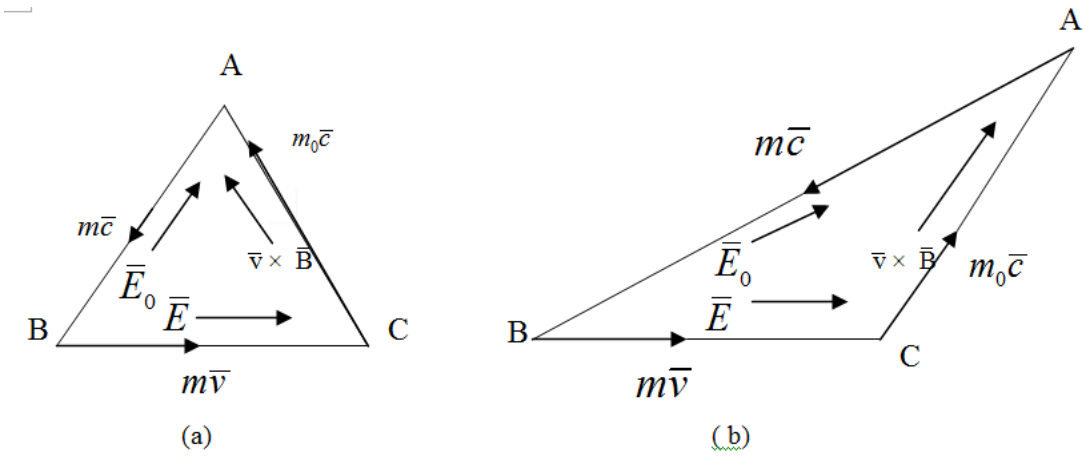


Fig. 2 Assumed momentum distribution of a charged particle moving (a) in an acute-angled triangle; (b) in an obtuse-angled triangle.

Momentum along DA= $m_0 c\sin C$

or $m^2c^2 = m_0^2c^2 + m^2v^2 - 2mm_0cv.\cos C$

$$m^2(c^2 - v^2) + 2mm_0cv.\cos C - m_0^2c^2 = 0$$

$$m = \frac{-2m_0cv\cos C \pm \sqrt{(2m_0cv\cos C)^2 + 4(c^2 - v^2)m_0^2c^2}}{2(c^2 - v^2)}$$

$$= \frac{-m_0c}{(c^2 - v^2)}\{v.\cos C \pm \sqrt{c^2 - v^2\sin C}\} \qquad (12)$$

There are two values of m say m_1 and m_2

$$m_1 = \frac{-m_0c}{(c^2 - v^2)}\{v.\cos C + \sqrt{c^2 - v^2\sin C}\} \qquad (13)$$

negative for c>v

$$m_1 = \frac{-m_0c}{(c^2 - v^2)}\{v.\cos C - \sqrt{c^2 - v^2\sin C}\} \qquad (14)$$

positive for c>v

Since m_2 is positive it represents the relativistic mass of a positron or hole, On the other hand, m_1 being negative it represents the relativistic mass of an electron.

It can be checked that

$$m_2v_1 = -m_1v_2 = m_0c \qquad (15)$$

where,

$$v_1 = \sqrt{c^2 - v^2\sin C} + v\cos C \qquad (16)$$

$$v_1 = \sqrt{c^2 - v^2\sin C} - v\cos C \qquad (17)$$

$$v_1v_2 = c^2 - v^2 \qquad (18)$$

$$m_2 - m_1 = \frac{2m_0c\sqrt{c^2 - v^2\sin C}}{c^2 - v^2} = 2ms \qquad (19)$$

where m is the relativistic mass expressed by eqn (1) and s is a scaling factor given by

$$s = \left(\frac{c^2 - v^2\sin C}{c^2 - v^2}\right)^{\frac{1}{2}} \qquad (20)$$

The scaling factor s modifies the relativistic mass 2m of a hole-electron pair. Since c>v, then s>1 until $\angle C$ is a right angle. 2ms will have greater volume or less density than that of 2m, until $\angle C = 90^0$

$$m_1m_2 = \frac{(m_0c)^2}{(c^2 - v^2)^2}\{v^2.\cos^2 C - c^2 + v^2\sin^2 C\}$$

$$= \frac{(m_0c)^2}{(c^2 - v^2)^2}\{v^2 - c^2\} \qquad (21)$$

$$= -\frac{(m_0c)^2}{(c^2 - v^2)}$$

The product is negative because m_1 is negative. Also for c>>v, $m_1m_2 \approx -m_0^2$

This forms a basis of the force of attraction between two moving bodies.

B. 2. Case-2 Analysis for $\angle$C an Obtuse Angle

Referring to Fig.2(b)

$$-\cos C = \frac{m_0^2c^2 + m^2v^2 - m^2c^2}{2.\ mv.\ m_0c}$$

Or, $m^2c^2 = m_0^2c^2 + m^2v^2 + 2mm_0cv.\cos C$

Or, $m(c^2 - v^2) - 2mm_0cv.\cos C - m_0^2c^2 = 0 \qquad (22)$

$$m = \frac{2m_0cv\cos C \pm \sqrt{(2m_0cv\cos C)^2 + 4(c^2 - v^2)m_0^2c^2}}{2(c^2 - v^2)}$$

$$= \frac{m_0c}{(c^2 - v^2)}\{v.\cos C \pm \sqrt{c^2 - v^2.\sin^2 C}\} \qquad (23)$$

There are two values of m say m_3 and m_4

$$m_3 = \frac{m_0c}{(c^2 - v^2)}\{v.\cos C + \sqrt{c^2 - v^2.\sin^2 C}\} \qquad (24)$$

positive for c>v

$$m_3 = \frac{m_0 c}{(c^2 - v^2)}\{v.\cos C - \sqrt{c^2 - v^2.\sin^2 C}\} \quad (25)$$

negative for c>v

$$m_0 c = m_3 v_2 = -m_4 v_1 = -m_1 v_2 = m_2 v_1 \quad (26)$$

Also,

$$m_3 - m_4 = m_2 - m_1 = \frac{2m_0 c\sqrt{c^2 - v^2 \sin^2 C}}{c^2 - v^2} = 2ms \quad (27)$$

III. Physical Interpretation of the Four Roots Arising from the Momentum-Balance

All the four roots of the quadratic equations observed above involving the relativistic mass m of a particle can be configured in a vector diagram shown in Fig.3 .The positive values of m i.e. m_2 and m_3 represent the relativistic masses of two positrons or holes moving on the arms of two acute-angled triangles ABC and A'BC and the negative values of m i.e. m_1 and m_4 represent the relativistic masses of two electrons moving on the arms of two complementary obtuse angled triangles A'''CB and A"CB. Electric current flow is established combinedly by all the momentum- vectors shown in Fig.3.Every positron has an image electron. But the conventional image theory does not hold good in the present case , because while a positron or hole travels on the arms of an acute- angled triangle,. its image electron travels simultaneously on a complementary obtuse-angled triangle having the same base. That means while a hole of relativistic mass m_2 travels on the arms of the acute-angled triangle ABC , its unconventional image-electron of mass m_1 travels simultaneously on the arms of the complementary obtuse- angled triangle A"BC. The concept of unconventional images in an anisotropic plasma has been studied by the author in [7]. The multidirectional flow of current is attributed to multidirectional momenta of charged particles arising from non-perpendicularity of the Hall e.m.f. $\overline{V} \times \overline{B}$ with respect to $\overline{E}$., .

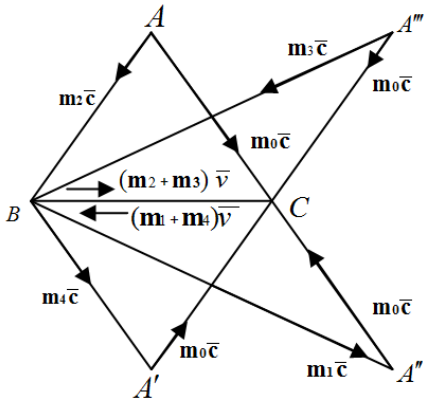


Fig.3 Momentum distribution on the arms of four triangles corresponding to the four roots of the quadratic. equations involving the relativistic mass m moving in Lorentz field, . Since m_1 and m_4 are negative, they represent electrons and since m_2 and m_3 are positive, they represent positrons or holes. The conventional image theory does not apply

IV. Energy Calculation

The conservation of energy requires that
Total positron energy E = total electron energy - E , leading to a state of equilibrium or zero configuration. Referring to triangles ABC and A'''BC of Fig.3, the applied field $\overline{E}_0$ conserves potential energy along the arms AB and A'''B.

So, total P.E. for positron = energy in AB branch due to $m_2\overline{c}$ + energy in A'''B branch due to $m_3\overline{c}$

$$= \frac{1}{2}m_2 c^2 + \frac{1}{2}m_3 c^2$$

$$= \frac{1}{2}\frac{m_0 c^3}{v_1} + \frac{1}{2}\frac{m_0 c^3}{v_2}$$

$$[\because m_2 v_1 = m_3 v_2 = m_0 c]$$

$$= \frac{1}{2}\frac{m_0 c^3}{v_1 v_2}(v_2 + v_1)$$

$$= \frac{m_0 c^3}{c^2 - v^2}\sqrt{c^2 - v^2 \sin C} = msc^2 = \frac{1}{2}m'c^2$$

where,

$$m' = \frac{2m_0 c}{\sqrt{c^2 - v^2}}.\frac{\sqrt{c^2 - v^2 \sin C}}{\sqrt{c^2 - v^2}} = 2ms \quad (28)$$

Total K. E. for positron = energy in AB branch due to $(m_2 + m_3)\overline{v}$ + Energy in AC branch due to $m_2\overline{v}_1$ + energy in A'''C branch due to $m_0\overline{c}$ or

$$= \frac{1}{2}(m_2 + m_3)v^2 + \frac{1}{2}m_2 v_1^2 + \frac{1}{2}m_3 v_2^2$$

$$= \frac{v^2}{2}.2\frac{m_0 c}{c^2 - v^2}\sqrt{c^2 - v^2 \sin^2 C} + \frac{1}{2}m_0 c(v_1 + v_2)$$

$$= \frac{v^2}{2}.\frac{2m_0 cs}{\sqrt{c^2 - v^2}} + \frac{1}{2}m_0 c.\frac{2\sqrt{c^2 - v^2 \sin^2 C}}{c^2 - v^2}c^2 - v^2 \quad (29)$$

$$= \frac{v^2}{2}m' + \frac{1}{2}.2ms(c^2 - v^2) = \frac{1}{2}m'c^2$$

So, total positron energy is given by, E =

$$P.E. + K.E. = \frac{1}{2}m'c^2 + \frac{1}{2}m'c^2 \quad (30)$$

$$= m'c^2$$

From the triangles ABC and A"BC of Fig.3 , total electron energy from m_1 and m_4 can be calculated as

$$-E = m'c^2 \quad (31)$$

Hence the total energy of a stable system is conserved by zero configuration Also m' and c defined above stand for m and c in eqn (13) confirming the consistence of classical electromagnetic wave theory with Schrödinger's wave-particle theory. Since m' = 2ms indicates the relativistic mass of a hole-electron pair ,then U_0 in eqn(13) represents appropriately the potential energy of

a hole- electron pair participating in the wave-particle energy exchange, and given by

$$U_0 = eV_0 = \frac{1}{2}m'c^2 \tag{32}$$

where e is the charge of an electron.

The scaling factor s suggests also that the basic momentum triangle ABC shown in Fig.1(a) expands when the applied field is such that the Hall field $\bar{V} \times \bar{B}$ is not perpendicular to the internal field $\bar{E}$ (as sketched in Fig.2). Fig.4 illustrates this phenomenon where an equivalent EMP is shown by the right-angled triangle A'B'C' as an expanded version of the EMP represented by the right-angled triangle ABC . This phenomenon of expansion is attributed to the additional loss arising from the non-perpendicularity of the Hall field with respect to the internal field. It explains very well the phenomenon of thermal expansion of a conducting body resulting from Ohmic dissipation.

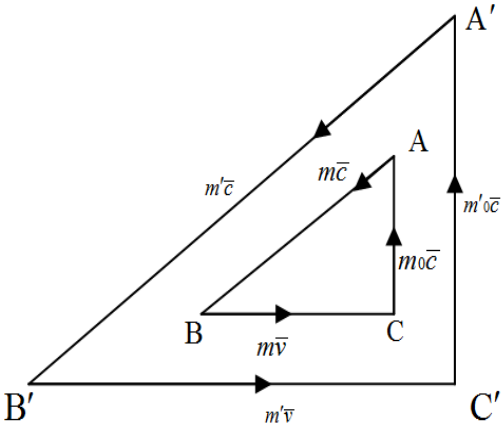


Fig. 4 Sketch of an equivalent EMP representing momentum –balance by the right-angled triangle A'B'C' which is an expanded version of the basic right-angled triangle ABC (shown primarily in Fig1(a)). $m'_0 = 2m_0 s$

Total energy for $\angle C$ an acute angle is given by E= energy in AB branch + energy in BC branch + energy in CA branch.

$$= \frac{1}{2}mc^2 + \frac{1}{2}mv_1^2 + \frac{1}{2}mv^2$$
$$= mc^2 + m\bar{v}.\bar{v}_1 \tag{33}$$

It becomes equal to the mean value of the energy or energy at equilibrium when $\angle C = 90^o$ giving E=mc^2 as the maximum energy

When $\angle$C is an obtuse angle, then

$$-\cos C = \frac{m_0^2c^2 + m^2v^2 - m^2c^2}{2.mv.m_0c}$$

$$Or,\ m^2c^2 = m_0^2c^2 + m^2v^2 + 2.mm_0cv.\cos C \tag{34}$$

$$Or, m^2(c^2 - v^2) - 2mm_0cv.\cos C - m_0^2c^2 = 0$$

$$m = \frac{2m_0cv\cos C \pm \sqrt{(2m_0cv\cos C)^2 + 4(c^2 - v^2)m_0^2c^2}}{2(c^2 - v^2)}$$
$$= \frac{m_0c}{(c^2 - v^2)}\{v.\cos C \pm \sqrt{c^2 - v^2.\sin^2 C}\} \tag{35}$$

There are two values of m say m_1 and m_2

$$m_1 = \frac{m_0c}{(c^2 - v^2)}\{v.\cos C + \sqrt{c^2 - v^2.\sin^2 C}\} \tag{36}$$

positive for c>v

$$m_1 = \frac{m_0c}{(c^2 - v^2)}\{v.\cos C - \sqrt{c^2 - v^2.\sin^2 C}\} \tag{37}$$

negative for c>v

It can be checked that

$$m_1v_2 = m_2v_1 = -m_0c \tag{38}$$

$$m_1m_2 = \frac{(m_0c)^2}{(c^2 - v^2)^2}\{v^2.\cos^2 C - c^2 + v^2\sin^2 C\}$$
$$= \frac{(m_0c)^2}{(c^2 - v^2)^2}\{v^2 - c^2\}$$
$$= -\frac{(m_0c)^2}{(c^2 - v^2)}$$

The product is negative because m_2 is negative. Also for c>>v, $m_1m_2 \approx -m_0^2$

The product yields the following identity:

$$c^2 - v^2 = \left(\sqrt{c^2 - v^2\sin^2 C} + v\cos C\right)\left(\sqrt{c^2 - v^2\sin^2 C} - v\cos C\right) \tag{39}$$

For a positive mass in the case of $\angle C$ an obtuse angle let $m_1 = m$. Hence

$$m = \frac{m_0c}{c^2 - v^2}(\sqrt{c^2 - v^2\sin^2 C} + v\cos C)$$
$$= \frac{m_0c}{v_2} \tag{40}$$

Here v_2 is the velocity of m in the CA branch

Total energy for $\angle$C an obtuse angle

E=energy in AB branch + energy in BC branch + energy in CA branch

$$= \frac{1}{2}mc^2 + \frac{1}{2}mv_2^2 + \frac{1}{2}mv^2$$
$$= mc^2 - m\bar{v}.\bar{v}_2 \tag{41}$$

It becomes equal to the mean value of the energy or energy at equilibrium when $\angle C = 90^0$ giving E=mc^2. Note that

$$v_1 - v_2 = 2v\cos C \tag{42}$$

The mass of an electron and that of a positron become equal if and only if $\angle C = 90^0$.

V. Conclusion

Wave particle interaction in plasma is a challenging field of study. Works are going on the formulation of the process of instability in a plasma medium. A four particle approach has been developed along with necessary physical interpretation here in this paper. It is found that , not only the momentum distribution of the particles in a right angled triangle has been considered but also the acute angled triangle and obtuse angled triangle of momentum distribution have been considered here to interpret the physical mechanism of the propagation of instability in a plasma medium.

References

[1] A Antoniazzi et al, "Wave-particle interaction: from plasma physics to the free-electron laser," J. Phys.: Conf. Ser. **7** pp. 143-153, 2005

[2] V. V. Krasnoselskikh et. Al., "Beam-plasma Introduction in randomly inhomogeneous plasmas and statistical properties of small amplitude Lanmuir Waves in the solar wind and electron foreshock", J. Geophys, Research Vol. 112, 2007.

[3] J. G. Dong, "Introduction of Wave-Particle Resonance in Tokomaks," August 16-18, 2007, South Western Institute of Physics, Chengdu, China.

[4] Wave particle interaction in a space where particle speed approaches the speed of the wave. By Md. Abdul Matin, Md. Zahid Hossain and Shaikh Asif Mahmood 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December, 2004, Dhaka, Bangladesh. pp. 502-505.

[5] "Energy conversion from Lorentz's field in a straight electric dipole antenna" by Md. Abdul Matin, Abdul Matin Patwari, Sekendar Ali and Rummana Matin, Journal of Electrical Engineering The Institution of Engineers, Bangladesh Vol. EE 33, No. I & II, December, 2006

[6] C. L. Hemenway, R. W. Henry and M. Caulton, "Physical Electronics," John Wiley & Sons, N. Y., 1967.

[7] M. A. Matin, K. Sawaya, T. Ishizone and Y. Mushiake, "Impedance of a monopole antenna over a ground plane and immersed in a magnetoplasma," IEEE Trans. Antennas Propagat., Vol. AP-28, pp. 332-341, 1980.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Rain Fade Analysis on Earth-Space Microwave Link in a Subtropical Region

Md. Rafiqul Islam, Md. Arafatur Rahman, SK. Eklas Hossain and Md. Saiful Azad
Kulliyyah of Engineering, International Islamic University Malaysia
Jalan Gombak, 53100 Kuala Lumpur, Malaysia
E-mail: rafiq@iiu.edu.my

Abstract

The rapid development in communication systems has forced system designer to explore higher and higher frequencies. Rain is a dominant source of attenuation at higher frequencies in tropical and subtropical regions. The knowledge of rain fade is essential in order to optimize system capacity and meet quality and reliability. This paper has predicted rain fades in a subtropical country Bangladesh. The rain intensity data are derived from forty years measured annual rainfall data. The converted rain intensity data is used to estimate rain fade for earth-space link at C, Ku and Ka-bands. The rain fade is also estimated using ITU-R recommended rain intensity and compared with those predicted based on converted data.

I. INTRODUCTION

The rapid development in communication systems has brought saturation to the most desirable frequency band (1 to 10 GHz). This fact has led to the utilization of higher frequencies extending the radio frequency spectrum into the millimeter wavelength region. Rain is a dominant source of attenuation at higher frequencies. Attenuation due to rain at frequencies above 10 GHz, mainly leads to outages that compromise the availability and quality of service, making this one of the most critical factors in satellite link design. The design of new telecommunication systems requires the knowledge of rain fade in order to optimize system capacity and meet quality and reliability criteria [1].

The intensity and distribution of rainfall greatly affects transmission quality and limits system's availability. Rain effects on microwave systems are more critical in tropical and equatorial zones, where rainfall is higher than in temperate zones. In temperate regions the rain effects becomes significant above 10 GHz, while in tropical climates in general and in equatorial climate particularly, since the rain drops are larger than in temperate climates, the incidence of rainfall on radio links becomes important for frequencies as low as about 7 GHz [5]. When designing microwave links operating above such frequencies, the major problem in link design is to determine the excess attenuation due to rainfall.

In the design of a radio communication system, one of the major concerns is to assess system unavailability also called outage time. Unavailability or outage time is the amount of time during which the system's performance will be below some threshold value or that it will not be usable. For designing a reliable system, the amount of outage time has to be kept below some objective. In microwave systems, outages can occur either due to equipment failure or it can be propagation outage. In modern systems, the equipment outage time can be made negligibly small by using standby equipment and automatic protection switching systems. However, at higher frequencies, attenuation by rain can cause an outage, which can not be easily protected. Therefore the practical way of achieving a reliable radio system at those frequencies where there is substantial rain attenuation is to design the system in such a way that the expected amount of rain outage is below some objective [2].

For a reliable communication system, unavailability time during a year has to be kept at 0.01 percent [6]. This corresponds to availability time of 99.99 percent during a year. Therefore rainfall with one-minute integration time is very important parameter to predict attenuation at 0.01% of time availability. The long term cumulative annual rainfall data are available for most of the countries of the world. This paper has presented the cumulative rainfall data collected for forty years in four different parts of Bangladesh. Using appropriate conversion model, the long-term annual rainfall data has been converted to rain intensity data. The rain intensity proposed by International Telecommunication Union (ITU-R) as well as converted data are used to predict the rain fade for earth-to-satellite at C, Ku and Ka-Bands. This paper has predicted the rain fade for earth-to-satellite links at different frequency bands operating in Bangladesh. It has also investigated the performance of Ku and Ka-Bands during rains in a subtropical region.

II. ANNUAL RAIN STATISTICS

Bangladesh has a subtropical monsoon climate characterized by wide seasonal variations in rainfall, moderately warm temperatures, and high humidity. Heavy rainfall is characteristic of Bangladesh. With the exception of the relatively dry western region of Rajshahi, where the annual rainfall is about 1600 mm, most parts of the country receive at least 2000 mm of rainfall per year. It locates at 88° to 93° longitude (East) and 20° to 27° Latitude (North). Because of its location just south of the foothills of the Himalayas, where monsoon winds turn west and northwest, Bangladesh receives the heavy average precipitation which classify into four distinct regions. They are as North-West, North-East, South-West and South-East. The annual rainfall data collected from 1928 to 1948 for four regions are shown in from Figure 1 to Figure 4 [7].

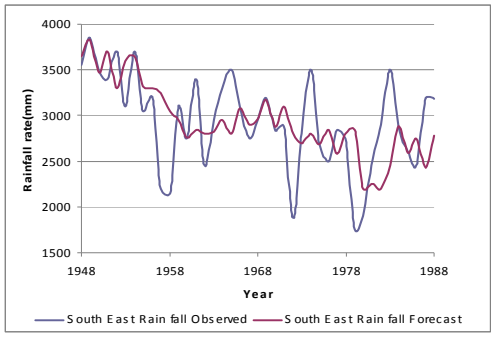


Figure 1: Variations of annual rainfall data measured and forecasted for forty years in South East region of Bangladesh.

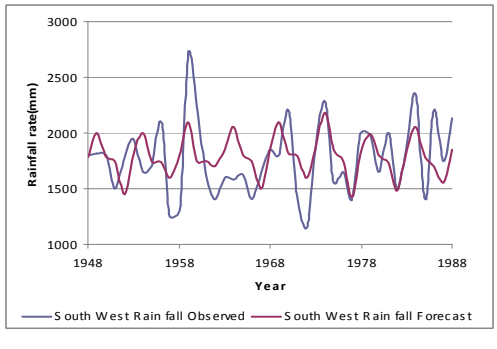


Figure 2: Variations of annual rainfall data measured and forecasted for forty years in South West region of Bangladesh.

The mean annual rainfall of North-East is 2635 mm, North-West is 1821 mm, South-East is 2937 mm and South-West is 1768 mm[7].

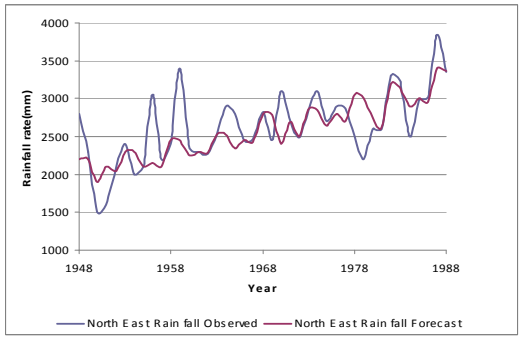


Figure 3: Variations of annual rainfall data measured and forecasted for forty years in North East region of Bangladesh.

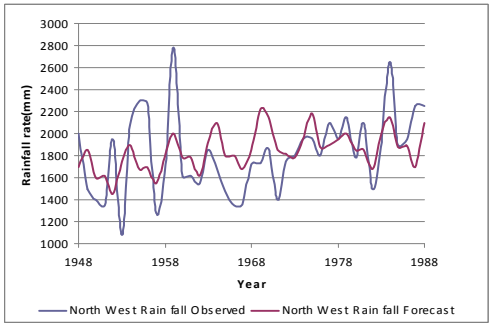


Figure 4: Variations of annual rainfall data measured and forecasted for forty years in North West region of Bangladesh.

III. CONVERSION OF RAIN STATISTICS TO RAIN INTENSITIES

Common rain attenuation prediction methods require 1-min rain rate data, which is scarce in the tropical and subtropical region [2]. However, yearly rainfall data are available at many meteorological stations. **A** method for converting the available rainfall data to the equivalent 1 min rain rate cumulative distribution (CD) would be very useful for radiowave engineers. For this reason 1 min rain rate CD can be estimated by the use of the refmed Moupfouma model and long-term mean annual rainfall data.

Several studies have shown that the Moupfouma model with refined parameters can best describe the 1 min rain rate distribution in tropical regions [3, 4]. Moupfouma found that the 1 min rain rate CD could be expressed as [5]

$$p(R \geq r) = 10^{-4}\left(\frac{R_{0.01}+1}{r+1}\right)^{b} \exp(u[R_{0.01}-r]) \qquad (1)$$

where ***r*** [mmih] represents the rain rate exceeded for a fraction of the time, and *b* is approximated by the following expression:

$$b = \left(\frac{r-R_{0.01}}{R_{0.01}}\right)\ln\left(1+\frac{r}{R_{0.01}}\right) \qquad (2)$$

The parameter *U* in eqn. 1 governs the slope of rain rate CD, and depends on the local climatic conditions and geographical features. For tropical localities

$$u = \frac{4\ln 10}{R_{0.01}}\exp\left(-\lambda\left[\frac{r}{R_{0.01}}\right]^{\gamma}\right) \qquad (3)$$

where *and* are positive constants. Based on the measured **1**min rain rate CD at several locations in Malaysia, Singapore and Indonesia [3], it was found that in tropical regions the best values for the parameters λ and γ are given in table 1:

Table 1: Parameters λ and γ

	λ	γ
M < 3000 mm	0.707	0.060
M > 3000 mm	0.398	-0.125

Thus, the Moupfouma model requires three parameters λ, γ and $R_{0.01}$. M is the mean annual rainfall. The first two parameters are easily determined from Table 1. To estimate $R_{0.01}$,it is suggested that it be derived from the value of M at the location of interest.

Several techniques have been described for the estimation of $R_{0.01}$ from the long-term mean annual rainfall *M.* These include the Morita model, Hosoya *et al.* model, Ajayi *et al.* model, Tropical India regression model and Chebil model [3,4]. All these five models use the power law relationship

$$R_{0.01} = \alpha M^{\beta} \qquad (4)$$

where α and β are regression coefficients. Chebil has made a comparison between the five models based on measured values of *M* in Malaysia, Indonesia, Singapore, Brazil and Vietnam. He showed that his model is the best estimate of the measured data [3,4]. In Chebil model the regression coefficients α and β are defined as [3]

$$\alpha = 12.2903 \text{ and } \beta = 0.2973 \qquad (5)$$

Using Chebil model, long-term mean annual rainfall data has been converted to 1 min rain rate data and are presented in Table 2.

Table 2: Measured mean annual rainfall and converted corresponding rain intensity at four regions in Bangladesh.

Name of Region	Measured Mean Annual Rainfall, mm	Converted Rainfall rate $R_{0.01}$,mm/hr
South East	2937	132
South West	1768	114
North East	2635	128
North West	1821	115
ITU Map [6]	-	95

The highest rain intensity is observed at south east region of Bangladesh at 132 mm/hr and lowest at south west is 113 mm/hr. The rain intensity recommended by ITU-R map is found 95 mm/hr for Bangladesh[6] which is far lower than converted rain intensity from measured long term annual rainfall.

IV RAIN ATTENUATION STATISTICS FROM RAINFALL RATE

The following procedure provides estimates of the long-term statistics of the slant-path rain attenuation at a given location for frequencies up to 55 GHz. The following parameters are required:

$R_{0.01}$:point rainfall rate for the location for 0.01% of an average year (mm/h)

h_s :height above mean sea level of the earth station (km)

θ :elevation angle (degrees)

φ :latitude of the earth station (degrees)

f: frequency (GHz)

R_e :effective radius of the Earth (8 500 km).

The geometry is illustrated in Figure 5 where
A = frozen precipitation
B = rain height
C = liquid precipitation

D = Earth-space path

Step 1: Determine the rain height, h_R, as given in Recommendation ITU-R P.839.

Step 2: For $\theta \geq 5°$ compute the slant-path length, L_s, below the rain height from:

$$L_S = \frac{(h_R - h_S)}{\sin\theta} \quad \text{km} \tag{6}$$

For $\theta < 5°$, the following formula is used:

$$L_s = \frac{2(h_R - h_s)}{\left(\sin^2\theta + \frac{2(h_R - h_s)}{R_e}\right)^{1/2} + \sin\theta} \quad \text{km} \tag{7}$$

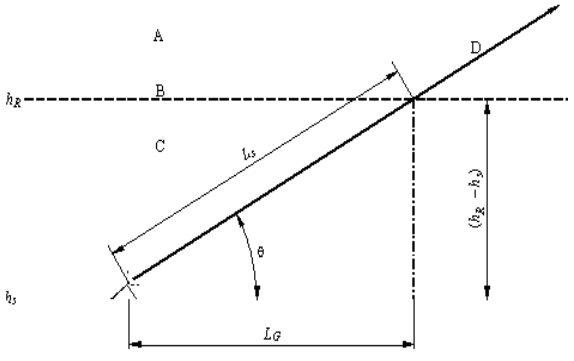


Figure 5: Schematic presentation of an earth-space path given the parameters to be input to the attenuation prediction process.

If $h_R - h_s$ is less than or equal to zero, the predicted rain attenuation for any time percentage is zero and the following steps are not required.

Step 3: Calculate the horizontal projection, L_G, of the slant-path length from:

$$L_G = L_s \cos\theta \quad \text{km} \tag{8}$$

Step 4: Obtain the rainfall rate, $R_{0.01}$, exceeded for 0.01% of an average year (with an integration time of 1 min). If this long-term statistic cannot be obtained from local data sources, an estimate can be obtained from the maps of rainfall rate given in Recommendation ITU-R P.837. If $R_{0.01}$ is equal to zero, the predicted rain attenuation is zero for any time percentage and the following steps are not required.

Step 5: Obtain the specific attenuation, γ_R, using the frequency-dependent coefficients given in Recommendation ITU-R P.838 and the rainfall rate, $R_{0.01}$, determined from Step 4, by using:

$$\gamma_R = k\,(R_{0.01})^{\alpha} \quad \text{dB/km} \tag{9}$$

Step 6: Calculate the horizontal reduction factor, $r_{0.01}$, for 0.01% of the time:

$$r_{0.01} = \frac{1}{1 + 0.78\sqrt{\frac{L_G\,\gamma_R}{f}} - 0.38\left(1 - e^{-2L_G}\right)} \tag{10}$$

Step 7: Calculate the vertical adjustment factor, $v_{0.01}$, for 0.01% of the time:

$$\zeta = \tan^{-1}\left(\frac{h_R - h_s}{L_G\,r_{0.01}}\right) \quad \text{degrees}$$

For $\zeta > \theta$, $L_R = \dfrac{L_G\,r_{0.01}}{\cos\theta}$ km

Else, $L_R = \dfrac{(h_R - h_s)}{\sin\theta}$ km

If $|\varphi| < 36°$, $\chi = 36 - |\varphi|$ degrees

Else, $\chi = 0$ degrees

$$\nu_{0.01} = \frac{1}{1 + \sqrt{\sin\theta}\left(31\left(1 - e^{-(\theta/(1+\chi))}\right)\frac{\sqrt{L_R\,\gamma_R}}{f^2} - 0.45\right)}$$

Step 8: The effective path length is:

$$L_E = L_R\,\nu_{0.01} \quad \text{km} \tag{11}$$

Step 9: The predicted attenuation exceeded for 0.01% of an average year is obtained from:

$$A_{0.01} = \gamma_R\,L_E \quad \text{dB} \tag{12}$$

Step 10: The estimated attenuation to be exceeded for other percentages of an average year, in the range 0.001% to 5%, is determined from the attenuation to be exceeded for 0.01% for an average year:

This method provides an estimate of the long-term statistics of attenuation due to rain. Considering Singapore Satellite ST1 located at 88 degree East longitude and earth station at Dhaka at 90° longitude (East) and 24° Latitude (North), the elevation angle is 61.8^0. Using above methodology from step1 to step10

and the rain rates derived in Table 2, rain fades have been estimated for all four regions as well as those recommended by ITU-R and presented in from Fig. 6 to Fig. 9. All cases the signals are assumed as vertically polarized and the regression coefficients are given as table Table 3.

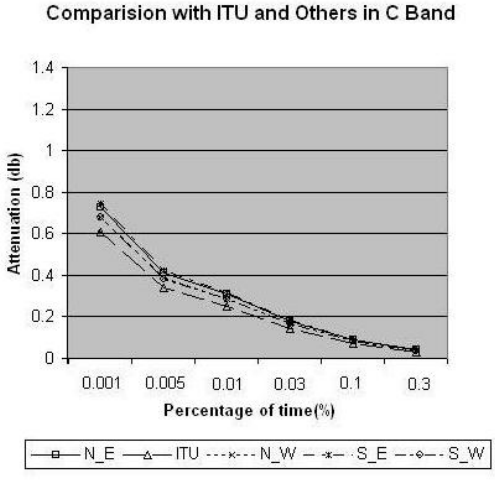


Figure 6: Variations of predicted rain fade for C-Band at different percentage of time of the year.

At C-band, attenuation due to rain is predicted 0.2 dB for 0.01% of the time of the year. But the same at Ku-band and Ka-band are predicted as 10 dB and 30 dB respectively. The difference between C and Ku-band rain fade is almost 10 dB and C and Ka-band is about 30 dB. for the design of earth-to-satellite microwave link with 99.99% reliability. Hence to design reliable earth-to-satellite microwave link is very critical at Ku and Ka-bands and needs careful and accurate estimation of rain attenuation in Bangladesh.

All three bands, rain attenuation are estimated based on ITU-R recommended rain rate as well as converted rain rate from long term measured data. It is obvious that ITU-R predicted rain attenuation are lower than those predicted using measured rain rate in all cases. The differences are 0.1 dB at C-band, 3 dB at Ku-band and 5 dB at Ka-band. At C-band, attenuation due to rain is predicted 0.3 dB for 0.01% of the time of the year. But the same at Ku-band and Ka-band are predicted as 13.8 dB and 39.2 dB respectively as shown in Table 4.

The difference between C and Ku-band rain fade is almost 13.5 dB and C and Ka-band is about 38.8 dB for the design of earth-to-satellite microwave link with 99.99% reliability. Hence to design reliable earth-to-satellite microwave link is very critical at Ku and Ka-bands and needs careful and accurate estimation of rain attenuation in Bangladesh.It is observed that ITU-R

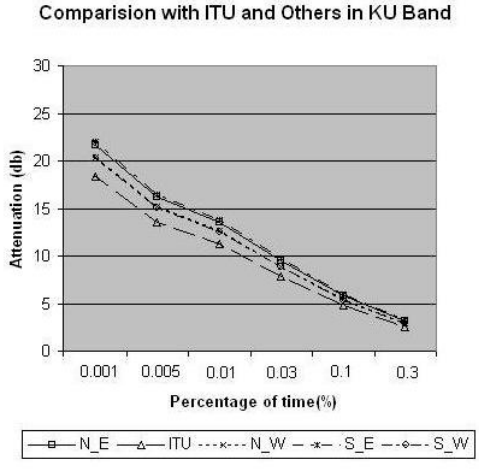


Figure 7: Variations of predicted rain fade for Ku-Band at different percentage of time of the year.

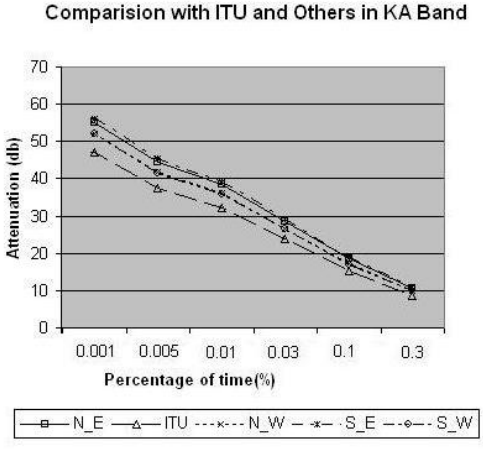


Figure 8: Variations of predicted rain fade for Ka-Band at different percentage of time of the year.

recommendation underestimates the rain rate measured in Bangladesh and consequently the link budget estimation introduces significant errors specially at Ku and Ka-bands. But the measured rain rate data is converted from measured long term annual rainfall data, it is too preliminary to comment correctly. Therefore it is recommended to measure rain intensity and rain drop

size distribution urgently for the design of reliable microwave link in Bangladesh.

Table 3: Regression Coefficients for Estimating Specification Attenuation

vertical	C-band (6GHz)	Ku-band (14GHz)	Ka-band (20GHz)
α	1.265	1.128	1.065
k	0.00155	0.0168	0.0691

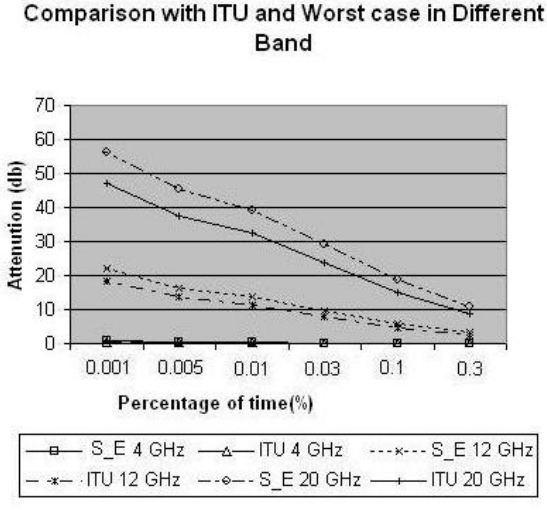


Figure 9: Variations of predicted rain fade for all three frequency bands at different percentage of time of the year.

Table 4: Estimated rain fades for all three frequency bands.

	4 GHz	12 GHz	20 GHz
$A_{0.01\%}$	dB	dB	dB
ITU-R	0.25	11.3	32.3
Measured	0.31	13.8	39.2

V. CONCLUSION

Rain is a dominant source of attenuation at higher frequencies in tropical and subtropical regions. Therefore accurate estimation of rain fade is very essential in order to design reliable microwave links in such regions. This paper has predicted rain fades in a subtropical country Bangladesh using rain intensity data derived from forty years measured annual rainfall data. It is observed that rain fade is not significant at C-band, but is very critical at Ku and Ka-bands. From observation, the estimated rain fade for earth-to-space is 10 dB at Ku-band and 30 dB at Ka-band which is really challenge to use these bands in such a subtropical country. The rain fade is also estimated using ITU-R recommended rain intensity and found significantly lower than those predicted based on converted data. Hence it is highly recommended to measure rain intensities for the design of reliable microwave links at higher frequency bands.

ACKNOWLEDGEMENT

The authors wish to acknowledge the support from the International Islamic University Malaysia - Research Management Centre by funding this research through IIUM/504/RES/G/14/3/07/EDW B0803-96.

REFERENCES

[1] ITU-R P.618-8, "Propagation data and prediction methods required for the design of Earth-space telecommunication systems", ITU, Geneva, Switzerland, 2003.

[2] ONG, J.T., and ZHU, C.N.: 'Rain rate measurements by rain gauge nerwork in Singapore', *Electron. Lett.*. 1997, 33, (3), pp. 240-242.

[3] CHEBIL, J., and RAHMAN, T.A.: 'Rain rate statistical conversion for the prediction of rain attenuation in Malaysia', *Electron. Lett.,* 1999, 35**,** (12), pp. 1019-1021.

[4] J. Chebil and **T.A.** Rahman, "Development of 1 min rain rate contour maps for microwave applications in Malaysian Peninsula", ELECTRONICS LETTERS 30th September I999 Vol. 35 No. 20, pp.1172-1174.

[5] MOUPFOUMA, F., and MARTIN, L.: 'Modelling of the rainfall rate cumulative distribution for the design of satellite and terrestrial communication systems', *Int. J. Sat. Commun.,* 1995, **13,** pp. 105- 115.

[6] ITU-R, P. RECOMMENDATION ITU-R P.837-4, "Characteristics of precipitation for propagation modeling", ITU, Geneva, Switzerland, 2003.

[7]http://www.sdnpbd.org/sdi/international_days/wed/2006/bangladesh/image/rainfall.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Design, Simulation and Fabrication of a Microstrip Patch Antenna for Dual Band Application

Md. Fokhrul Islam, M. A. Mohd. Ali, B. Y. Majlis and N. Misran

Department of Electrical, Electronic and System Engineering
Faculty of Engineering
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, MALAYSIA
Email: engr_tutul96@yahoo.com

Abstract - There is an increasing demand for newer microwave and millimeter-wave systems to meet the emerging telecommunication challenges with respect to size, performance and cost. Microstrip antennas offer the advantages of thin profile, light weight, low cost, ease of fabrication and compatibility with integrated circuitry. This paper presents a coaxilly-fed single-layer compact microstrip patch antenna for achieving dual-polarized radiation suitable for applications in the IEEE Radar Band C and X. Simultaneous use of both frequencies should dramatically improve data collection and knowledge of the targets in an airborne synthetic aperture radar system. The designed antenna consists of three rectangular patches which are overlapped along their diagonals. The design and simulation of the antenna were performed using 3D full wave electromagnetic simulator IE3D. The antenna with a bandwidth of VSWR<2 reaches 154MHz (f_0=6.83GHz) and 209MHz (f_0=9.73GHz) was designed and simulated successfully.

I. Introduction

Remote Sensing is the general science of gathering data and information about features, objects and classes on the Earth's land surface, oceans and atmosphere from sensors located beyond the immediate vicinity of such source. One such sensor that has captured the interest of the scientific community is the Synthetic Aperture Radar (SAR) [1]. It is capable of producing high-resolution imagery in microwave bands by using a special processing technique that synthesises a very long antenna aperture, thus the name synthetic aperture. Microwave frequencies are preferred as it can penetrate clouds; certain wavelength can even penetrate forest canopy. SAR sensors are active sensors, thus day and night operation is possible. SAR is usually carried on board satellites or aircrafts, as it requires relative motion between the sensor and the surface being imaged.

The radar group at the Goodyear research facility in Litchfield, Arizona is credited with building the first airborne SAR, back in 1953. It operated at 930MHz using a Yagi antenna with a very wide beamwidth (100^0). Subsequently many more airborne SAR systems was developed, notable among them are the AIRSAR by Jet Propulsion Laboratory (JPL), E-SAR by German Aerospace Research Establishment (DLR), C/XSAR [2] by Canadian Centre for Remote Sensing (CCRS) and EMISAR by Danish Centre for Remote Sensing (DCRS). These airborne SAR systems employed many types of antennas, ranging from Yagi, slotted-waveguide to microstrip. However, modern civilian SAR system generally operates in L-, C- and X-band, where microstrip antenna dominates [3].

Recently, dual-band and dual-polarized antennas have been studied using different techniques for satellite and wireless communication applications [4-7]. In particular, since microstrip antennas have attractive features such as low profile, light weight, and easy fabrication [8], the antennas are widely used to satisfy demands for polarization diversity and dual-frequency. The work described in this paper focuses on the design of a C-band and X-band dual-polarized synthetic aperture radar (SAR) antenna sharing the same physical aperture.

II. Antenna Design and Operating Principle

The basic configuration of the proposed patch antenna for exciting dual-band dual-polarization is illustrated in Fig. 1. Three rectangular patches are overlapped along their diagonals. The dimensions of the patches are $(W \times L)$ mm^2. S_1 and S_2 indicate the overlapping dimensions of the patches. The radiating patch is fed by a coaxial probe type feed in this design. As shown from Fig. 1, the inner conductor of the coaxial connector extends through the dielectric and is connected to the radiating patch, while the outer conductor is connected to the ground plane. The main advantage of this type of feeding scheme is that the feed can be placed at any desired location inside the patch in order to match with its input impedance. This feed method is easy to fabricate and has low spurious radiation. The center of the patch is taken as the origin and the feed point location is given by the co-ordinates (X_f,Y_f) from the origin. The feed point must be located at that point on the patch, where the input impedance is 50 ohms for the resonant frequency. Hence, a trial and error method was used to locate the feed point. For different locations of the feed point, the return loss (R.L) was compared and that feed point was selected where the R.L was most negative. The structure has three different resonant lengths as follows:

$$l_1 = W + (W - S_2) + 2\Delta l_1 \quad (1)$$

$$l_2 = L + (L - S_1) + 2\Delta l_2 \quad (2)$$

$$l_3 = W + L - (L - S_1) + 2\Delta l_3 \quad (3)$$

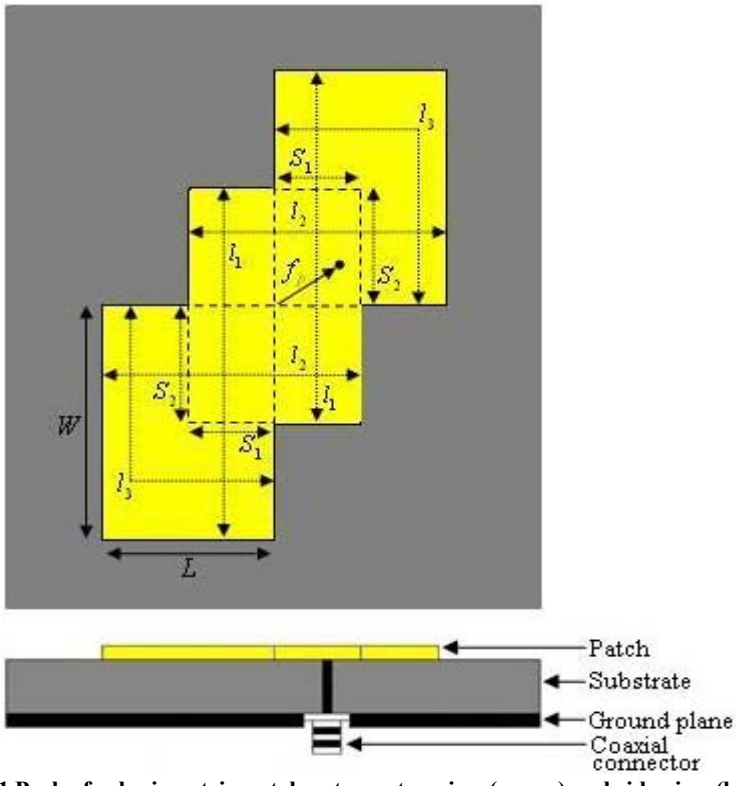


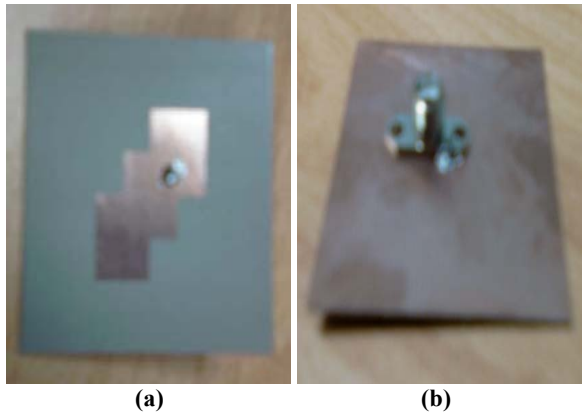
Fig. 1 Probe feed microstrip patch antenna top view (upper) and side view (lower)

The increments to the lengths, Δl_1, Δl_2 and Δl_3 are due to the fringing fields and can be computed from

$$\Delta l = 0.412h \frac{\left(\varepsilon_{reff} + 0.3\right)\left(\frac{W}{h} + 0.264\right)}{\left(\varepsilon_{reff} - 0.258\right)\left(\frac{W}{h} + 0.8\right)} \quad (4)$$

As an example, an antenna with the following dimensions was designed: three rectangular patches of dimension $(13\times 9)\,\text{mm}^2$ with overlapping dimensions $S_1 = 4.5$ mm and $S_2 = 6.5$ mm; a dielectric substrate of relative permittivity $\varepsilon_r = 2.2$ and thickness $h = 1.58$ mm and the feed location $f_p = 4.43$ mm was used.

In the extreme edges, it illustrates the curved paths along the mean dimension l_3 and thus confirms the corresponding resonant frequencies given by equation (3). The dual-polarized behavior is explained as follows: At resonance frequencies f_1 and f_2, the antenna has two radiating strips perpendicular to each other, which radiate in vertical and horizontal polarizations (Fig. 1). At resonance frequency f_3, the radiating strip has a bend and its radiation is due to two perpendicular edges, which provides dual polarization [9]. Fig. 2 shows the photograph of the fabricated microstrip patch with SMA connector. The patch is fabricated on Rogers RT 5880 with dielectric constant of 2.2 and with thickness of 1.5748mm. IE3D *em* CAD simulator is used to optimize the design parameters and the antenna is measured using the VNA-40GHz.



(a) **(b)**

Fig. 2 Photo of the fabricated antenna: (a) top view, (b) backside view

III. Results and Discussion

The results tabulated in Table 1were obtained after varying the feed location along the diagonal length of the patch from the origin (center of patch) to its right most edge. The coaxial probe feed used was designed to have a radius of 0.5mm. A frequency range of 6–12 GHz was selected and 241 frequency points were selected over this range to obtain accurate results. Table 1 shows the calculated results for different feed locations. Fig. 3 shows the return loss plots for some of the feed point locations. The center frequency is selected as the one at which the return loss is minimum. As described in Section II, the bandwidth can be calculated from the return loss (RL) plot. The bandwidth of the antenna can be said to be those range of frequencies over which the RL is greater than -9.5 dB (-9.5 dB corresponds to a VSWR of 2 which is an acceptable figure).

Table 1 Effect of feed point on center frequency, return loss and bandwidth

No.	Feed Location (X_f,Y_f)	C-Band			X-Band		
		Center Frequency (GHz)	Return Loss (dB)	Bandwidth (MHz)	Center Frequency (GHz)	Return Loss (dB)	Bandwidth (MHz)
1	(2.0,2.0)	6.92	-9.31	-	9.69	-15.32	141
2	(2.5,2.5)	6.90	-11.96	143	9.70	-23.21	183
3	(3.0,3.0)	6.85	-18.49	162	9.73	-30.01	203
4	**(3.25,3.25)**	**6.83**	**-24.13**	**154**	**9.73**	**-23.86**	**209**
5	(3.5,3.5)	6.83	-23.53	136	9.74	-21.90	213
6	(3.75,3.75)	6.80	-14.55	104	9.75	-21.22	217
7	(4.0,4.0)	6.79	-10.30	39	9.75	-20.92	220

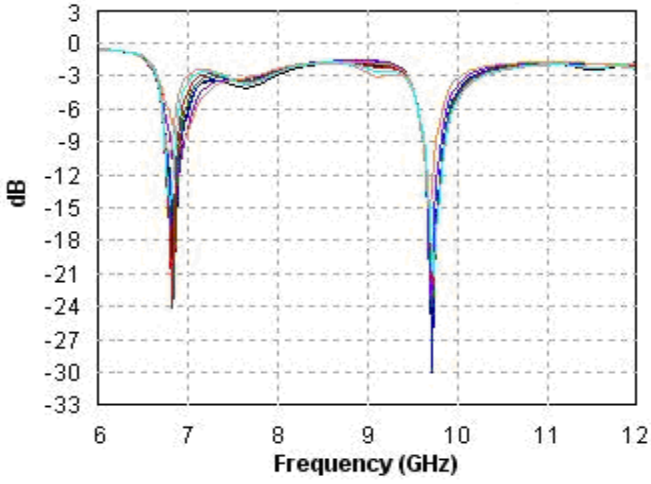


Fig. 3 Return loss for feed located at different locations

From Table 1, the optimum feed point is found to be at (X_f,Y_f) = (3.25,3.25) where the RL of -24.13 dB and -23.86 dB are obtained for C-band and X-band respectively. The bandwidth of the antenna for this feed point location is calculated to be 154 MHz (f_0=6.83GHz) and 209 MHz (f_0=9.73). It is observed from Table 1 that, as the feed point location is moved away from the center of the patch, the center frequency starts to dicrease in the low frequency whereas slightly increase in the high frequency. It is also seen that though the maximum return loss is obtained at (X_f,Y_f) = (3.0,3.0), the maximum bandwidth is obtained at (X_f,Y_f) = (4.0,4.0). The measurement result of the return loss of the antenna is shown in Fig. 4.

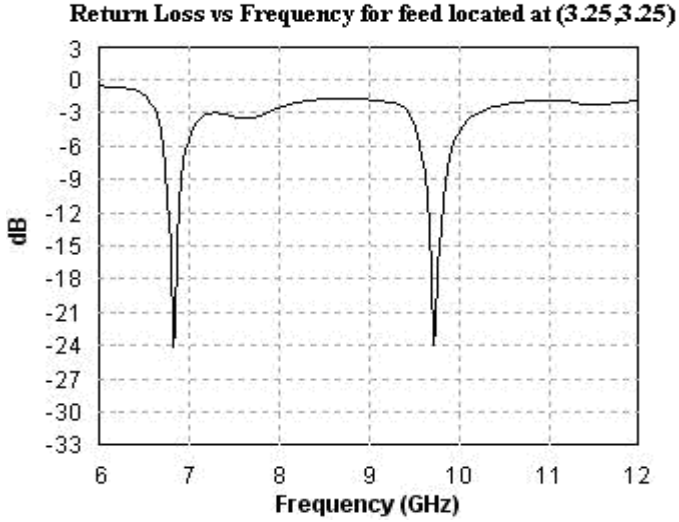


Fig. 4 Return loss for feed located at (3.25, 3.25)

IV. Conclusion

A dual-band dual-polarized coaxially-fed single-layer microstrip patch antenna with a compact structure has been demonstrated and numerically studied. The proposed antenna with a bandwidth of 154MHz (f_0=6.83GHz) and 209MHz (f_0=9.73GHz) was designed and simulated successfully. From the results obtained, it can be seen that this novel antenna is capable of satisfying some of the requirements of an airborne SAR, for example, the radiation pattern and bandwidth requirements. In additions, For high data collections and accurate knowledge of targets missions which could need several frequency bands, this concept allows to share between C and X bands SAR system. Beside this, the concept can be simply adapted to design other antenna operating at different frequency band.

Acknowledgement

The authors would like to thank the Ministry of Science, Technology and Innovation (MOSTI) of Malaysia for supporting this work under the eScienceFund 03-01-02-SF0254.

References

[1] Y. K. Chan, V. C. Koo & T. S. Lim, "Conceptual Design of a High Resolution, Low Cost X-Band Airborne Synthetic Aperture Radar System", Progress In Electromagnetics Research Symposium, Beijing, China, March 26-30, 1704-1708, 2007.

[2] F. Stuhr, R. Jordan and M. Werner, "SIR-C/X-SAR: A multifaceted radar", Aerospace and Electronic Systems Magazine, IEEE, 1995.

[3] M. Cyril and L. Jerome, "Dual band dual polarized radiating subarray for synthetic aperture radar", Antennas and Propagation Society International Symposium, IEEE, 1999.

[4] Y. J. Kim, W. S. Yun and Y. J. Yoon, "Dual-frequency and dual-polarisation wideband microstrip antenna", Electron. Lett., vol. 35, no. 17, pp. 1399-1400, 1999.

[5] E. Lee, P. S. Hall and P. Gardner, "Compact dual-band dual-polarisation microstrip patch antenna", Electron. Lett., vol. 35, no. 13, pp. 1034-1036, 1999.

[6] T. W. Chiou and K. L. Wong, "Broad-band dual-polarized single microstrip patch antenna with high isolation and low cross polarization", IEEE Trans. Antennas Propagat., vol. 50, no. 3, pp. 399-401, 2002.

[7] D. H. Choi and S. O. Park, "Dual-Band and Dual-Polarization Patch Antenna with High Isolation Characteristic", Proceedings of Asia-Pacific Microwave Conference, 2006.

[8] C. A. Balanis, Antenna Theory: Analysis and Design. New York: Wiley, 1997.

[9] K. Rambabu, M. Alam, J. Bornemann and M. A. Stuchly, "Compact Wideband Dual-Polarized Microstrip Patch Antenna", Antennas and Propagation Society International Symposium, IEEE, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

An Experimental Approach of DLC Film Deposition on Metal Substrates

Md. Mahmud Hasan[1], Muhammad Athar Uddin[2, 3], S. M. Mominuzzaman[2]

[1]Department of Electrical and Computer Engineering, University of Waterloo
Waterloo, N2L 3G1, ON, Canada
[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET)
Dhaka 1000, Bangladesh
[3]Department of Electrical & Electronic Engineering, International Islamic University Chittagong, Dhaka Campus,
Dhanmondi -3, Dhaka 1205, Bangladesh

E-mail: mahmud96@ieee.org, momin@eee.buet.ac.bd

Abstract – In this paper an experimental approach of depositing diamond like carbon (DLC) on metal substrates has been described. By electrolysis of 10% (by mass) camphoric solution in methanol, attempts were made to deposit DLC films on Copper (Cu) and Aluminum (Al) substrates at room temperature. Solution is prepared using camphor ($C_{10}H_{16}O$), a natural source in methanol solvent. At first we applied this approach on Cu substrate and then on Al substrate. The surface morphologies of deposited films were examined by scanning electron microscopy (SEM). A comparison between Cu and Al substrates has been also presented under this approach.

I. Introduction

Interest in depositing diamond like Carbon (DLC) films has been motivated by the properties of the materials and the demand of modern technologies. These properties include chemical inertness, high electrical resistivity, high dielectric strength, optical transparency, biological compatibility, and high thermal conductivity. These economically and technologically attractive properties have drawn almost unparalleled interest towards DLC film coatings on various materials. Many studies have been reported on the deposition of DLC films [1-8]. In [1-4], the attempts were taken to deposit DLC films on Si substrate. In [1], [3], [4], liquid phase deposition techniques were applied. Electrolytic properties of organic solutions had been used. The approaches described in [5-8], were for metal substrates. The techniques, which were described in [2], [5-8], are generally known as the vapour phase deposition techniques. These include high intensity ion beam (HIPIB) ablation, plasma enhanced chemical vapour deposition (PECVD), ion beam assisted deposition ion beam assisted deposition (IBAD) etc. Since DLC films synthesized in the liquid phase have significant scientific and technological implications, it is worth pursuing research with electrolytes. Deposition of DLC films on metal substrates in the liquid phase is seldom reported. There are experimental evidences that materials that can be deposited from the vapour phase can also be deposited in the liquid phase using electroplating techniques and vice versa [9]. Based on the observations in [10], it was suggested that camphor ($C_{10}H_{16}O$) and camphor like other precursors might be the best-suited candidates as starting materials for semiconducting carbon films for electronics applications.

In this study, by electrolysis of 10% (by mass) camphoric solution in methanol, an attempt is made to deposit DLC film on Cu substrate at room temperature. Solution is prepared using camphor, a natural source in methanol solvent. We followed the same procedure for Al substrate as well.

II. Experimental Details

A schematic diagram of the experimental set up is shown in Fig. 1. Cu or Al substrate with a size of 3.2×1.5×0.1 cm^3, have mounted on the negative electrode in this electroplating technique. The pure water, CH_3COCH_3, CH_3OH cleans the substrates successively. The average distance between the substrate and positive electrode was 1.5 cm. The DC voltage applied to between electrodes could be varied from 0 to 3500 volts. A thermometer is adjusted to the system to measure the temperature of the solution during the deposition occurs. The temperature was kept below 60 0C. A DC voltmeter is connected across the power supply to measure applied voltage. A mili-ammeter is also connected to the supply line to measure current.

The current density of substrate plays an important role in film formation from an organic solution [11]. Higher current density indicates more polarized charged particles move from solution to electrode, which may have some effect on the growth rate of film. Fig. 2 shows comparative plots of current density versus applied voltage betwenn Cu and Al substrates. The temperature of the solution was always kept below 60 ^{0}C, at this condition maximum current density (10 mA/ cm^2) was achieved at DC 1200

volt. According to Fig. 2, higher current density is observed in case of Cu than that of Al under same applied voltage.

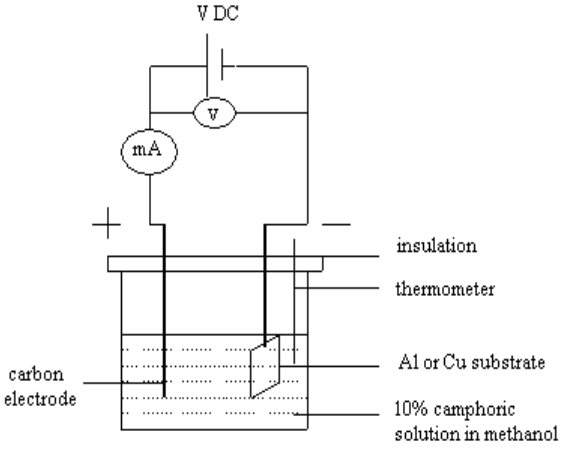


Fig. 1. Experimental set up of DLC film deposition on Cu and Al substrates.

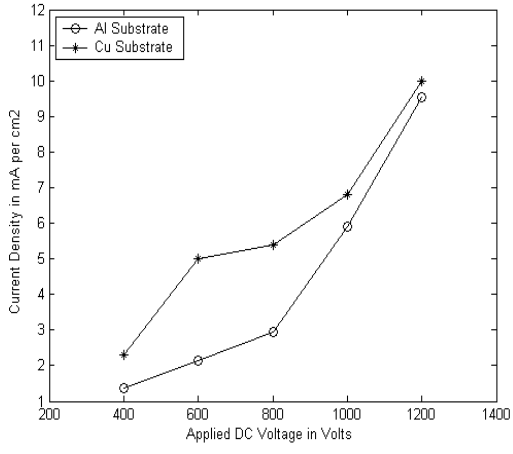


Fig. 2. Comparative plots of current density versus applied voltage for Cu and Al substrates.

III. Results and Discussions

We obtained two samples of Cu substrate. The first one was taken for 3 hours and the second one was taken for 10 hours of deposition time. The surface morphology of the deposited films was examined by scanning electron microscopy (SEM). Fig. 3 and Fig. 4 show the micrograph of front and back surface of the first sample respectively. Fig. 5 and Fig. 6 show the micrograph front surface and cross section of the second sample respectively. It is clear that the micrograph of front surface (see Fig. 3) and that of back surface (see Fig. 4) are totally different. The films were deposited only on front surface of Cu substrate. The micrograph of cross section (see Fig. 6) shows almost no film. However it was taken under deposition process for 10 hours. Let us observe the micrographs of the front surfaces shown in Fig. 3 and Fig. 5. The density of the deposited films in Fig. 5 (deposition time 10 hours) is higher than that of in Fig. 3 (deposition time 3 hours). It reveals that deposition density increased with increasing deposition time. According to the micrographs (see Fig. 3 and Fig.5), it can be seen that film is composed of small, compact grains.

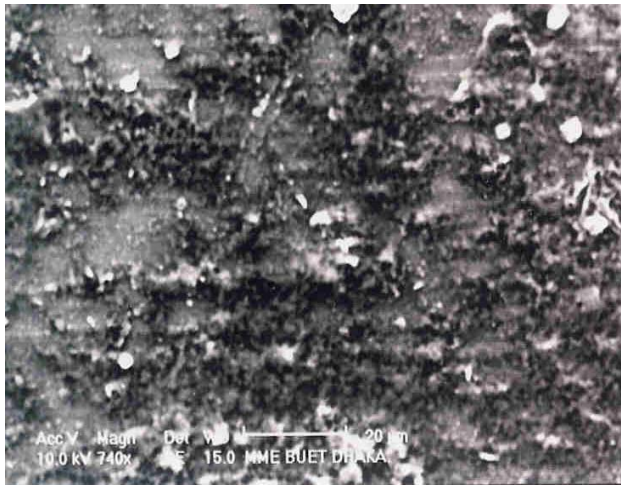


Fig. 3. SEM micrograph of the front surface of Cu substrate with 3 hours deposition.

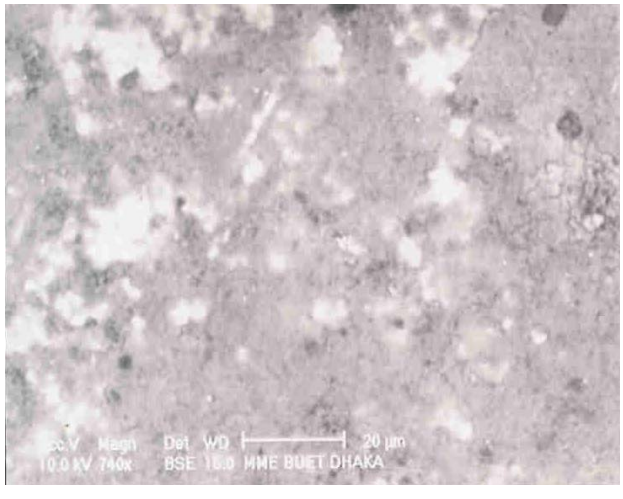


Fig. 4. SEM micrograph of the back surface of Cu substrate with 3 hours deposition.

Fig. 5. SEM micrograph of the front surface of Cu substrate with 10 hours deposition.

During the deposition of film, the carbon atoms might combine at the surface to form all possible combination of sp^1, sp^2, and sp^3 bonds. The trigonal (sp^2) or tetrahedral (sp^3) configurations dictate graphite or diamond structure. Amorphous carbon refers to carbon network that has sp^2, sp^3 bonding structures and almost no sp^1 bond. Similar types of micrographs were also observed in the previous attempts of DLC film deposition. From the configuration of the deposited film that seen in the micrographs (see Fig. 3 and Fig.5), we assume that the films deposited on Cu substrate are DLC films.

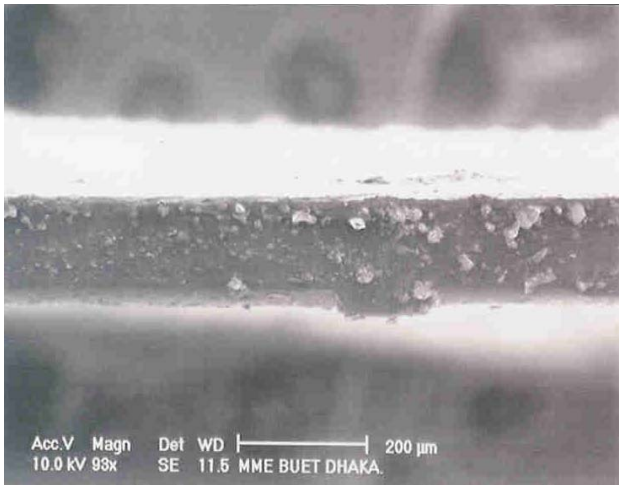


Fig. 6. SEM micrograph of the cross section of Cu substrate with 10 hours deposition.

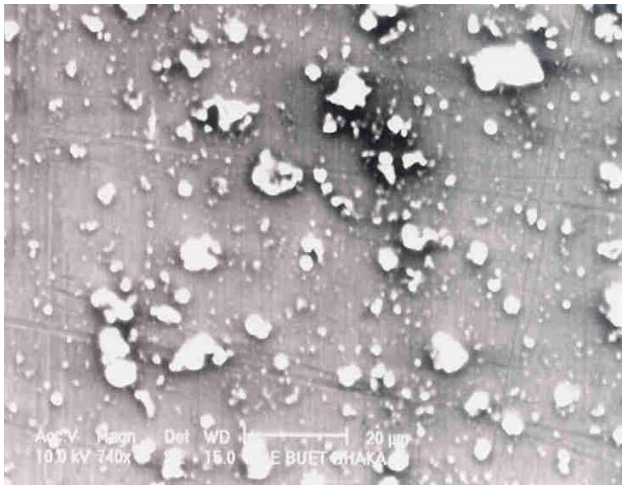


Fig. 7. SEM micrograph of the front surface of Al substrate with 4 hours deposition.

Fig. 7 and Fig. 8 show the micrographs of Al substrates (front surfaces) for 4 hours and 10 hours respectively. According to Fig. 5 and Fig. 8, it is clear that density of deposited films on Cu substrate is much higher than then of Al substrate (for both deposition time was 10 hours). Earlier, it was predicted from the plots shown in Fig. 2. The density of deposited films on Cu substrate shown in Fig. 5 (deposition time 10 hours) is much higher than that of Fig. 3 (deposition time 3 hours). It reveals that deposition density increased with increasing deposition time. The technique which is described here, works better on Cu substrates as shown in the micrographs of Fig. 3 and Fig. 5. We do not find any significant change between micrographs of Fig. 7 and Fig. 8, since the Al substrates were used for film deposition.

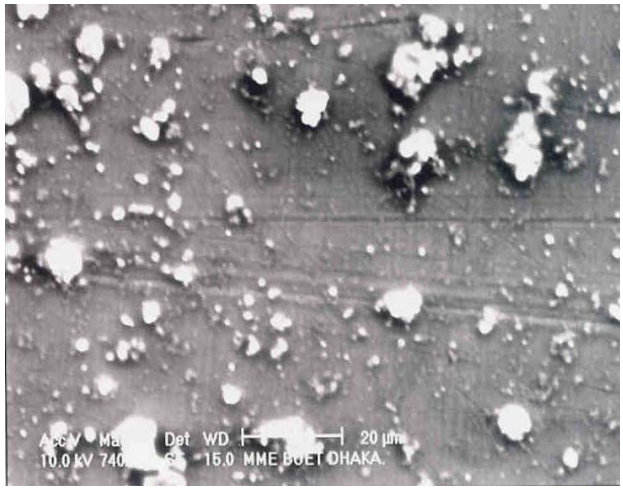


Fig. 8. SEM micrograph of the front surface of Al substrate with 10 hours deposition.

IV. Conclusions and Future Works

In this study, we described a novel approach of DLC film deposition on metal substrates, which is a liquid phase deposition technique. We applied the technique on Cu and Al substrates. We examined surface morphology of the obtained samples by using SEM. It is found that the technique is efficient for depositing DLC films on Cu substrate. Future work can be carried out on Raman spectroscopy of the deposited samples of Cu substrate. It will help to know the structure (sp^3 and sp^2 ratio) and the quality of the deposited film.

References

[1] Jing-Ting Jiu, Li-Ping Li, Chuan-Bao Cao, and He-Sun Zhu, "Deposition of diamond-like carbon films by using liquid phase electrodeposition technique and its electron emission properties," Journal of Material Science, Vol. 36, pp. 5801-5804, December 2001.

[2] MEI Xian-Xiu, LIU Zhen-Min, MA Teng-Cai, and Dong Chuan, "Deposition of Diamond-Like Carbon Films by High-Intensity Pulsed Ion Beam Ablation at Various Substrate Temperatures,"Chinese Physics Letters, Vol. 20, pp. 1619-1621, September 2003.

[3] S. C. Ray, B. Bose, J.W. Chiou, H.M. Tsai, J.C. Jan, Krishna Kumar, W.F. Pong, D. DasGupta, G.Fanchini and A. Tagliaferro, "Deposition and characterization of diamond-like carbon thin films by electro-deposition technique using organic liquid," Journal of Material Research, Vol. 19, pp. 1126-1132, April 2004.

[4] X B Yan, T Xu, S R Yang, H W Liu, and and Q J Xue, "Characterization of hydrogenated diamond-like carbon films electrochemically deposited on a silicon substrate," Journal of Physics D: Applied Physics, Vol. 37, pp. 2416-2424, September 2004.

[5] J. Wang, W.Z. Li, and H.D. Li, "Influence of the Bombardment Energy of CH^{n+} ions on the Properties of Diamond-like Carbon Films," Elsevier Surface and Coatings Technology, No. 122, pp. 273-276, 1999.

[6] Q. Jun, L. Jianbin, W. Shizhu, W. Jing, and L. Wenzhi, "Mechanical and Tribological Properties of Non-hydrogenated DLC Films Synthesized by IBAD," Elsevier Surface and Coatings Technology, No. 128-129, pp. 324-318, 2000.

[7] V. Bursikova, V. Navratil, L. Zajickova, and J. Janca, "Temperature Dependence of Mechanical Properties of DLC/Si Protective Coatings Prepared by PECVD," Elsevier Material.Science and Engineering: A, No. 324, pp.251-256, 2002.

[8] S. Takeuchi, A. Tanji, H. Miyazawa, and M. Murakawa, "Synthesis of Thick DLC Film for Micromachine Components," Elsevier Thin Solid Films, No. 447-448, pp. 208-211, 2004.

[9] E. G.. Spencer, P. H. Schmidt, D. C. Joy, and F. J. Sansalone, "Ion-beam-deposited polycrystalline diamondlike films," AIP: Applied Physics Letters, Vol. 29, pp.118-120, July 1976.

[10] S. M. Mominuzzaman, K. M. Krishna, T. Soga, T. Jimbo and M. Umeno, "Optical Absorption and Electrical Conductivity of Amorphous Carbon Thin Films from Camphor: a Natural Source," Japan Journal of Applied Physics, Vol. 38, pp. 658-663, February 1999.

[11] Y. Namba, "Attempt to Grow Diamond Carbon Films from an Organic Solution,"Journal of Vacuum Science and Technology A: Vacuum Surfaces and Films, Vol. 10, pp. 3368-3370, September 1992.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Crystalline and the luminescence characteristics of β-$FeSi_2$ in photonics formed by pulsed laser deposition

M. Zakir Hossain, T. Mimura and S. Uekusa
Department of Electronics and Bioinformatics, School of Science and Technology, Meiji University,
1-1-1 Higashimita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan.
Corresponding author E-mail: zakirh76@isc.meiji.ac.jp

Abstract - Compound semiconducting silicide β-$FeSi_2$ has been formed on FZ - Si (111) substrates by means of pulsed laser deposition (PLD) method using ArF (λ = 193 nm) excimer laser. In photoluminescence (PL) measurements at 8K detected by a Ge detector, the PL spectra of the samples annealed at 900 °C for 1, 5, 8 and 20 hrs showed that the PL intensities of the A-and peak increased depending on annealing time in comparison with those of the as-deposited sample. The intrinsic PL intensity of the A-band peak from 20 hr annealed sample was investigated. The dependence of the PL excitation power density of the 20-hr-annealed sample also showed A-band peak at 0.808 eV that almost constant peak position. The temperature dependence PL intensity of the 20-hr-annealed sample in the ranges 15K~150K by an InGaAs detector showed the PL peak was at 0.808 eV. This peak confirmed the intrinsic PL peak of A-band of the β-$FeSi_2$. The transmittance of the thin films was measured in the range 0-2500 nm. Moreover, it is calculated the refractive indices using the reflectance. We report an application of β-$FeSi_2$ with crystalline and luminescence characteristics with a high refractive index in photonics.

I. Introduction

Semiconducting silicide β-$FeSi_2$ has attracted much attention as a candidate for silicon based optoelectronic materials for last one decade. Single-phase β-$FeSi_2$ with high crystal quality fabrication process has chemical, physical stability, absorption coefficient indicating that it has potential as photovoltaic and solar energy cell [1, 2]. Moreover, due the band gap of the β-$FeSi_2$ at about 0.80-0.89 eV [3], the fabrication process of β-$FeSi_2$ devices is compatible with silicon CMOS technology. Therefore β-$FeSi_2$ has great potentiality as a silicon-based light emitting material [4]. Various techniques such as molecular beam epitaxy (MBE) [5], electron beam evaporation and ion beam synthesis (IBS) [6, 7] has been developed to investigate the thin films of β-$FeSi_2$. Still there are some difficulties to it's prepare conditions such as chemical and physical stability, cleanness, deposition efficiency and maintain the optical behavior. Among all of these methods, PLD is unique for its deposition efficiency, mechanism and cleanness while preparing the thin films. In addition, PLD is more appropriate for the preparing better crystalline structure and luminescence characteristics of the thin films [8, 9]. An author of this paper previously reported the β-$FeSi_2$ prepared by IBS method [6, 9]. In this study from view point of the large future application of the β-$FeSi_2$ in photonics, optical fiber communications, photovoltaic, solar energy and light emitting diode (LED), we studied the large variety of experiments [10, 11] of XRD, Photoluminescence (PL), the transmittance and refractive indices (n). Based on these experimental results and data we investigated that the long time high temperature thermal annealing process has been developed and shown to be able to significant improvement of the crystalline and luminescence characteristics of the β-$FeSi_2$ in photonics, optical fiber communications.

II. Experimental procedure

Prior to laser ablation, the FZ n-Si (111) (1000-20000 Ω · cm) substrates were cleaned with organic solvents. The substrates were dipped in a dilute HF solution with a ratio of (HF: H_2O = 2:40) for one min [10]. The etched Si (111) substrates were then rinsed in deionized water and subsequently loaded into the growth chamber with a base pressure 10^{-5} Pa. The β-$FeSi_2$ films were grown at a substrate temperature of 600 °C with the sintering $FeSi_2$ target (99.99 %). The laser source used was an ArF excimer laser (wavelength; 193 nm) and the laser fluence was 4.0 J/cm^2. The β-$FeSi_2$ film was subjected to high-temperature long-time annealing using an infrared lamp in a continuous-gas-flow and N_2-atmosphere (99.9995 %). Moreover to improve the crystalline structure quality and to enhance the optical and epitaxial growth of the β-$FeSi_2$ sample, they were annealed for 1, 5, 8, 10 and 20 hrs. The annealing temperature was 900 °C [12]. The chemical composition, thickness and crystalline characteristics of both as-deposited and annealed samples were characterized by X-ray diffraction (XRD) measurements using CuKα (λ = 1.54178 Å) radiation with glancing angle incidence has been employed in the range 10 to 80 degrees to identify the phase of as-deposited and annealed samples. PL measurements at 8K were performed using the wavelength of 514.5 nm line of an Ar ion laser with an average excitation power of 200 mW for exciting light source. The emission was dispersed by a single monochromatic and was detected using a liquid nitrogen-cooled Ge p-i-n photodiode. The temperature dependence PL of 20-hr-annealed sample was also examined at temperature ranges of 15 K~160 K. The sample also

carried out by single monochromatic and also using a liquid nitrogen cooled InGaAs detector. In this case the PL measurements performed using the wavelength of 488 nm line of an Ar ion laser with the focal length 32cm. At room-temperature (RT), optical transmittance (T) measurements were performed to determine the absorption coefficient of the β-$FeSi_2$ thin films. Moreover, using the transmittance of the samples we also calculated the reflectance (R) and the refractive (n) indices.

III. Results and Discussion

Figure 1 shows the XRD pattern of the β-$FeSi_2$ films as-deposited, after annealing to investigate the structural properties and the crystallographic orientation. The thin films show the XRD spectra from (202) or (220) signals of the β-$FeSi_2$ as deposited and subjected to high-temperature (900^o C) thermal annealing for a long time (5~20 hrs). The XRD peaks can be observed in as-deposited to 20-hr-annealed samples. There are no obvious signal peaks corresponding to the other β-$FeSi_2$ and correlated phases except that from Si substrates. This indicates that the epitaxial growth of β-$FeSi_2$(202 or 220)/Si(111) and single-crystalline β-$FeSi_2$ can be prepared using PLD.

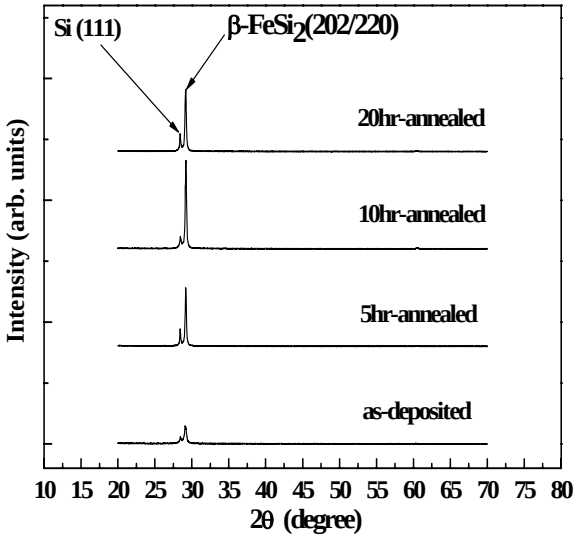


Fig. 1. The XRD pattern of the β-$FeSi_2$ thin films as-deposited and annealed at 900^{o}C with the different annealing time in the ranges 5, 10 and 20 hrs.

This result can be explained as follows. The crystallinity of the β-$FeSi_2$ samples formed by the PLD method depends on ablation condition and the long time annealing increased 5 to 20 hr at 900 ^{o}C. Therefore, it is easy to form the crytallinity and epitaxial growth of the samples. From the above results in all stages of XRD measurements, it can be considered that the long time annealing 5 to 20 hr and the annealed temperature 900 ^{o}C would be the appropriate condition for the fabrication of the semiconducting single crystalline silicide β-$FeSi_2$.

Figure 2 shows the PL spectra of β-$FeSi_2$, all the samples shown in the figure were as-deposition and annealed at 900 ^{o}C for 1, 5, 8 and 20 hrs. It is clearly shown that increasing the annealing time markedly increased the PL intensity. Maeda et al. [13] reported on the basis of ion beam synthesis (IBS) method sample that the PL spectra at A-band 0.805 eV is the intrinsic luminescence of β-$FeSi_2$ film, the B-band of the PL spectra at 0.841 eV is caused by the defects of the Si interface and C-band of the PL spectra at 0.766 eV is caused by defects in β-$FeSi_2$ film. We have also investigated the PL spectra of the PLD prepared samples confirmed to the results of Maeda et al [13]. The PL spectra of the as-deposited sample showed the A-band peak and a broad peak at around 1.1 eV. The broad peak is considered to be the PL peak of Si substrates. The PL spectra of the samples annealed for 1, 5, 8 and 20 hrs showed that the PL intensities of the intrinsic main peak A-band and C-band peaks increased depending on annealing time in comparison with those of the as-deposited sample. The PL intensity of the 20-hr-annealed sample shows that the intrinsic A-band peak was significantly improved and became stronger corresponding in comparison with those of others time-dependent samples. This result shows that the enhanced optical and the luminescence characteristics of the thin film can be attributed to the improvement of the PL intensity due to the long-time annealing at the temperature of 900 ^{o}C for 20 hr. Katsumata and Uekusa et al. [6] previously reported by IBS method that the PL spectra of the β-$FeSi_2$ at 0.805 eV to 0.807 eV. We have also investigated A- band peak (0.808 eV) is emitting from the β-$FeSi_2$ by the PLD method which is very close to the IBS method. This main peak (A-band peak) of the PL spectra will discuss in details on following.

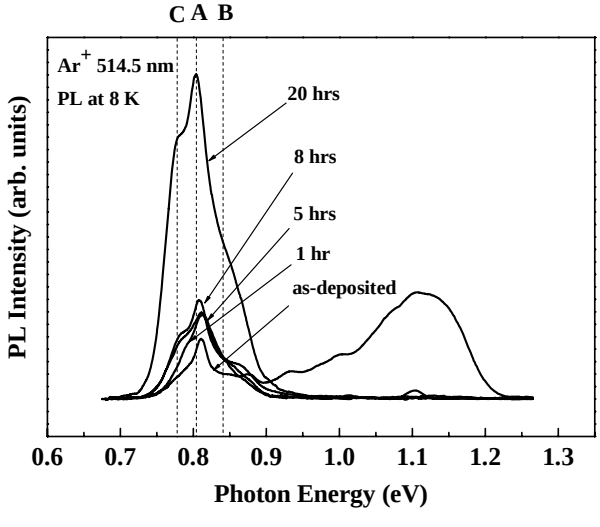


Fig. 2. The PL spectra of the β-$FeSi_2$ thin films as deposited and annealed at 900^{o}C with the different annealing time for the ranges 1, 5, 8 and 20 hrs.

Figure 3 shows the dependence of the intrinsic A-band PL peak energy on the excitation power density curve. In order to investigate the properties of intrinsic PL peak (A-band) of the β–$FeSi_2$, we examined the excitation power density of the 20-hr-annealed sample at ranges of 0.01 W/cm^2 to 10 W/cm^2. In this step it found the PL intensity of the A-band peak is almost constant and stable while the other band peaks corresponding the defect related band of D and B (not shown in Fig.3) has large dependence on the laser power excitation of the PL measurements. The A band peak is sharp and no change its peak position (0.808 eV). This band peak can be attributed to β-$FeSi_2$ sample's originating and confirmed ever first time by PLD. The PL's at 0.84 eV ~0.89 eV and 0.77 eV~0.79 eV may be attributed to dislocation related lines [16].

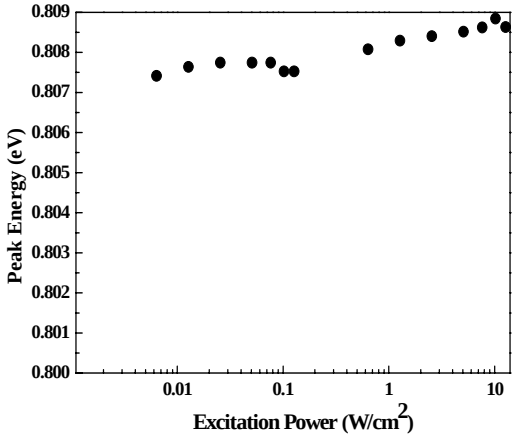


Fig.3. The dependence of the intrinsic A-band PL peak energy on the excitation power density.

Figure 4 shows the temperature dependence of the PL intensity of 20-hr-annealed sample. The aim of the temperature dependent PL spectra is also to investigate the intrinsic PL i.e. the A-band peak at 0.808 eV is originating from the β-$FeSi_2$. The temperature of the PL measurements was carried out in ranging of 15 K~150 K and laser source was Ar^+ ion laser with the wavelength of 488 nm detected by an InGaAs detector. After 150 K the PL intensity was not observed on the sample. It is investigated that A-band peak energy decreased after increasing the PL measurements temperature of the 20-hr-annealed sample.

Figure 5 shows the temperature dependence of the A-band PL peak energy (E_p) for the 20-hr-annealed sample. To characterize the PL peak energy we have been used the Varshini's formula (as an empirical law) by the following equation (1)

$$E_p(T) = E_p(0) - \alpha T^2/(\beta+T) \qquad (1)$$

where $E_p(0)$ is the peak energy at 0K, α is the energy shift for temperature and β is the constant. In the equation (1) Martineli et al. [16] $Ep(0)$ = 0.8052 eV, β = 600 and $\alpha = 4.2\times10^{-4}$ eV/K also reported the same calculation by IBS sample. In our calculation we obtained that the Varshini's parameters are $Ep(0)$ = 0.8085eV, $\alpha = 1.8\times10^{-4}$ eV/K and β = 300. Their results reported very close similarity to our results obtained from 20-hr-annealed sample. The 20-hr-annealed sample showed very small temperature dependence of the PL peak energy. In 20hr annealed sample the small α value has been obtained and can be ascribed the intrinsic PL (A-band) of the β-$FeSi_2$. This result has been found for the first time in PLD prepared sample.

Figure 6 shows a schematic diagram explanation of the radiative process resulting from the dissociation of excitons at the band gap calculated at Y and Λ points. The excitons dissociation need the emission of phonon

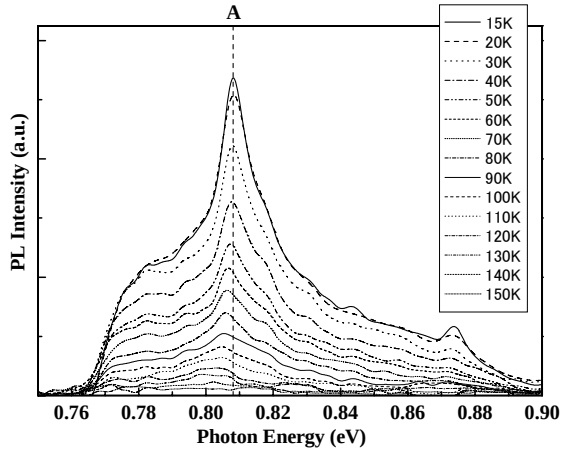


Fig.4. Temperature dependence of PL intensity of the 20-hr-annealed sample showed the A-band peak.

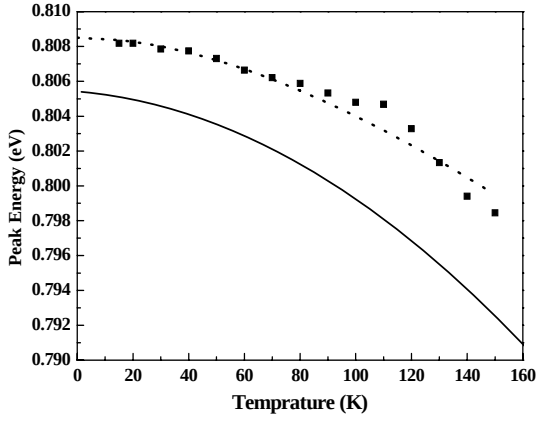


Fig.5. The temperature dependence of the A-band PL peak energy (E_p) for the 20-hr-annealed sample.

E_{ph} = 31 meV, E_{ph} is the phonon and E_x is the exciton of the band edge [12]. As we have obtained the A-band peak at 0.808 eV which is smaller to direct band gap [16]

therefore we have calculated the indirect band-gap energy by the following equation (2)

$$E_g^{ind} = E_{PL} + E_{ex}^{1st}(r) + E_{ph} \quad (2)$$

After considering the intrinsic A-band peak (0.808eV), the phonon E_{ph} is 31 meV and E_{ex} = 0.003 meV, we have obtained the band gap E_g^{ind} is 0.842eV which can be compare as an indirect band-gap energy of Giannini et al. (0.83 eV) [17] and Maeda et al. 0.839 eV [12]. From these experimental results and discussion by different groups, it can be considered the energy band-gap obtained by PLD prepared sample is very close compare to other groups.

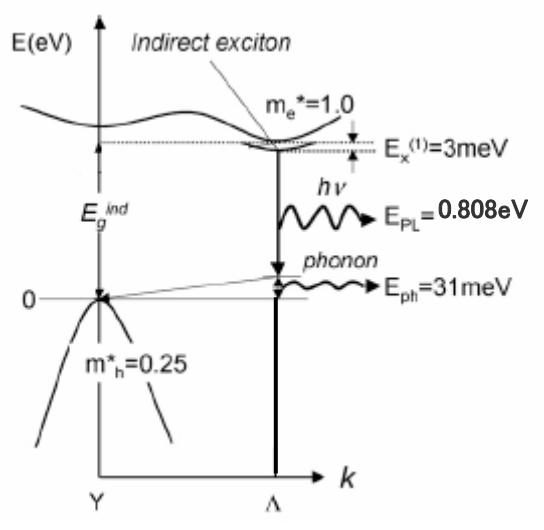


Fig. 6. Schematic diagram explanation of the radiative process resulting from the dissociation of excitons with phonons emission at the indirect band gap at Y and Λ points.

Figure 7 shows the transmittance of the thin films of the as-deposited and at ranges of 5 to 20 hr annealed samples. In the wavelength region at 0 to 2500 nm, there was no transmittance of the samples due to the absorption of the samples. At 1200 nm region (1.033 eV), it is observed that the transmittance of the Si substrates. Moreover, at 1500 nm region (0.827 eV), the transmittance of annealed samples decreased. It is considered that the absorption of the β-$FeSi_2$ occurred because of corresponding of the PL spectra A-band (0.808 eV). In the transparent region at 2000 to 2500 nm, it clearly shown that after increasing the annealing time ranges from 5 to 20 hr the transmittance percentage also increased. Especially at 2500 nm the transmittance of the thin films show that the transmittance percentages for as-deposited and 5, 10 and 20 hrs annealed samples were 53.37 %, 56.24 %, 60.35 % and 72.34 % respectively. From the above results of transmittance at 2500 nm, we also calculated the reflectance and the refractive index of the samples by using the following equation

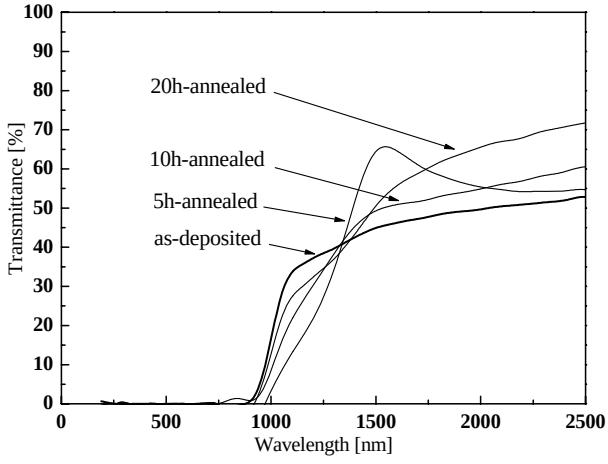


Fig. 7 The transmittance spectra of the β-$FeSi_2$ /Si (111) thin films of as-deposited, 5, 10 and 20 hrs annealed samples.

$$1 - T = R = \left(\frac{n-1}{n+1}\right)^2 \quad (3)$$

where T be the transmittance, R be the reflectance and n be the refractive index. However no absorption is considered in the transparent region. Therefore, the reflectance for as-deposited and 5, 10 and 20 hrs annealed samples are 0.4663, 0.4376, 0.3965 and 0.2766 respectively. After calculating the equation, it is found the refractive indexes (n) are 5.3, 4.9, 4.4 and 3.2 respectively for as-deposited, 5, 10 and 20 hr annealed samples. As per our knowledge this result has been investigated for the first time by PLD oriented thin films. Moreover, this result (n = 3.2~5.3) is compare to very near with Maeda et al [12] reported simulation results using a finite difference time domain (FDTD) method with high refractive index (n~5.6) at an infrared region. Development of Photonic crystals (PhCs) for optical fiber communications is proceeded by materials such as mainly InGaAsP/InP, InGaAs/GaAs or Si on insulator (SOI). For optical fiber communication we need PhCs that can control 1.3 to 1.55 μm light propagation. As semiconducting β-$FeSi_2$ has a high refractive indices 3.2~5.3 in the transparent range, it can be reported the possible application of high refractive semiconducting silicide of high index contrast photonic crystal (PhC).

IV. Conclusions

The maximum intrinsic A-band PL intensity by Ge detector from the 20-hr-annealed sample has been investigated, compare to those of other time dependent annealed samples. The 20-hr-annealed sample also showed that intrinsic A-band PL peak (0.808 eV) detecting by an InGaAs detector and confirmed the A band peak is originating from the β-$FeSi_2$. The XRD shows that the crystalline structure enhanced after increasing the annealing time. From the band edge

calculation β-$FeSi_2$ the indirect band gap (E_g^{ind} = 0.842 eV) is similar to others group report can be applicable in photonics and optical fiber communication. Moreover, it is found that the refractive indices (n) calculated using the results of the reflectance is very near compare to the reported simulation results. From the above results and discussions, it is suggested that the β-$FeSi_2$ formed by PLD is highly effective semiconducting material for the application in photonics and optoelectronics. Moreover it is also an effective process for the better crystalline and luminescence characteristics of the semiconducting silicide β-$FeSi_2$.

Acknowledgement

We would like to thank Dr. Yoshiaki Hara for his kind support in PL measurements by Ge detector at Ibaraki National College of Technology, Japan.

References

[1] K. Lefki, P. Muret, N. Cherif and R.C. Cinti: J. Appl. Phys. 69 (1991) 325.
[2] E. Grob, M. Riffel and U. Stonrer: J. Mate. Res. 10 (1995) 34.
[3] M.C. Bost and J.E. Mahan: J. Appl. Phys. 64 (1988) 2034.
[4] D. Leong, M. Harry, K. J. Resson and K. P. Homewood: Nature 387 (1997) 686.
[5] T. Suemasu, T. Fuji, K. Takakura M. Tanaka and F. Hasegawa: Journal of Luminescence 80 (1998) 473.
[6] H. Katsumata, Y. Makita, N. Kobayashi, H. Shibata, M. Hasegawa, I. Aksenov, S. Kimura, A. Obara and S. Uekusa: J. Appl. Phys. 80 (1996) 5995.
[7] Y. Maeda, K. Umezawa, Y. Hayashi, K. Miyake, K. Ohashi: Thin Solid Films 381(2001) 256.
[8] T. Yoshitake, T. Nagamoto and K. Nagayama: Thin Solid Films 381 (2001) 236.
[9] S. Uekusa, M. Yamamoto, K. Tsuchiya and N. Miura: Mat. Res. Soc. Symp. proc.Vol.744 (2002).
[10] S. Uekusa, M. Zakir Hossain, K. Aoki, T. Fukuda and N. Miura: Mater. Res. Soc. Symp. Proc. Vol. 0891-EE03-24 (2006) 163.
[11] M. Zakir Hossain, K. Aoki, T. Fukuda, N. Miura and S. Uekusa: WIP-Renewable Engineers. 1 AV. 2.39 (2006) 215.
[12] Y. Maeda: Applied Surface Science 254 (2008) 6242.
[13] Y. Maeda: Funct. Mater. 10 (2005) 5 (in Japanese).
[14] Y. Terai, M. Itakura, N. Kuwano and Y. Maeda: Thin Solid Films 461 (2004) 160.
[15] A.G Birdwell, T.J Shaffner, D.Chandler Horowitz and S. Collins: J Appl.Phys. 95 (2004) 2441.
[16] C. Giannini, S. Lagomarison, F. Scarinici, P. Castrucci: Phys. Rev. B 45 (1997) 8822.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Interpretation of Cu(111)//Nb(110) Growth on SiO_2 by Transmission Electron Microscopy

Md. Maniruzzaman[1] and Atsushi Noya[2]

[1] Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh.
[2] Department of Electrical and Electronic Engineering, Kitami Institute of Technology, Kitami 090-8507, Japan.
[1] E-mail: mz.ece@ku.ac.bd

Abstract - The [110]-textured extremely thin Nb barrier layer is prepared on an amorphous SiO_2 substrate, on which the preferentially oriented Cu [111] texture is obtained. Transmission Electron Microscopy (TEM) observation indicates that nanocrystalline grains no larger than ~10 nm with a spread [110] orientation along the substrate normal direction is characteristic of a 10 nm thick Nb barrier on SiO_2, whereas for the 100 nm thick barrier, the well-oriented columnar structure of a Nb [110] layer is obtained by ordering the orientation during the increase in the thickness of the Nb film by the coalescence of the initially deposited nanocrystalline grains. The quality of Cu [111] layer is much better when deposited on 100 nm thick Nb layer than that on a 10 nm thick barrier.

I. Introduction

Presently, Cu is used as an interconnecting material for Si ultra-large-scale-integrated (Si-ULSI) circuits mainly due to its lower electrical resistivity and high resistance against electromigration (EM). But, EM problem is still a major concerning issue in the Cu interconnect technology because of the increasing current density along fine interconnects, as the on-chip interconnect dimensions are continuously scale-downed. As well known in Al interconnects [1], it is also true that the excellent EM resistance is generally obtained in Cu interconnects when the most closely packed (111) plane of Cu as an fcc metal is placed in the normal direction to the current flow direction [2]. So, development of stronger [111] texture of Cu seed layer on a barrier metal within a trench formed in a field insulating layer (Cu dual-damascene process) is essential, which in turn will enhance the texture of the Cu layer grown on the seed layer in interconnect lines/ vias. The barrier layer is employed to prevent fast Cu diffusion into nearby insulating layer, e.g., SiO_2, which causes the degradation of device or integrated circuit performance. As the textures of Cu layers depend on the microstructures of the underlying barrier metal, an appropriate barrier layer, on which Cu [111] texture will be grown, is required for Cu interconnects in downscaled circuits.

Based on theoretical and experimental analyzes [3, 4], it is reported that a closely packed fcc (111) plane can grow epitaxially on a closely packed bcc (110) plane by minimizing the interfacial energy between the planes, Cu $(111)_{fcc}$/Nb $(110)_{bcc}$ and Cu $(111)_{fcc}$/Mo $(110)_{bcc}$ [5], for example.

However, Ta or Ta-based material is known as an extensively used barrier material for Cu interconnects, and it is reported that Cu [111] texture can usually be obtained on bcc Ta [110] [6, 7]. But, it is difficult to obtain bcc Ta [110] especially in thin films, as Ta shows a polymorphic meta-stable tetragonal phase (β-Ta) together with the bcc phase (α-Ta) in thin films [8].

Under this pretext, in this work, we have taken into account Nb as the requisite barrier material because of its low resistivity, good structural stability of bcc Nb phase compared to Ta and good lattice match with Cu alike bcc Ta. Besides, chemical properties of Nb are almost identical to those of Ta, as both the metals belong to the same group 5 in the periodic table.

In this study, firstly, we have shown the successful formation of the preferentially oriented Cu [111] texture on the [110] preferentially oriented Nb barrier deposited on traditional insulating layer of SiO_2, where we have used a rather thick Nb barrier (~100 nm). This has been based on an already mentioned epitaxial relationship of $[111]_{fcc}$// $[110]_{bcc}$.

However, the barrier layer thickness should be as thin as possible to obtain sufficient cross-sections of Cu interconnects – a 10 nm thick barrier is defined for the 90 nm node technology [9], for example. Therefore, in the present study, secondly, we have interpreted the growth and the quality of the Cu [111] layer employing the extremely thin Nb [110] barrier (~10 nm) in details.

II. Experiments

Specimens of Cu/Nb/SiO_2/Si were prepared by using a tetrode dc sputtering system with a base pressure of less than 5×10^{-7} Torr. On a thermally oxidized SiO_2 layer with 100 nm in thickness grown on a p-type Si (100) wafer, an Nb layer (10 – 100 nm) and a subsequent Cu layer (200 – 300 nm) were deposited at 400 °C and room temperature, respectively, without breaking the vacuum. The target voltage and current were 500 V and 80 mA, respectively, and the pressure of Ar gas was maintained at 2×10^{-3} Torr during sputtering. Crystallographic orientations between Cu and thin Nb layers have been examined using transmission electron microscopy (TEM). The TEM specimens were prepared by a conventional Ar-ion thinning technique.

III. Results and Discussion

A. Analyses of Cu/Nb (100 nm)/SiO_2 system

The crystallographic textures of the Nb and Cu layers in the Cu/Nb (100 nm)/SiO_2/Si specimen were examined by TEM. Figure 1 shows the cross-sectional TEM view of the specimen. In the figure, the deposited Nb barrier on SiO_2 and the subsequently grown Cu layer, both are uniform in thickness, are seen with sharp transition at the interfaces. The columnar structure is characteristic of the Nb and Cu layers. The column size of Nb is smaller than that of Cu. It is plausible since the Nb grains first nucleated at the nucleation sites, which are distributed at random with an appropriate density on amorphous SiO_2, the grain growth in random orientation in the lateral direction will result in the restriction of the base of the Nb column in size. The column sizes in the lateral direction are typically about 20 – 50 nm for the Nb barrier and 100 – 200 nm for the Cu layer. The columns consist of some mosaic grains small in sizes as a result of nucleation-type growth of the layer. For Cu grains on Nb (110)/SiO_2, Cu (111) planes are almost perpendicular to the thin film growth direction. The relationship of subsequent growth, Cu (111)//Nb (110), suggested by an epitaxial relationship [3, 4], is confirmed by nano-spot electron diffraction in typical combinations of Cu/Nb columns under observation, as typically shown in Fig. 1. In the lateral direction, however, Nb columns are grown at random in the orientation.

A typical high-resolution TEM (HRTEM) image around the Nb/SiO_2 interface is shown in Fig. 2, where the Nb (110) grains are grown on SiO_2. The image shows a distinct interface layer (1.5 – 2.5 nm) between Nb and SiO_2, which is developed at the first stage nucleation of the Nb atoms on amorphous SiO_2. The fluctuation of the growth direction from the normal direction from every Nb grain is seen.

At the Cu/Nb interface from the specimen, as shown in Fig. 3(a), the lattice fringes are observed close to each other at the interface, indicating the absence of the inter-diffused and/or reaction layer, although the overlap of the grains make the interface obscure in the observed view. At the Cu/Nb interface, rather small Cu grains are seen, indicating the nucleation-type growth of Cu is dominant at the first stage of the Cu deposition. However, as shown in Fig. 3(b), it is seen that the shape of the Nb grains affects the Cu/Nb interface in morphology. The lattice fringes of Cu (111) are seen to be parallel to those of Nb (110) lattice in the individual grains, indicating the relationship of subsequent growth of Cu (111) on Nb (110). A very thin amorphous interlayer (1.5 – 2.5 nm) probably caused by intermixing of elements (Cu and Nb) is seen at the interface. Kwon et al. [10] reported the existence of an amorphous interlayer consisting of a mixture of Cu and Ta at the as-deposited Cu/tetragonal Ta interface, while the interlayer was not evident at the Cu/bcc Ta interface, where the Cu deposition was carried out at room temperature (the same condition as that in this study). In this study, the Cu/Nb interface with an amorphous interlayer and that free from solid-phase amorphization are simultaneously observed in the as-deposited specimen depending on the area under observation.

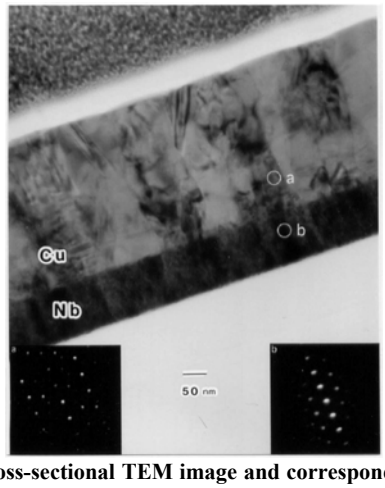


Fig. 1 Cross-sectional TEM image and corresponding nano-probe electron diffraction patterns of Cu/Nb (100 nm)/ SiO_2/Si specimen.

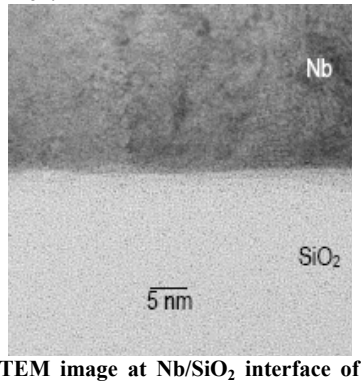


Fig. 2 HRTEM image at Nb/SiO_2 interface of the Cu/Nb (100 nm)/ SiO_2/Si specimen.

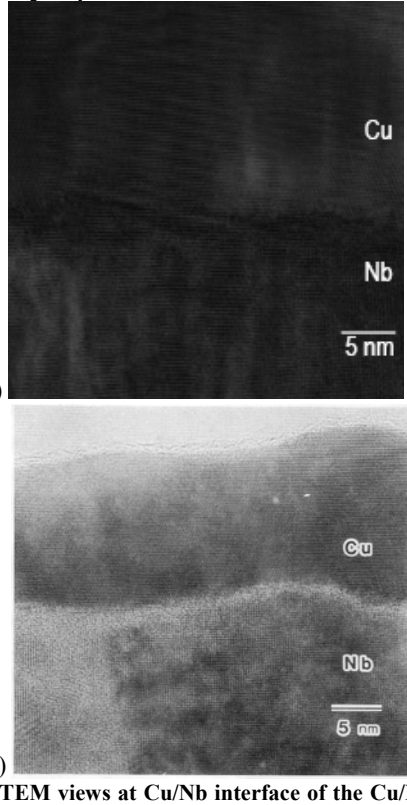


Fig. 3 HRTEM views at Cu/Nb interface of the Cu/Nb (100 nm)/SiO_2/ Si specimen.

B. Analyses of Cu/Nb (10 nm)/SiO_2 system

The crystallographic textures of the Cu and Nb layers in the Cu/Nb (10 nm)/SiO_2/Si specimen were examined by TEM to clarify the growth of Cu [111] texture on the Nb barrier of ~10 nm in thickness. Figure 4 shows a whole view of the Cu/Nb (10 nm)/SiO_2/Si system, where pronounced columnar structure of the Cu layer consisting of mosaic grains is characteristic, and the Nb barrier is uniform and continuous with smooth interfaces of Cu/Nb and Nb/SiO_2.

A typical enlarged HRTEM image around the Nb barrier in the specimen is shown in Fig. 5. The HRTEM image exhibits the ~10 nm thick Nb barrier with a columnar structure, and Cu grains adjoining the Nb barrier are also seen. An Nb column consists of one or two grains at most in the direction substrate normal is characteristic. In the lateral direction, grains in ~10 nm size are observed, which are almost the same as the barrier thickness. The HRTEM image shows the lattice fringe spacing of 0.235 nm in Nb and that of 0.209 nm in Cu, which are well corresponding to the lattice spacing of Nb (110) (0.233 nm) [11] and Cu (111) (0.208 nm) [12], respectively. In the Nb barrier, the Nb (110) lattice fringes are seen to be nearly parallel to the substrate for the most part of the barrier, although the inclination of lattice fringes from the level parallel to the substrate surface is somewhat much to connive at in every grain, indicating that the quality of [110] orientation in the present Nb [110] texture is inferior to that previously described 100 nm thick barrier.

In the deposition process of Nb on SiO_2, initially deposited Nb atoms on SiO_2 have low surface mobility due to high chemical affinity of Nb to oxygen (O) in SiO_2. This becomes a cause of insufficient surface diffusion of deposited atoms during Nb grain nucleation, resulting in a nanocrystalline phase of Nb and inferior preferred orientation due to insufficient relaxation of interfacial energy in every grain. If we prepared a thick Nb film, good [110] preferred orientation was obtained by relaxation of interfacial energy during growth and coalescence of grains, as already described in the previous section for 100 nm thick Nb layer.

On the other hand, Cu grains are sufficiently larger than those of Nb, as already seen in Fig. 4. The lattice fringes of Cu (111) are seen to be mainly parallel to the substrate surface direction, however, relatively large inclination of lattice fringes from the substrate parallel is evident, for example, a tilting angle of ~7°, as shown in white in Fig. 5. Observation of the Cu layer indicates that grains adjoining the Nb barrier are small in size with rather larger inclination angles of fringes from the horizontal level parallel to the substrate surface as compared with those in the upper part of Cu columns, suggesting that initially nucleated Cu grains are strongly affected by the state of crystalline surface of Nb (110) so as to lower the interfacial energy of the Cu/Nb interface dominated by local chemistry. This is because initially grown Cu [111] grains nucleate on Nb [110] nanograins with relatively large mosaicity to satisfy the previously mentioned relationship of Cu (111)//Nb (110), resulting in the formation of Cu grains near the Cu/Nb interface with rather large deviation of [111] orientation of each grain around the direction of substrate normal. Subsequent growth of the Cu layer during deposition brings about the improved growth of [111] preferred orientation by coalescence of small grains to form relatively large columns consisting of well ordered [111] mosaic grains.

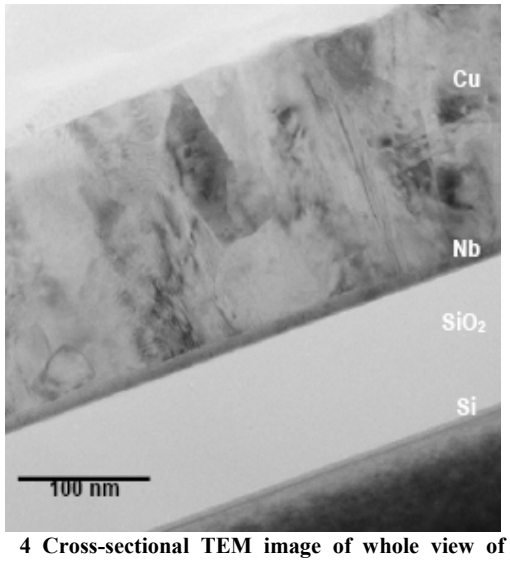


Fig. 4 Cross-sectional TEM image of whole view of the Cu/Nb (10 nm)/SiO_2/Si specimen.

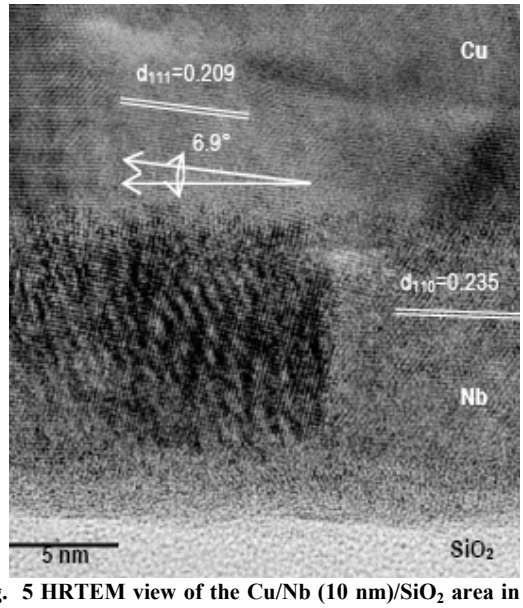


Fig. 5 HRTEM view of the Cu/Nb (10 nm)/SiO_2 area in the specimen.

Interestingly, the authors can barely find an Nb grain in random orientation in the barrier, as seen in Fig. 6, where the Cu grains on the noted Nb grain in random orientation shows Cu [111] preferred orientation setting the expected relationship at naught. As already seen, Cu grains are generally larger than those of Nb, which is just the case for the Cu grain grown on the noted Nb grain, as pointed out in Fig. 6. Therefore, the orientation of Cu grains is probably dominated by the surface structure of the Nb barrier, at which the Nb (110) plane is major in surface area ratio.

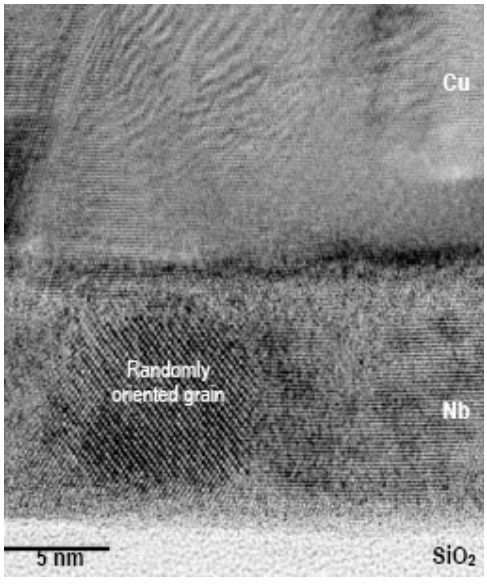


Fig. 6 HRTEM view of the Cu/Nb (10 nm)/SiO_2 area in the specimen, showing Cu (111) grain growth on a randomly oriented Nb grain.

The TEM observation reveals that the quality of the Cu [111] layer is mainly dominated by the quality of Nb [110] preferred orientation. In the present case, the Nb [110] orientation is not always good in quality, because the film growth of Nb in nucleation type is quenched in the initial stage of film growth without a decrease of interfacial energy due to extremely thin thickness, nevertheless, we can obtain the Cu [111] texture of good quality, which is up to our expectation. It is interesting that the mosaic spread of Cu [111] orientation around the direction of substrate normal is always superior to that of the Nb barrier underneath, although the quality of the Cu layer is also dependent on that of Nb underneath. From a point of view of the film growth, it is very difficult to obtain the textured film of a preferred orientation with extremely thin thickness on an amorphous insulating layer. Under the circumstances, the present Cu [111] layer prepared on the Nb barrier as thin as ~ 10 nm in thickness is a successful result for the Cu [111] texture on an extremely thin barrier in the forthcoming metallization technology.

In Figs. 5 and 6, the authors (and also the readers) can see a very thin interlayer of 1 – 2 nm in thickness at the interface of Nb/SiO_2, and also at the Cu/Nb interface. As already speculated, these interfacial layers are probably attributed to intermixing and/or reaction of elements at the interfaces. Especially, reaction of oxidation/reduction is suggested at the Nb/SiO_2 interface. Although these interfacial layers are sometimes good for the adhesion of layers at interfaces, those formed by diffusion or reactions at interfaces in thin film assemblies generally possess a typical value of several nm in thickness. Ideally, formation of interfacial layers should be eliminated from the Cu metallization system with an extremely thin barrier, because the thin barrier is sacrificed due to formation of interfacial layers. Furthermore, at the worst, the growth of interfacial layers upon annealing exhausts the whole of the thin barrier. Fortunately, both of these phenomena are not the case in the present Cu/Nb/SiO_2/Si system. This is fatefully important for the extremely thin barrier application of the present thin Nb [110] film.

IV. Conclusion

Preferentially oriented Cu [111] texture deposited on a thin Nb layer is characterized in a thin film stacked structure of Cu/Nb/SiO_2/Si as a challenge for preparation of a Cu [111] seed layer of interconnects on an extremely thin diffusion barrier. The Cu [111] layer is obtained on Nb films of [110] orientation in various thickness; however, mosaic spread of Cu [111] texture depends on the thickness of Nb film underneath. TEM reveals the Nb layer (10 nm thick) consisting of relatively mosaic spreaded [110] fine grains, which is a result of an initial stage of nucleation type growth of Nb film in columnar structure on SiO_2. It is revealed that the Cu [111] texture of relatively good mosaicity in columnar structure is obtained on the 10 nm thick Nb layer of inferior mosaicity.

References

[1] S. Vaidya, and A. K. Sinha, “Effect of texture and grain structure on electromigration in Al-0.5%Cu thin films”, Thin Solid Films, vol. 75, no. 3, pp. 253-259, January 1981.

[2] C. Ryu, K.-W. Kwon, A. L. S. Loke, H. Lee, T. Nogami, V. M. Dubin, R. A. Kavari, G. W. Ray, and S. S. Wong, “Microstructure and reliability of copper interconnects”, IEEE Trans. Electron. Dev., vol. 46, no. 6, pp. 1113-1120, June 1999.

[3] Y. Gotoh, M. Uwaha, and I. Arai, “Interpretation of the epitaxial orientation relationship at bcc(110)/fcc(111) interfaces ”, Appl. Surf. Sci., vol. 33/34, pp. 443-449, September 1988.

[4] Y. Gotoh, and H. Fukuda, “Interfacial energy of the bcc (110)/fcc (111) interface and energy dependence on its size”, Surf. Sci., vol. 223, no. 1/2, pp. 315-325, December 1989.

[5] Y. Nakasaki, G. Minamihaba, K. Suguro, and H. Itow, “Interfacial energy calculation at interconnect-metal/barrier-metal interfaces for grain orientation control”, J. Appl. Phys., vol. 77, no. 6, pp. 2454-2461, March 1995.

[6] C.-Y. Yang, J. S. Jeng, and J.S. Chen, “Grain growth, agglomeration and interfacial reaction of copper interconnects”, Thin Solid Films, vol. 420/421, pp. 398-402, December 2002.

[7] J.-W. Lim, K. Miyake, and M. Isshiki, “Characteristics of ion beam deposited copper thin films as a seed layer: effect of negative substrate bias voltage”, Thin Solid Films, vol. 434, no. 1/2, pp. 34-39, June 2003.

[8] L. G. Feinstein, and R. D. Huttemann, “Factors controlling the structure of sputtered Ta films”, Thin Solid Films, vol. 16, iss. 2, pp. 129-145, May 1973.

[9]. http://www.itrs.net/Common/2004Update/2004Update.htm.

[10] K.-W. Kwon, H.-J. Lee, and R. Sinclair, “Solid-state amorphization at tetragonal-Ta/Cu interfaces”, Appl. Phys. Lett., vol. 75, no. 7, pp. 935-937, August 1999.

[11] JCPDS-ICDD Card File No. 35-0789.

[12] JCPDS-ICDD Card File No. 04-0836.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Vibrational Modes in $Ga_xMn_{1-x}Sb$ Studied by Raman Spectroscopy

M. M. Hasan[1,], M. R. Islam[1], N. F. Chen[2], and M. Yamada[3]*
[1]Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
[2]Laboratory of Semiconductor Materials Science, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China
[3]Department of Electronics, Kyoto Institute of Technology, Kyoto 606-8585, Japan
*E-mail:mahbub01@eee.kuet.ac.bd

Abstract- Raman scattering study on vibrational modes has been reported in ferromagnetic semiconductor $Ga_{1-x}Mn_xSb$ grown by Mn ions implantation, deposition, and post-annealing. The Raman experiments are performed in the implanted and unimplanted regions of the sample before and after etching. Only GaSb-like phonon modes are observed in the spectra measured from the unimplanted region. However, in addition to GaSb-like phonons, some extra modes are observed in the spectra measured from the implanted region of the sample. The experimental results demonstrate that the 115, 152, and 437 cm^{-1} modes are appeared due to surface defects and crystal disorder caused by Mn ions implantation and deposition processes. The origin of the weak structure observed approximately at 269 cm^{-1} is not so clear. However, the frequency position of MnSb-like LO phonon mode (266.4 cm^{-1}) determined by reduced-mass model is found to be close to the experimentally observed mode at 269 cm^{-1}. The phonon mode appeared approximately at 659 cm^{-1} is found to be associated with blackish layer formed on the surface of the sample from the annealing process and is assigned to Mn_3O_4-like. Furthermore, existence of coupled LO-phonon plasmon mode is found in the spectra measured from the implanted and close to the implanted regions of the sample.

I. Introduction

Ferromagnetic semiconductor $Ga_{1-x}Mn_xSb$ can integrate semiconducting properties with magnetic properties and is attractive for the development of new class spintronic devices due to its room temperature ferromagnetism [1-5]. For the last few years, the structural and room temperature ferromagnetic properties have been investigated by x-ray diffraction and vibrating sample magnetometer in this material [4,5]. The structural properties of this material grown by digitally doped Mn are also analyzed theoretically and experimentally using cross-sectional scanning tunneling microscope [6]. There is, however, still not much known about the details of vibrational properties of the $Ga_{1-x}Mn_xSb$ grown by Mn ion implantation, deposition and post-annealing.

In our recent study [7], Raman spectroscopic determination of Mn composition and strain has been reported in $Ga_{1-x}Mn_xSb$ prepared by liquid phase epitaxy, where only GaSb-like LO phonon mode was observed in the spectra. This is because the Mn composition of the samples was very low. Although few Raman scattering studies have been performed in GaN prepared by ion-implantation technique [8-10], to the best of our knowledge, no body has made Raman scattering study on the vibrational modes in $Ga_{1-x}Mn_xSb$ grown by Mn ions implantation and deposition, which is more useful in fabricating various devices. This article presents Raman scattering study on the vibrational properties of $Ga_{1-x}Mn_xSb$ single crystals. The samples were prepared by the Mn ions implantation, deposition, and post-annealing technique. To understand the optical phonon modes clearly, the samples were measured before and after etching. By correlating the Raman spectra obtained from the implanted and unimplanted surfaces before and after etching the samples, some additional phonon modes are identified. From the experimental observations, the origin of these modes is discussed in details.

II. Experimental Procedures

The zinc-blende $Ga_{1-x}Mn_xSb$ single crystals were prepared by a low-energy ion-beam deposition (LEIBD) system [5]. There are magnetic analyzers in the LEIBD system, with which the manganese can be purified as pure as isotope. First, the manganese ions with energy of 1 keV were implanted into a small region of an unintentionally doped p-type (001) oriented GaSb wafer in the depth of about 100 nm at 200 ^oC. Then, the manganese ions with energy of 100 eV were deposited on the surface of the wafer, which formed a thin layer about 5 nm thick. After the Mn-ions implantation and deposition, the wafers were annealed at 400 ^oC in an argon ambience for 30 minutes. A portion of the sample-surface was found to be blackish after annealing. To clean the sample-surface, it was etched using diluted H_3PO_4 for 15 minutes. The crystalline structure and room temperature magnetic properties of the $Ga_{1-x}Mn_xSb$ samples were studied [5] using x-ray diffraction and vibrating sample magnetometer, respectively. Energy dispersive x-ray was used to evaluate the Mn concentration. The maximum Mn

concentration was found to be $x = 0.1$ at the surface of the sample.

Raman experiments were performed at room temperature in the backscattering configuration employing the 514.5 nm line of an argon-ion laser. The laser beam was focused and then the scattered light was collected with a 20×objective lens. In order to eliminate elastic diffusion, a suitable notch-filter was used. The slit width was reduced to about 100 μm to prevent background noise. The laser power was low enough to prevent the local heating of the sample. The scattered light was dispersed with a Renishaw-2000 model spectrometer and detected with a cooled CCD detector.

III. Experimental results and Discussion

The first-order Raman scattering from a semiconductor typically shows longitudinal and transverse optical (LO and TO) phonon modes [11]. However, depending upon the crystal orientation and experimental geometry, one of them may be optically forbidden. The compound semiconductors such as $Ga_{1-x}In_xAs$ show two-mode behavior in the first-order Raman scattering where LO and TO phonon modes are found corresponding to binary end materials [12]. The appearance and frequency position of these phonons are found to dependent on the crystal orientation and In composition. Alike of $Ga_{1-x}In_xAs$ crystal, GaSb- and MnSb-like LO and TO phonon modes should be appeared in the first-order Raman spectra measured from the $Ga_{1-x}Mn_xSb$ crystals. Fig. 1 (i) and (ii) show some of the Raman spectra measured from the inside and outside of the implanted region before and after etching the sample. The sample schematic is shown in the inset of Fig. 1(ii) where the filled white circles represented by the letters a, b, and c indicate laser probing positions. It is found in Fig. 1 that the spectra measured from the position a before and after etching the sample show phonon modes only at 235.6 and 226 cm^{-1}, which are identified as GaSb-like LO and TO modes, respectively [9,13]. Besides the GaSb-like LO and TO modes, some additional modes are observed at about 115 and 152 cm^{-1} in the spectra measured from the positions b and c before etching the sample. These modes are found to be almost disappeared in the spectra measured after etching the samples as seen in Fig. 1 (ii).

Furthermore, the intensity of the GaSb-like phonon modes is found to be changed in Fig. 1 as moving the probing laser beam from the position a to c. The intensity of GaSb-like LO phonon is found to be screened in the spectra measured from the position b due to free carrier plasma, which couple with the LO phonon via their macroscopic electric fields. In our sample the free carriers are holes generated due to the incorporation of acceptor Mn into GaSb matrix. It is to be mentioned here that some portion of the sample-surface was blackish due to annealing. Some of the Raman spectra measured from the blackish and clean regions are shown in Fig. 2. The spectra are truncated from the top and plotted in high frequency range to observe the weak phonon modes at high frequency. Several spectra from the blackish and the clean regions are shown in Fig. 2 to confirm the reproducibility of the weak phonon modes. The spectra measured from the clean region show weak phonon modes at 269 and 437cm^{-1}. On the other hand, comparatively a strong phonon mode is observed at 659 cm^{-1} in the spectra measured from the blackish region.

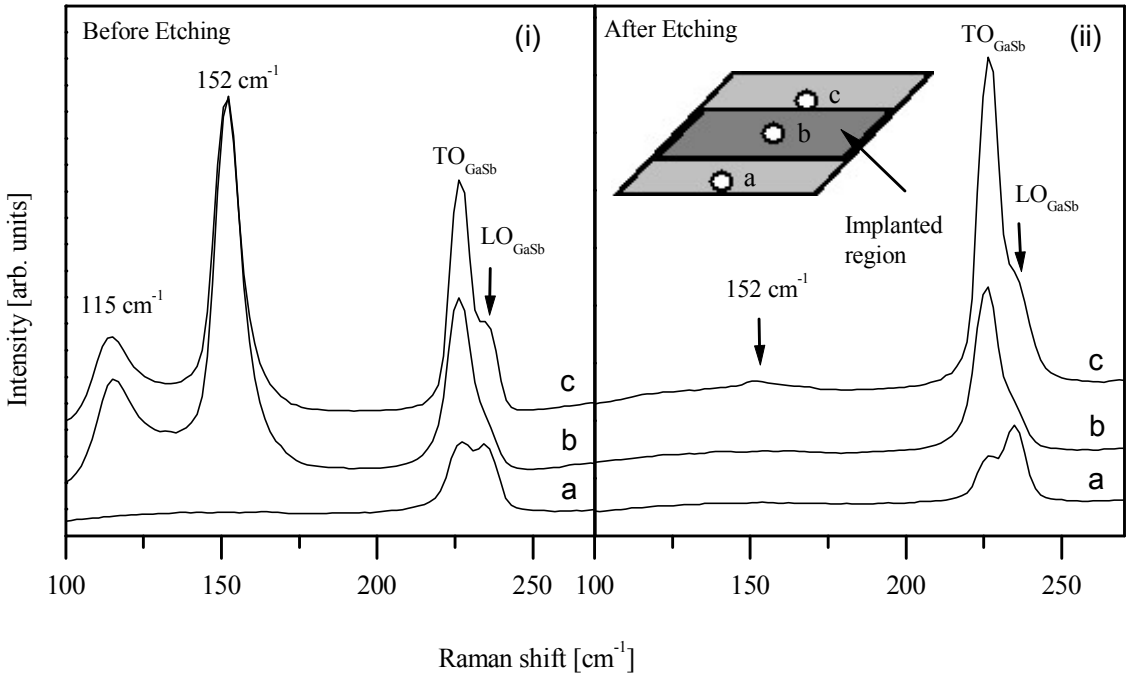


Fig. 1. Room temperature Raman spectra measured from $Ga_{1-x}Mn_xSb$ sample (i) before etching and (ii) after etching. Inset shows experimental schematic where filled circles indicate measurement points. The spectra designated by the letters a, b, and c are obtained from the corresponding points shown in the inset.

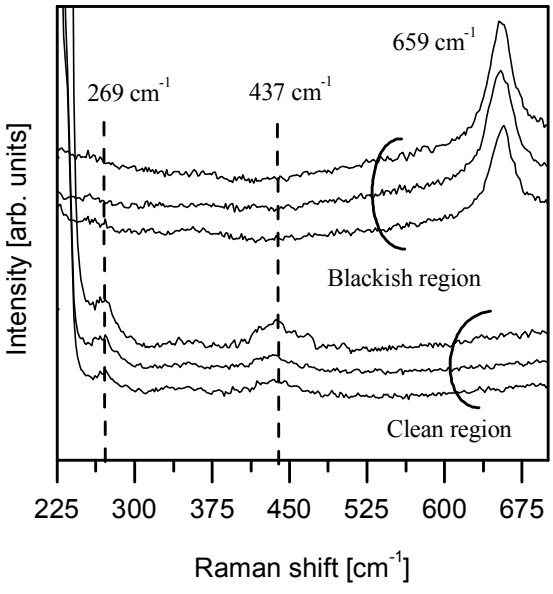


Fig. 2 Comparison of the Raman spectra measured from blackish and clean regions of the sample. Several spectra from both of these regions are presented to confirm the reproducibility of the phonon modes.

The 437 cm^{-1} mode is found to broad in the spectra and can be assigned to as disorder activated mode. The phonon dispersion may experience disorder due to the implantation of Mn atoms in GaSb matrix resulting in the appearance of disorder activated mode. The existence of disorder phonon mode was also found in the Raman spectra measured from GaN layers implanted with Mn, Ar, P, C, Mg, and Ca ions [8-10]. Similar broad structure was found at 440 cm^{-1} in a previous study [13] performed in GaSb epilayer doped with dimethylzinc and assign to as second order combinational mode. However, we think that the 440 cm^{-1} structure in their study is associated with crystal disorder induced from the doping process. As seen in Fig. 2, the 659 cm^{-1} mode is completely disappeared in the spectra measured from the clean region. This indicates that the appearance of this phonon mode is connected with the blackish layer. It is thought that manganese oxide may form on the surface of the sample due to annealing. Therefore, the phonon mode observed at 659 cm^{-1} can be assigned to the vibration of manganese oxide. In a previous study [15], the phonon mode observed in the Raman spectra at 560 cm^{-1} was suggested to be Mn_3O_4 like. The origin of the phonon modes observed at 115, 152 and 269 cm^{-1} is discussed below.

A. Phonon modes at 115 and 152 cm^{-1}

The origin of the phonon modes observed at 115 and 152 cm^{-1} is not so clear. It is anticipated that these modes are associated with Mn ions implantation with higher energy. It was reported [16] that the Raman spectra measured from GaN doped with Mg show low frequency phonon mode (132 cm^{-1}) and assigned to as local vibrational mode (LVM) of Mg in GaN matrix. As the concentration of Mn-ions in our samples is quite high, the LVM of Mn in the GaSb matrix should be considered. The frequency position of Mn LVM in the GaSb matrix can be estimated within the simple mass defect approximation [16] and found to be 245 cm^{-1}, which does not agree with the experimentally observed phonon modes at 115 and 152 cm^{-1}. To understand the origin of these modes, Raman experiments were further performed after etching the sample with the same schematic as shown in the inset of Fig. 1. The Raman spectra obtained from the etched sample are shown in Fig. 1 (ii). It is clearly observed from Fig. 1 (ii) that the 115 and 152 cm^{-1} phonon modes are almost disappeared after etching the sample. We therefore conclude that these modes are associated with the surface defects induced by Mn ions implantation and deposition processes.

B. Phonon mode at 269 cm^{-1}

Since Mn atom takes Ga site in GaSb lattice, the GaMnSb crystal should follow the $Ga_{1-x}Mn_xSb$ structure [6,17,18]. Raman spectra measured from ternary compound semiconductor $Ga_{1-x}In_xAs$ show GaAs-and InAs-like phonon modes [12]. Alike of $Ga_{1-x}In_xAs$, Raman spectra measured from $Ga_{1-x}Mn_xSb$ should show GaSb- and MnSb-like phonon modes. To best of our knowledge, the frequency positions of MnSb-like phonon modes are unknown. In order to extrapolate unknown phonon modes in ternary compound semiconductors, reduced-mass model was used [19]. Figure 3 shows the plots of TO and LO phonons of GaSb, AlSb and InSb crystals in dependence of $1/\sqrt{\mu}$, where μ is the reduced-mass of Ga, Al and In atoms. The frequency positions of MnSb-like TO and LO phonons are extrapolated from the simple fits as shown in Fig. 3 and found to be approximately at 253.2 and 266.4 cm^{-1}, respectively. Since the extrapolated MnSb-like mode (266.4 cm^{-1}) shows good agreement with the phonon peak experimentally observed at 269 cm^{-1}, the peak appeared in the spectra at 269 cm^{-1} may be assumed

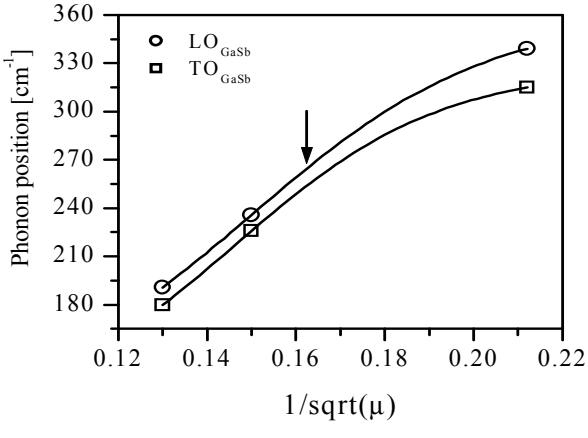


Fig. 3. The circles and squares represent LO and TO phonon frequencies, respectively, for GaSb, AlSb and InSb in dependence of $1/\sqrt{\mu}$. The LO and TO phonon frequencies for MnSb can be obtained from extrapolations of the fits.

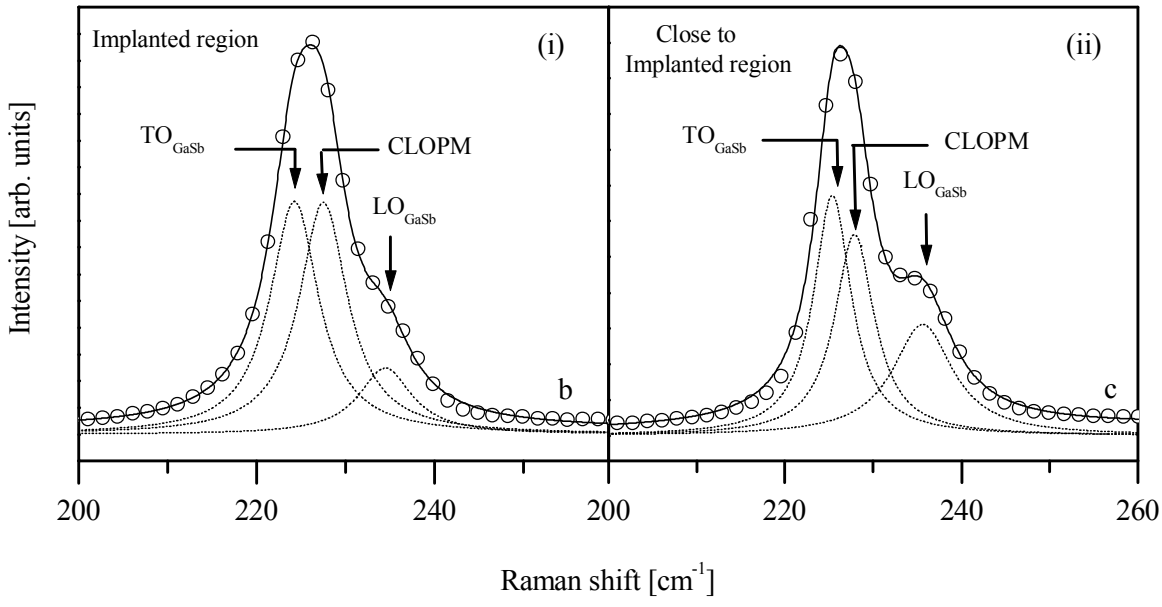


Fig. 4. Decomposition of Phonon modes by best fitting the Raman spectra measured from the (i) implanted and (ii) close to the implanted regions using Lorentzian function and a proper background. The open circles indicate experimental data points and the solid and dotted lines are obtained from lineshape analysis. The peak frequency and peak intensity of CLOPM changes due to change in hole density in these regions.

to be MnSb-like LO mode. But we are not certain at this stage. Similar phonon mode was found [13] in p-type GaSb epilayer grown on semi insulating GaAs substrate by vapor phase epitaxy and assign to as second order combinational mode. However, we think that this structure was appeared in their spectra from the GaAs substrate and its frequency position is the same as the TO_{GaAs} phonon [12].

C. Coupled-LO phonon Plasmon mode

In a polar semiconductor, an LO phonon and the plasmon formed by the free carriers interact via their macroscopic electric fields [20]. The frequency position of coupled LO-phonon plasmon mode (CLOPM) in $Ga_{1-x}Mn_xSb$ crystal varies from the frequency position of LO_{GaSb} to the TO_{GaSb} depending on the carrier concentration [20]. The existence of CLOPM should be found in the spectra b due to the implantation of acceptor Mn in GaSb matrix. The Mn atoms can be diffused from the implanted region due to its high diffusivity; consequently, CLOPM may also appear in the spectra c. In our case, the CLOPM is convoluted with LO_{GaSb} and TO_{GaSb} modes in the form of broad structures as seen in Figs. 4(i) and (ii). In order to decompose the convoluted phonon modes, the spectra b and c are best fitted using Lorentzian function and a proper background. The results are shown in Fig. 4 where open circles indicate experimental data points, solid lines are from lineshape fitting, and dotted lines represent each component of vibrational mode. It is found in Fig. 4 that the CLOPM appeared approximately at 227.66 and 228.72 cm^{-1} in the spectra b and c, respectively. The peak frequency position of the CLOPM demonstrates that hole concentration is more in the implanted region than close to the implanted region.

IV. Conclusion

The vibronic properties of ferromagnetic semiconductor $Ga_{1-x}Mn_xSb$ grown by Mn ions implantation and deposition have been studied using Raman spectroscopy. In order to understand various features in the spectra, we have made Raman experiments from both implanted and unimplanted regions of the samples before and after etching. By correlating the Raman spectra obtained under different measurement conditions, GaSb-like LO and TO phonons were found to appear approximately at 235.6 and 226 cm^{-1} in the spectra obtained from the implanted and unimplanted regions. In addition to GaSb-like phonons, some extra structures are found approximately at 115, 152, 269, 437, and 659 cm^{-1} in the spectra obtained from the implanted region. The 115 and 152 cm^{-1} modes, and 437cm^{-1} mode are found to be associated, respectively, with surface defects, and crystal disorder caused by Mn ions implantation and deposition processes. The origin of the 269 cm^{-1} mode is not so clear. However, the frequency position of MnSb-like LO mode (266.4 cm^{-1}) extrapolated by reduced-mass model is found to be close to the experimentally observed mode at 269 cm^{-1}. Comparing the spectra measured from the blackish and clean regions of the sample, it is confirmed that the 659 cm^{-1} mode is due Mn_3O_4 vibration. Furthermore, the existence of CLOPM is found in the spectra measured from the implanted and close to the implanted regions. The free carrier hole plasmon formed due to implantation

of acceptor Mn in GaSb matrix interact with LO phonon resulting in the appearance of CLOPM in the spectra.

References

[1] M. L. Roukes,"Electronics in a spin," Nature, vol 411, pp.747-748, June 2001.

[2] S. Datta and B. Das, "Electronic analog of the electro-optic modulator," Appl. Phys., Lett. vol.56, no.7, pp. 665-667, February 1990.

[3] Y. D. Park, B.T. Jonker, B. R. Bennet, G, Itzkos, M. Furis, G. Kioseoglou and A. Petrou, "Electrical spin injection across air-exposed epitaxially regrown semiconductor interfaces,"Appl. Phys. Lett., vol. 77, no. 24, pp. 3989-3991, December 2000.

[4] X. Chen, M. Na, M. Cheon, S. Wang, H. Luo, B. D. McCombe, "Above room temperature ferromagnetism in GaSb/Mn digital alloys," Appl. Phys. Lett., vol. 81, no. 3, pp. 511-513, july 2002.

[5] N. F. Chen, F. Q. Zhang, J. L. Yang, Z. K. Liu, S. Y. Yang, C. L. Chai, Z. G. Wang, W. R. Hu, and L. Y. Lin, "Room temperature ferromagnetic semiconductor $Mn_xGa_{1-x}Sb$," Chinese Science Bulletin, vol. 46, pp. 516-520, 2003.

[6] G. I. Boishin, J. M. Sullivan, and L. J. Whitman, "Structure of GaSb digitally doped with Mn," Phys. Rev. B, vol.71, no.19, pp.193307 (4 pages), May 2005.

[7] M. R. Islam, N. F. Chen, and M. Yamada, "Raman Scattering Characterization of Mn Composition and Strain in $Ga_{1-x}Mn_xSb$/GaSb epitaxial layers," Cryst. Res. Technol, DOI 10.1002/crat.2008XXXXX (2008) in press.

[8] W. Limmer, W. Riteter, R. Sauer, B. Mensching, C. Liu, and B. Rauscchenbach, "Raman scattering in ion-implanted GaN," Appl. Phys. Lett., vol.72, no.20, pp. 2589-2591, May 1998.

[9] M. R. Islam, N. F. Chen, and M. Yamada, "Raman scattering study on vibrational modes in $Ga_{1-x}Mn_xN$ prepared by Mn-ion implantation," Mater. Sci. in Semi. Processing, vol. 9, pp. 184-187, February 2006.

[10] V. Yu. Davydov, Yu. E. Kitaev, I. N. Gocharuk, A. N. Smirnov, J. Graul, O. Semchinova, D. Ufmann, M. B. Smirnov, A. P. Mirgorodsky, and R. A. Evarystov, "Phonon dispersion and Raman scattering in hexagonal GaN and AlN," Phys Rev. B, vol.58, no.19, pp. 12899-12907, June 1998.

[11] M. Cardona and G. Untherodt, *Light Scattering in Solids*, Springer-verlag, Berlin, 1983.

[12] M. R. Islam, P. Verma, M. Yamada, M. Tatsumi and K. Kinoshita, "Micro-Raman Characterization of Starting Material for Traveling Liquidus Zone Growth Method," Jpn. J. Appl. Phys., Vol. 41, pp. 991-995, August 2002.

[13] J. E. Maslar, W. S. Hurst and C. A. Wang, "Raman spectroscopic determination of hole concentration in p-type GaSb,"J.Appl.Phys., vol.103, no.1, pp. 013502 (11 pages), January 2008.

[15 K. W. Nam and K. B. Kim, "Manganese Oxide Film Electrodes Prepared by Electrostatic Spray Deposition for Electrochemical Capacitors,"Journal of the Electrochemical Society, vol.153, no.1, pp.A81-A88, November 2006.

[16] A. Hoffmann, A. Kaschner, C. Thomsen, "Local vibrational modes and compensation effects in Mg-doped GaN," phys. stat. sol. (c), vol.0, no.6, pp.1783-1794, April 2003.

[17] K. M. Yu, W. Walukiewicz, T. Wojtowicz, I. Kuryliszyn, X. Liu, Y. Sasaki, and J. K. Furdyna, "Effect of the location of Mn sites in ferromagnetic $Ga_{1-x}Mn_xAs$ on its Curie temperature," Phys. Rev. B, vol. 65, no.20,pp. 201303 (4 pages), April 2002.

[18] J. Masek, J. Kudrnovsky, and F. Maca, "Lattice constant in diluted magnetic semiconductors (Ga,Mn)As," Phys. Rev. B, vol. 67, no.15,pp.153203 (4 pages), April 2003.

[19] P. Verma, K. Oe, and M. Yamada, H. Harima, M. Herms, G. Irmer, "Raman studies on $GaAs_{1-x}Bi_x$ and $InAs_{1-x}Bi_x$,"Appl. Phys., vol. 89, no. 3, pp. 1657-1663, February 2001.

[20] R. Fukasawa and S. Perkowitz, "Raman-scattering spectra of coupled LO-phonon–hole-plasmon modes in p-type GaAs," Phys. Rev. B, vol.50, no.19, pp.14119-14124, July 1994.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Raman Spectroscopic Determination of Hole Density in Diluted Magnetic Semiconductor $Ga_xMn_{1-x}Sb$

M. M. Hasan[1,], M. R. Islam[1], N. F. Chen[2], and M. Yamada[3]*

[1]Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
[2]Laboratory of Semiconductor Materials Science, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China
[3]Department of Electronics, Kyoto Institute of Technology, Kyoto 606-8585, Japan
*E-mail:mahbub01@eee.kuet.ac.bd

Abstract- Raman scattering determination of hole density has been reported in diluted magnetic semiconductor $Ga_{1-x}Mn_xSb$ prepared by Mn ions implantation, deposition, and post annealing. The Raman spectra measured from the implanted region of the sample show coupled plasmon LO-phonon mode (CPLOM), which is found to be superimposed with GaSb-like phonon modes in the spectra. The spectral lineshapes are modeled using a dielectric function where interband and intraband transitions of free holes are included. In addition to CPLOM, the individual contribution arises from GaSb-like phonon modes are taken into account in the model. The hole density as a function of laser probing position is determined from the best fit parameters and is found to be in reasonable agreement with the results obtained from the electrochemical capacitance-voltage technique. Furthermore, the dependence of optical mobility, depletion width, and peak frequency of CPLOM with hole density is determined.

I. Introduction

In recent years, research interest has been highly concentrated in fabricating and investigating new semiconductor materials for the development of semiconductor spintronics technology [1-4]. Diluted magnetic semiconductors (DMSs) having room temperature ferromagnetic properties are required for spitronic devices. Compared with II-VI based DMSs, III-V based DMSs possess greater advantages, because doping control is far more difficult in II-VI than in III-V compounds. Most of the III-V based DMSs, the ferromagnetic phase transition occurs at maximum Curie temperature of less than 300 K, as reported so far [5,6]. $Ga_{1-x}Mn_xSb$ on the other hand is an attractive III-V based DMS for fabricating spintronic devices as it shows room temperature ferromagnetic properties [7-9]. The magnetic properties in $Ga_{1-x}Mn_xSb$ arise from the S = 5/2 Mn spin system. Since Mn acts as an acceptor in $Ga_{1-x}Mn_xSb$ system, the incorporation of Mn atoms into GaSb lattice sites generate high hole density. It is well established that the maximum Curie temperature of DMSs depends on hole density [6-9]. Therefore, determination of hole density is very much important for understanding ferromagnetic properties as well as electronic properties of $Ga_{1-x}Mn_xSb$ material.

Measurement of hole concentration is difficult in DMS by standard magneto-transport techniques (Hall measurement) due to anomalous Hall effect [10-13]. In addition, Hall measurements are not applicable to magnetically diluted samples that are insulating [10]. With increasing hole concentration the coupled plasmon-LO-phonon mode (CPLOM) moves from the LO to the TO frequency in the Raman spectra [12,13]. This leads to determine hole concentration by modeling the phonon lineshapes using coupled mode theory. Raman scattering determination of hole concentration has been studied in GaAs, GaSb, GaMnAs in the past [10-13]. To the best of authors knowledge, no body has made Raman scattering study on the hole concentration in DMS $Ga_{1-x}Mn_xSb$. This article presents Raman scattering study on the hole concentration in $Ga_{1-x}Mn_xSb$ single crystals prepared by the Mn ions implantation, deposition, and post annealing. To understand the optical phonon modes clearly, the sample was measured both from the implanted and unimplanted regions after etching. The hole concentration as a function of laser probing position is evaluated by analyzing the Raman lineshape using coupled-mode theory. The results are found to be within the range obtained by electrochemical capacitance-voltage (ECC-V) technique.

II. Experimental Procedures

The $Ga_{1-x}Mn_xSb$ single crystals studied in the present study were prepared by a low-energy ion-beam deposition (LEIBD) system [8]. The magnetic analyzers installed in the LEIBD system were used to purify manganese as pure as isotope. The samples were prepared in two-steps. In the first step, the manganese ions with energy of 1 keV were implanted into a small region of an unintentionally doped p-type (001) oriented GaSb wafer in the depth of about 100 nm at 200 °C. The deposition of manganese ions was done in the second step with energy of 100 eV, which formed a thin layer about 5 nm thick on the surface of the

wafer. After the Mn-ions implantation and deposition, the wafers were annealed at 400 ^{o}C in an argon ambience for 30 minutes. To remove various defects from the surface of the sample, it was etched using diluted H_3PO_4 for 15 minutes. The crystalline structure and hole density of the $Ga_{1-x}Mn_xSb$ samples were studied [8] by X-ray diffraction and ECC-V technique, respectively.

Raman experiments were performed at room temperature in the backscattering configuration employing the 514.5 nm line of an argon-ion laser. The laser beam was focused and then the scattered light was collected with a 20× objective lens. In order to eliminate elastic diffusion, a suitable notch-filter was used. The slit width was reduced to about 100 μm to prevent background noise. The laser power was low enough to prevent the local heating of the sample. The scattered light was dispersed with a Renishaw-2000 model spectrometer and detected with a cooled CCD detector.

III. Experimental Results and Discussion

Figure 1 shows some of the Raman spectra measured from the unimplanted, in the implanted and close to the implanted regions after etching the sample. The experimental schematic is shown in the inset where the white dotted line indicates laser probing direction. The spectra indicated by the letters a, b and f, and c to e are obtained from unimplanted, close to the implanted, and inside the implanted regions of the sample, respectively. It is found in Fig. 1 that the spectrum measured from the unimplanted region (spectrum a) show GaSb-like LO and TO phonon modes approximately at 235.6 and 226 cm^{-1} [10,11], respectively, in which LO phonon is found to be strong than TO phonon. However, the intensity of TO phonon starts increasing with decreasing the intensity of LO phonon in the spectra measured close to the implanted regions (spectra b and f) of the sample. Almost complete suppression of LO phonon with significant increasing of the intensity of TO phonon is found in the spectra measured from the implanted region (spectra c to e). It is also found in Fig. 1 that the position of GaSb-like phonons has not been shifted significantly in the spectra measured from these regions. In III-V ferromagnetic semiconductors, Mn+2 acts as an acceptor, generating free holes in the valance band [7,8]. The plasmon formed by free carriers coupled with LO phonon by their macroscopic field resulting in the appearance of CPLOM, whose frequency position in p-type semiconductors varies from LO to the TO depending on the hole concentration [12]. Therefore, lineshape analysis of the CPLOM including the individual contribution of GaSb-like phonons leads to determine hole concentration in the sample under investigation.

IV. Determination of Hole Density in $Ga_{1-x}Mn_xSb$

Two Raman active CPLOM $\omega+$ and $\omega-$ are observed in n-type semiconductors due to the coupling between LO

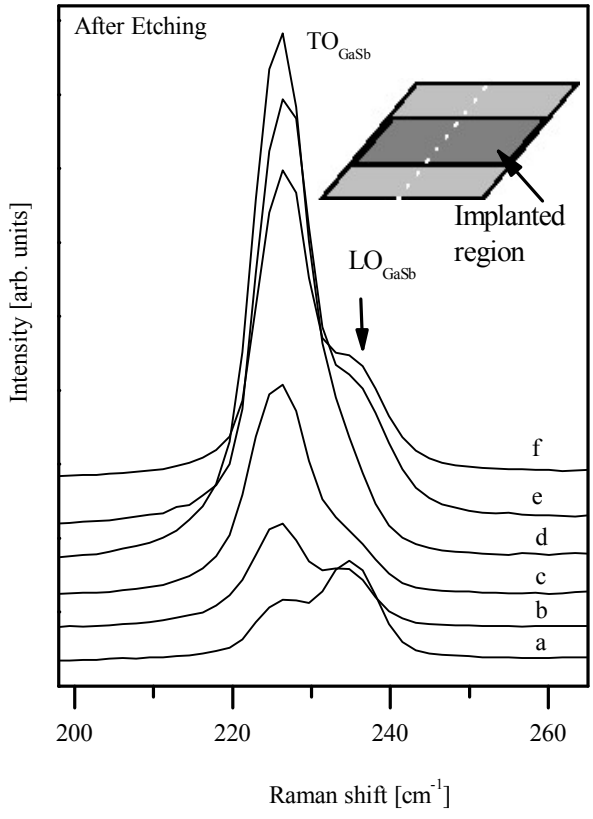


Fig. 1 Raman spectra measured from unimplanted, implanted and close to implanted regions after etching the $Ga_{1-x}Mn_xSb$ sample.

phonon and electron plasmon [14]. On the other hand, only one CPLOM is observed ($\omega+$ or $\omega-$) in p-type semiconductors due to a strong hole plasmon damping, moving from the LO to the TO frequency depending on the hole concentration [10-12]. According to the coupled mode theory, the frequencies of two CPLOM are given by [12]

$$\omega_{\pm}^{2} = \frac{1}{2}\{(\omega_{Lo}^{2} + \omega_{p}^{2}) \pm [(\omega_{Lo}^{2} + \omega_{p}^{2})^{2} - 4\omega_{p}^{2}\omega_{TO}^{2}]^{\frac{1}{2}}\}, \quad (1)$$

where ω_{Lo} , and ω_{To} are the LO and TO phonon frequencies. ω_p is the hole plasma frequency which can be defined by [9]

$$\omega_{p}^{2} = 4\pi p e^{2} / \varepsilon_{\infty} m_{h}^{*}, \quad (2)$$

where p , ε_{∞} , and m_h^* are the hole concentration, optical dielectric constant, and the effective mass of the free hole, respectively. The calculated $\omega_{\pm}$ frequencies of the coupled LO plasmon mode as a function of hole concentration are shown in Fig. 2 for $Ga_{1-x}Mn_xSb$ crystal. The material parameters used in this calculation are listed in Table 1. It is seen from Fig. 2 that upper mode frequency of $\omega+$ approximately coincide with LO phonon at low hole density and the lower mode frequency $\omega-$ coincide with TO phonon at high hole density. The upper and lower mode frequencies $\omega+$ and $\omega-$ are not observed in Raman spectra for high and low

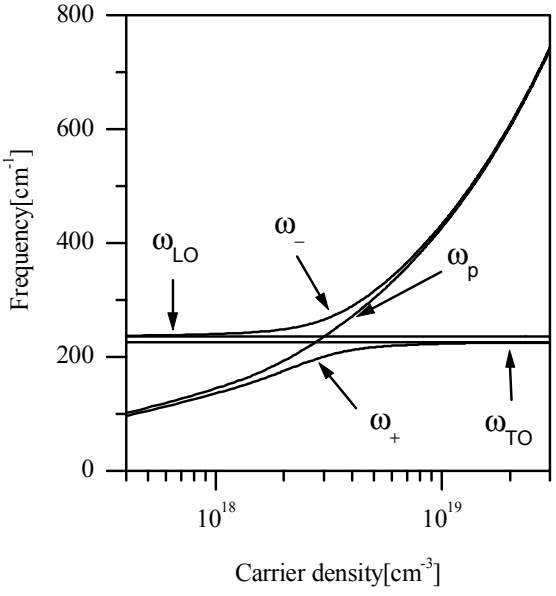


Fig. 2 The predicted frequencies of the coupled modes ($\omega+$ and $\omega-$) and the plasmon mode ω_p as function of carrier densities. ω_{Lo} and ω_{To} are the LO_{GaSb} and TO_{GaSb} phonon frequencies.

Table 1 The parameters values used in the Raman lineshape analysis are taken from Ref. [10].

Parameters	Value
LO-Phonon optical frequency, ω_{LO}	235.6 cm^{-1}
TO-Phonon optical frequency, ω_{TO}	226 cm^{-1}
High frequency dielectric constant, ε_∞	14.4
Faust-Henry coefficient, C	-0.23
Average hole effective mass, m_h^*	0.34m
Static dielectric constant, ε_0	15.7
Barrier height, φ	0.724eV

hole concentrations, respectively, as reported in the past [12]. Thus, modeling the Raman lineshapes leads to determine hole density using coupled mode theory.

In order to investigate the change in electronic and vibronic properties after Mn ions implantation and deposition, we have made a lineshape analysis of the Raman spectra. The lineshape of the CPLOM is modeled according to the formulations derived by Katayama and Murase [14]

$$g_T = g_1 + g_2 + g_3, \tag{3}$$

$$g_1 = \frac{C^2\omega^4{}_{To}}{(\omega^2{}_{Lo} - \omega^2{}_{To})}\hbar[\mathrm{n}(\omega)+1]\times \mathrm{Im}[(4\pi)^2 \chi_{ph}(\varepsilon_\infty + 4\pi\chi_{fc})/\varepsilon_\infty{}^2\,\varepsilon], \tag{4}$$

$$g_2 = \frac{2C\omega^2{}_{To}}{(\omega^2{}_{Lo} - \omega^2{}_{To})}\hbar[\mathrm{n}(\omega)+1]\times \mathrm{Im}[(4\pi)^2 \chi_{ph}/\varepsilon_\infty\varepsilon], \tag{5}$$

$$g_3 = \hbar[n(\omega)+1]\,\mathrm{Im}[-4\pi/\varepsilon], \tag{6}$$

where C is the Faust-Henry coefficient [10] and $n(\omega)$ is the Bose-Einstein factor. The total dielectric function ε for the phonon and the free charge system is given by [9,22]

$$\varepsilon = \varepsilon_\infty + 4\pi\chi_{ph} + 4\pi\chi_{fc}, \tag{7}$$

$$4\pi\chi_{ph} = \varepsilon_\infty \frac{\omega^2{}_{Lo} - \omega^2{}_{To}}{\omega^2{}_{To} - \omega^2 - i\omega\gamma}, \tag{8}$$

$$4\pi\chi_{fc} = -\varepsilon_\infty \frac{\omega^2{}_p}{\omega(\omega + i\Gamma_p)}, \tag{9}$$

where χ_{ph}, χ_{fc}, γ, and Γ_p are the ionic susceptibility, the free carrier susceptibility, the phonon damping constant, and the plasmon damping constant, respectively. In Eq. (3), g_1 and g_3 arise due to the modulation of the first-order Raman susceptibility, respectively, by the atomic displacement and by the macroscopic electric field associated with the coupled mode. g_2 is the cross correlation term between the atomic displacement and the macroscopic electric field [12]. Using Eq. (3), the peak frequency of the CPLOM in function of hole density is evaluated. The results are shown in Fig. 3. The peak frequency varies from 235.2 cm^{-1} to 227.0 cm^{-1} for the variation of the hole density from $2.18\times10^{18}cm^{-3}$ to 1.10×10^{20} cm^{-3}. This indicates that the peak frequency of the CPLOM shifted from LO_{GaSb} to TO_{GaSb} depending on the hole density.

Obviously, the Raman spectra shown in Fig.1 can not be modeled solely by the CPLOM band. Two further contributions correspond to the lineshape of GaSb-like

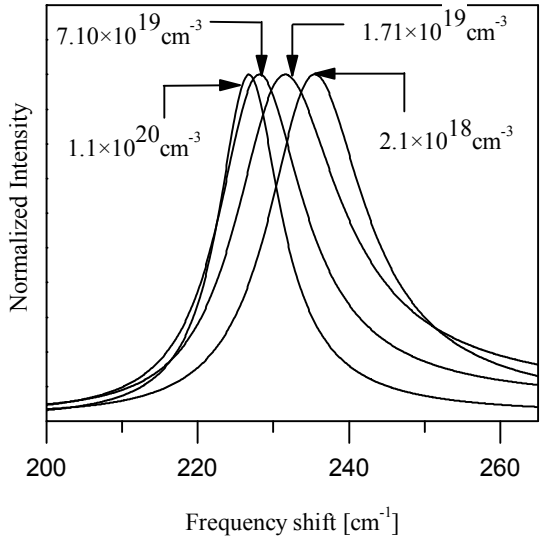


Fig. 3 Peak frequency variation of CPLOM evaluated with various hole densities using Eq. (3).

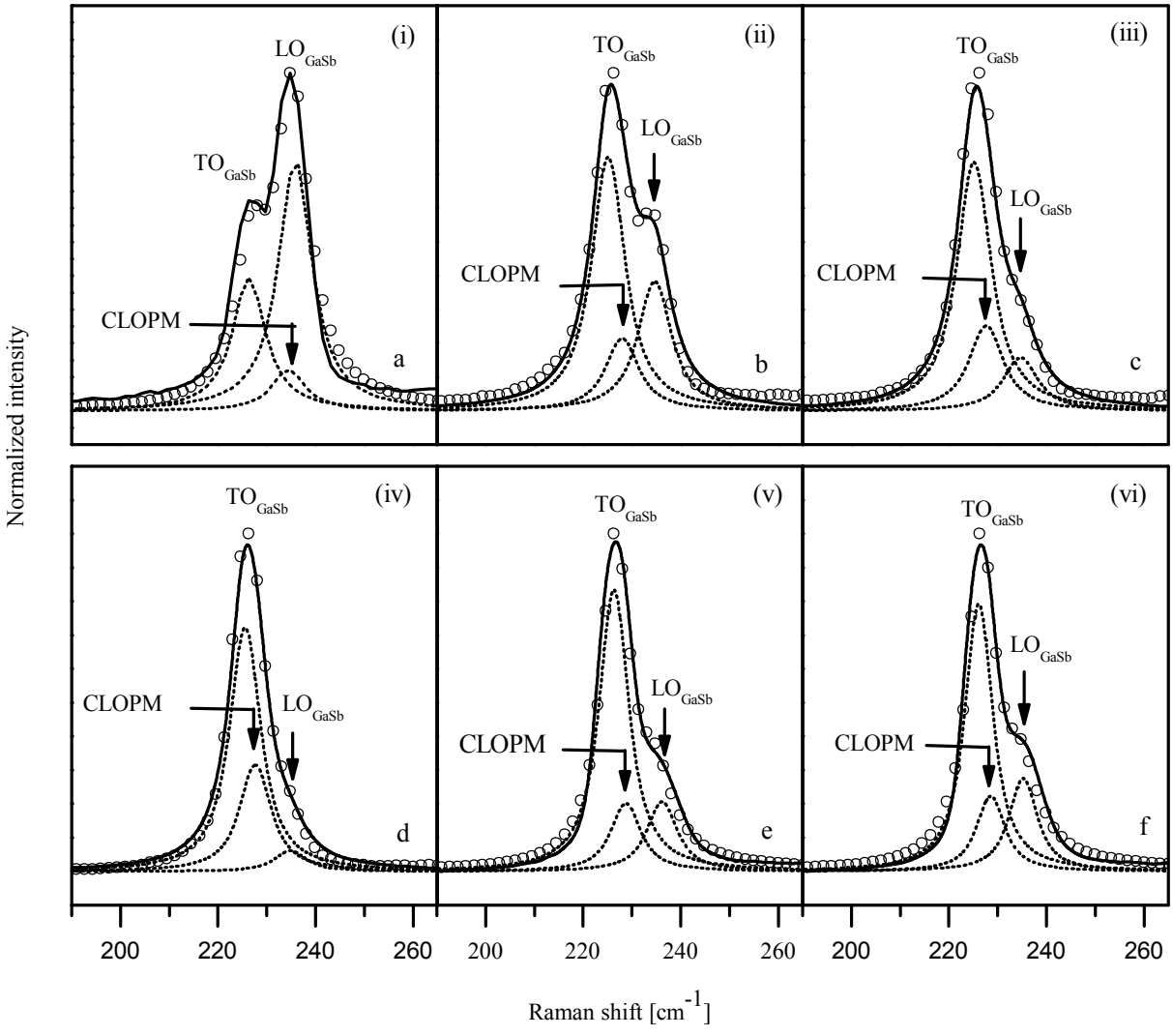


Fig. 4 Lineshape analysis of the Raman spectra shown in Fig. 1. The open circles and solid lines correspond to measured and calculated spectra, respectively. The dotted lines indicate individual phonon mode decomposed by lineshape modeling.

phonons are taken into account in the first-order Raman spectrum by the following formulation derived by Campbell and Fauchet [15]

$$I(\omega) = A\int \frac{|C(0,\vec{q})|^2 d\vec{q}}{[\omega - \omega(\vec{q})]^2 + (\Gamma_b)^2}, \qquad (10)$$

where $\omega(q)$ is the phonon dispersion function of the corresponding bulk material and $C(0,q)$ is the phonon confinement function, Γ_b is the half width half maximum of the phonon line in the bulk crystal (fitting value = 6 ~ 8cm^{-1}). The $C(0,q)$ can be given by [15]

$$|C(0,\vec{q})|^2 = \exp(-q^2 d^2_{av} / 4a^2) \qquad (11)$$

The wave vector $\vec{q}$ is expressed in unit of $2\pi/a$ (where a is the lattice constant= $6.09\times10^{-8} m$), d_{av} is the average diameter of the microcrystal (fitting value= $7.3\times10^{-8} m$). The average phonon dispersion in the bulk is $\omega(q) = \omega_0 - \Delta\omega\sin(q/4)$, ω_0 is the GaSb-like LO or TO phonon frequency, $\Delta\omega$ is the shift in GaSb-like LO or TO branches (fitting value ~ 1cm^{-1}). Using Eqs. (3) and (10), the lineshapes of the Raman spectra shown in Fig. 1 are analyzed taking into account of plasmon damping constant Γ_p , phonon damping constant γ , and plasma frequency ω_p as the main fitting parameters. The effect of other adjustable parameters ($\Delta\omega$, d_{av} , and Γ_b) on the phonon lineshapes is found to be not so significant. The parameters are adjusted until the best possible agreement between the calculated and the measured spectrum is achieved. The results are summarized in Figs. 4(i)-(vi) in which open circles and solid lines, respectively, correspond to the measured and calculated spectra. The dotted lines indicate individual phonon mode decomposed from the spectra. The frequency position of GaSb-like LO phonon is found to coincide with the CPLOM in the spectrum 'a' measured from outside the implanted region. Since this region was not implanted and it is away from

Table 2: Parameters for GaMnSb sample prepared by Mn ions implantation, deposition and post-annealing. The hole concentration p, optical mobility μ_{op}, depletion width d_L, phonon damping constant γ, hole-plasmon damping constant Γ_p, and peak frequency of coupled mode ω_R were evaluated by modeling the Raman spectra using coupled mode theory. The parameters are presented as a function of probing laser beam position.

Measured spectra	p (cm^{-3})	μ_{op} (cm^2/V s)	d_L (A^o)	γ (cm^{-1})	Γ_p (cm^{-1})	ω_R (cm^{-1})
a	7.82×10^{17}	54.56	399.6	3.6	500	235.60
b	1.34×10^{20}	24.60	30.50	5.2	1100	226.84
c	1.46×10^{20}	23.94	29.30	5.8	1140	226.77
d	1.60×10^{20}	22.75	27.98	7.0	1200	226.20
e	1.30×10^{20}	23.73	31.05	5.8	1150	226.90
f	1.10×10^{20}	24.80	33.70	5.0	1100	227.00

the implanted region, the hole concentration in this region is low. At lower hole concentration, the CPLOM and LO mode appeared in the same position which is clearly observed in Fig. 2. In contrast, position of the CPLOM shifted from the GaSb-like LO phonon mode in the spectra measured inside the implanted region as well as close to the implanted region due to the higher hole concentration in these regions. It was shown [16] that Mn atom diffuses from interstitial position to nearby region due to annealing. Therefore, the results obtained from close to the implanted regions caused by the diffusion of Mn atoms from the implanted region. In order to determine the hole concentration, the value of ω_p is determined from the best fit condition of the experimental spectra. Using Eq. (2), the hole concentration is determined corresponds to each probing position. We have also calculated depletion width, d_L and optical mobility, μ_{op} using $d_L = (\varepsilon_0 \varphi / 2\pi e^2 p)^{1/2}$ and $\mu_{op} = e / m^*_h \Gamma_p$ [12]. The results are listed in Table 2. Using ECC-V technique, the hole density along the thickness of the sample was measured [8]. The carrier concentration near the surface of the sample was about 1×10^{21} cm^{-3}, and it drops abruptly to 1×10^{19} cm^{-3} within a depth of 70 nm. Then the carrier concentration drops flatly to about 9×10^{17} cm^{-3} in the depth of 4 μm. It is found from Table 2 that the hole density evaluated from the implanted, close to the implanted as well as unimplanted regions are within in the range measured by the ECC-V technique.

It is well established that the shift in optical phonons in Raman scattering can be induced by the composition as well as by the strain in ternary compound semiconductors [17]. In our recent study [18], composition-dependent as well as strain-induced shift was found in GaSb-like LO phonon measured from $Ga_{1-x}Mn_xSb$/GaSb layers prepared by liquid phase epitaxy. However, coupled mode was not appeared in the spectra which may be due to the lower Mn concentration in the samples. The Mn composition-dependent shift in optical phonons was also found [10,19] in $Ga_{1-x}Mn_xAs$ prepared by molecular beam epitaxy and Mn ions implantation, deposition and post annealing. In the present study, we did not find remarkable shift in optical phonons as a function of probing position. The causes behind this are not clear at this stage. However, our previous results [18] indicate that the shift in optical phonons is smaller in strained $Ga_{1-x}Mn_xSb$ compared to unstrained $Ga_{1-x}Mn_xSb$. Since $Ga_{1-x}Mn_xSb$ layer was formed due to the Mn ions implantation and deposition processes, the existence of strain can be considered in the samples under investigation. It can be conceived that with the combined influence of strain and coupled mode, the GaSb-like phonon peaks may not show significant shift in the measured spectra. Since GaSb-like LO and TO phonon positions are found to be almost constant as a function of probing position, the parameters used in the present study correspond to GaSb material.

V. Conclusion

Raman spectroscopy has been used to determine hole density in DMS $Ga_{1-x}Mn_xSb$ prepared by Mn ions implantation, deposition and post annealing. The Raman spectra were recorded from both implanted and unimplanted regions after etching the samples. By comparing the Raman results obtained from the implanted and unimplanted regions, the existence of CPLOM in addition to GaSb-like phonons was found in the spectra. The intensity of these phonon modes found to have dependence with laser probing position due to the change in hole density in the sample. The spectral lineshapes were modeled using coupled mode theory taking into account of plasmon damping constant, phonon damping constant, and plasma frequency as the main fitting parameters. The hole density is evaluated under best fitting conditions and found that there is a variation of hole density in the implanted region as well as close to the implanted region of the sample. The typical hole density evaluated to be 7.82×10^{17}, 1.6×10^{20}, and 1.1×10^{20} outside the implanted, in the implanted and close to the implanted regions of the sample, respectively. The results obtained in the present study are compared with those obtained from ECC-V technique and found to be in reasonable

agreement. Furthermore, the optical mobility, depletion width, and peak frequency of the coupled mode are determined as a function of probing position.

References

[1] M. L. Roukes, "Electronics in a spin," Nature, vol.411, pp.747-748, June 2001.

[2] H. Ohno, H. Munekata, S. Molnar, and L. L.Chang, "New III-V diluted magnetic semiconductors," J. Appl. Phys, vol. 69, no.8, pp. 6103-6108, April 1991.

[3] S. Datta and B. Das, "Electronic analog of the electro-optic modulator," Appl. Phys. Lett., Vol.56, no.7,pp. 665-667, February 1990.

[4] Y. D. Park, B.T. Jonker, B. R. Bennet, G, Itzkos, M. Furis, G. Kioseoglou and A. Petrou, "Electrical spin injection across air-exposed epitaxially regrown semiconductor interfaces,"Appl. Phys. Lett., vol. 77, no. 24, pp. 3989-3991, December 2000.

[5] F. Matsukura, H. Ohno, A. Shen, and Y.Sugawara, "Transport properties and origin of ferromagnetism in (Ga, Mn)As," Phys. Rev. B, vol.57,no.4, pp. R2037-R2040, January 1998.

[6] T. Hayashi, M. Tanaka,T. Nishinaga, H. Shimada, "Magnetic and magnetotransport properties of new III-V diluted magnetic semiconductors: GaMnAs," J. Appl. Phys. , vol.81, no.8, pp. 4865-4867, April 1997.

[7] X. Chen, M. Na, M. Cheon, S. Wang, H. Luo, B. D. McCombe, "Above room temperature ferromagnetism in GaSb/Mn digital alloys," Appl. Phys. Lett., vol. 81, no. 3, pp. 511-513, july 2002.

[8] N. F. Chen, F. Q. Zhang, J. L. Yang, Z. K. Liu, S. Y. Yang, C. L. Chai, Z. G. Wang, W. R. Hu, and L. Y. Lin, "Room-temperature ferromagnetic semiconductor $Mn_xGa_{1-x}Sb$," Chinese Science Bulletin, vol. 46, pp. 516-520, 2003.

[9] G. I. Boishin, J. M. Sullivan, and L. J. Whitman, "Structure of GaSb digitally doped with Mn," Phys. Rev. B, vol.71, no.19, pp.193307 (4 pages), May 2005.

[10] M. J. Seong, S. H. Chun, H. M. Cheong, N. Samarth, A. Mascarenhas, "Spectroscopic determination of hole density in the ferromagnetic semiconductor $Ga_{1-x}Mn_xAs$," Phys. Rev. B, vol.66, pp.033202 (4 pages), July 2002.

[11] W. Limmer, M. Glunk, S. Mascheck, A. Koeder, D. Klarer, W. Schoch, K. Thonke, R. Sauer, and A. Waag, "Coupled plasmon–LO-phonon modes in $Ga_{1-x}Mn_xAs$," Phys. Rev. B vol.66, no.20, 205209(6 pages), November 2002.

[12] R. Fukasawa and S. Perkowitz, "Raman-scattering spectra of coupled LO-phonon–hole-plasmon modes in p-type GaAs," Phys. Rev. B, vol.50, no.19, pp.14119-14124 , July 1994.

[13] J. Masek, J. Kudrnovsky, and F. Maca, "Lattice constant in diluted magnetic semiconductors (Ga,Mn)As," Phys. Rev. B, vol. 67, no.15, pp.153203(2003) (4 pages), April 2003.

[14] S. Katayama and K. Murase, "Raman Scattering by Coupled LO Phonon-Plasmon Mode in *n*-GaAs," J. Phys. Soc. Jpn., vol. 42, pp. 886-894, march 1977.

[15] I. H. Campbell and P. M. Fauchet, "The effects of microcrystal size and shape on the one phonon Raman spectra of crystalline semiconductors," Solid State Commn.,Volume 58, no.10, pp. 739-741, June 1986.

[16] K. M. Yu, W. Walukiewicz, T. Wojtowicz, I. Kuryliszyn, X. Liu, Y. Sasaki, and J. K. Furdyna, "Effect of the location of Mn sites in ferromagnetic $Ga_{1-x}Mn_xAs$ on its Curie temperature," Phys. Rev. B, vol. 65, no.20, pp.201303 (4 pages), April 2002.

[17] M. R. Islam, P. Verma, M. Yamada, S. Kodama, Y. Hanaue, and K. Kinoshita, "The influence of residual strain on Raman scattering in $In_xGa_{1-x}As$ single crystals," Mater. Sci. and Eng., Vol. B B91-92, pp. 66-69, 2002.

[18] M. R. Islam, N. F. Chen, and M. Yamada, "Raman Scattering Characterization of Mn Composition and Strain in $Ga_{1-x}Mn_xSb$/GaSb epitaxial layers," Cryst. Res. Technol. ,DOI 10.1002/crat.2008XXXXX (2008) in press.

[19] M. R. Islam, N. F. chen, and M. Yamada, "Raman scattering study on diluted magnetic semiconductor $Ga_{1-x}Mn_xAs$ prepared by Mn-ion implantation," in Proc. XIIIth SIMC 2004, Beijing, China, IEEE Catalog No. 04CH37383, pp.185-188., 20-25 September, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Modeling and Numerical Analysis of Thermal Treatment of Granulated Porous Particles by Induction Plasma

M. Mofazzal Hossain[1], Y. Yao[2], M. Rafiqul Alam[3], M. Maksud Alam[3] and T. Watanabe[4]*

[1]Department of Electronics & Communications Engineering, East West University, Dhaka
[2] Kunming University of Science & Technology, Kunming 650093, Yunnan Province, China
[3]Chittagong University of Engineering & Technology, Chittagong-4349
[4]Tokyo Institute of Technology, G1-22, 4259 Nagatsuta, Yokohama 226-8502, Japan

*E-mail: dmmh@ewubd.edu

Abstract - In this paper it is aimed to describe the modeling and numerical analysis of thermal treatment of granulated porous particles by induction plasma. To investigate the heat exchange dynamics between plasma and particles during the flight of granulated porous particles through the hot plasma, a plasma-particle interactive flow model has been developed. This model solves the conservation equations to predict the temperature and flow fields of plasma, under local thermal equilibrium (LTE) conditions, and then computes the injected particles trajectories, temperature and size histories, and the particle source terms to incorporate the particle loading effects. It is found that the size and dose of injected particles greatly affect the particle trajectory and temperature, and hence the heat transfer to particles at higher powder feed-rate.

I. Introduction

Induction thermal plasma (ITP) has extensively been used for the synthesis and treatment of micro-particles since couple of decades as a clean reactive heat source [1]. Thermal plasma synthesis offers a versatile, cost-effective technology for the industrial-scale production of many advanced materials for demanding applications in high tech industries. The need for materials with improved physical and mechanical properties for demanding applications has gradually been increasing in such high tech industries as electronics, transportation, glass and nuclear power. ITP technology ensures essentially the in-flight, single-step, short time, and less pollution compared with the traditional technologies that have been using in the industries for the thermal treatment of granulated porous micro-particles. The thermal treatment of injected particles depends mainly on the plasma-particle heat exchange efficiency, which in turn depends to a large extent on the plasma temperature, particles trajectories, temperature and diameter histories. Experimentally, it is quite difficult to measure the particles trajectories, temperature and diameter histories during the flight through the hot plasma. Thus, the prediction of particles trajectories, temperature and diameter histories through numerical modeling is the prime concern. Boulos [2] developed a model and discussed the plasma-particle heat exchange dynamics for argon plasma. However, the particle porosity and its consequences during flight of the particle in the plasma were not discussed. The aim of this work is to demonstrate the modeling and simulation of the plasma-particle interactive flow in argon-oxygen plasma, for highly porous granulated micro-particles to optimize the discharge parameters that affect the treated particles diameter and compositions.

II. Modeling

Induction plasma is usually generated by radio frequency alternating current supply through some coils that surround a coaxial quartz tube and injecting plasma gas such as argon and sometimes mixtures of argon and molecular gases like oxygen, hydrogen, nitrogen etc. The discharged plasma flow can be treated as neutral fluid flow, although plasma itself is ionized, but it is electrically neutral. Again, for the thermal treatment of micro-particles, particles are injected into the plasma torch along with the carrier gas. Injected particles, penetrate the hot core of plasma, interact with that and exchange energy with plasma and get thermal treatment. Finally, treated particles are collected at the reaction chamber. Thus, to characterize the plasma-particle interaction behavior and to predict the plasma and particle parameters, that affect the size and morphology of the treated particles, we need to develop the plasma model and particle model, and correlate their interactions.

A. Plasma Model

The schematic geometry and dimensions of a typical ITP torch is presented in Fig.1 and Table 1 respectively. It is assumed that the plasma flow is 2-dimensioanl, axi-symmetric, laminar, steady, optically thin, and electromagnetic fields are 2-dimensional. With these assumptions the plasma flow is modeled by the conservation equations (1)-(4) and vector potential form of Maxwell's equation (5) [3]. This model solves the

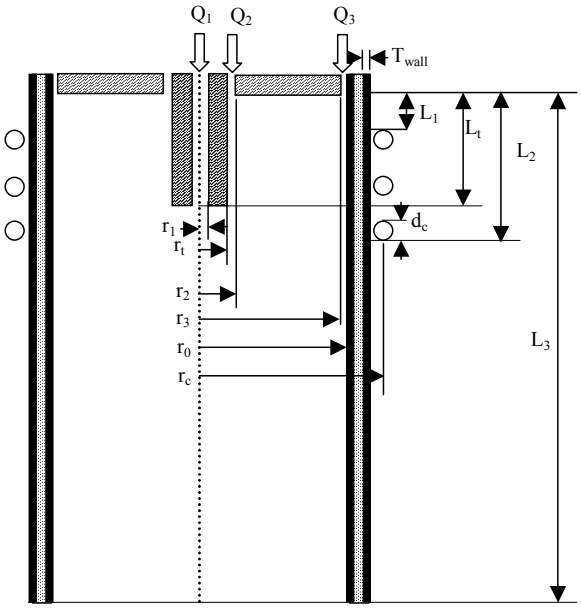


Fig.1 Schematic geometry and dimensions of ITP torch

conservation equations and vector potential form of Maxwell's equations simultaneously under LTE conditions, including a metal tube inserted into the torch. Adding the source terms to the conservation equations, the plasma-particle interaction and loading effects of particles injected into the torch, has been taken into account.

Mass conservation:

$$\nabla \cdot \rho \mathbf{u} = S_p^C \tag{1}$$

Momentum conservation:

$$\rho \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \nabla \cdot \mu \nabla \mathbf{u} + \mathbf{J} \times \mathbf{B} + S_p^M \tag{2}$$

Energy conservation:

$$\rho \mathbf{u} \cdot \nabla h = \nabla \cdot \left(\frac{\kappa}{C_p} \nabla h \right) + \mathbf{J} \cdot \mathbf{E} - Q_r - S_p^E \tag{3}$$

Species conservation:

$$\rho \mathbf{u} \cdot \nabla y = \nabla \cdot \left(\rho D_m \nabla y \right) + S_p^C \tag{4}$$

Vector potential form of Maxwell field equation [3]:

$$\nabla^2 \mathrm{A_c} = i \mu_0 \sigma \omega \mathrm{A_c} \tag{5}$$

Where, ∇: vector operator, $\mathbf{u}$: velocity vector, ρ: mass density, μ: viscosity, σ: electrical conductivity, κ: thermal conductivity, h: enthalpy, p: pressure, C_p: specific heat at constant pressure, D_m: multicomponent diffusion coefficient, y: mass fraction, $\mathbf{J}$: current density vector, $\mathbf{E}$: electric field vector, $\mathbf{B}$: magnetic field vector, Q_r: volumetric radiation loss, $\mathrm{A_c}$: complex amplitude of vector potential, μ_0: permeability of free space, ω: $2\pi f$ (f: frequency), i: complex vector ($\sqrt{-1}$). The particle source terms S_p^C, S_p^M and S_p^E are the contributions of particles to the mass and species, momentum and energy conservation equations respectively.

Table 1: Torch dimensions

Distance to initial coil position (L_1)	19 mm
Length of injection tube (L_t)	52 mm
Distance to end of coil position (L_2)	65 mm
Torch length (L_3)	190 mm
Coil diameter (d_c)	5 mm
Wall thickness of quartz tube (T_{wall})	1.5 mm
Inner radius of injection tube (r_1)	1 mm
Outer radius of injection tube (r_t)	4.5 mm
Outer radius of inner slot (r_2)	6.5 mm
Inner radius of outer slot (r_3)	21.5 mm
Torch radius (r_0)	22.5 mm
Coil radius (r_c)	32 mm

The boundary conditions, thermodynamic and transport properties of argon and oxygen gases are the same as those described in our previous work [4].

B. Particle Model

The following assumptions are made in the analysis of plasma-particle interactions: (i) the particle motion is two-dimensional; (ii) only the viscous drag force and gravity affect the motion of an injected particle; (iii) the temperature gradient inside the particle is neglected; (iv) the particle charging effect caused by the impacts of electrons or positive ions is negligible; (v) the electromagnetic drag forces caused by the particle charging of the injected particles are negligible compared with those by neutrals and charged particles. Thus, the momentum equations for a single spherical particle injected vertically downwards into the plasma torch can be expressed by equations (6), (7) and (8).

$$\frac{du_p}{dt} = -\frac{3}{4} C_D (u_p - u) U_R \left(\frac{\rho}{\rho_p d_p} \right) + g \tag{6}$$

$$\frac{dv_p}{dt} = -\frac{3}{4} C_D (v_p - v) U_R \left(\frac{\rho}{\rho_p d_p} \right) \tag{7}$$

$$U_R = \sqrt{(u_p - u)^2 + (v_p - v)^2} \tag{8}$$

The particle temperature, liquid fraction and particle diameter are predicted according to the energy balance equations (9), (10), (11) and (12).

$$Q = \pi d_p^2 h_c \left(T - T_p \right) - \pi d_p^2 \sigma_s \varepsilon \left(T_p^4 - T_a^4 \right) \tag{9}$$

$$\frac{dT_p}{dt} = \frac{6Q}{\pi \rho_p d_p^3 C_{pp}}, \quad T_p < T_b \tag{10}$$

$$\frac{dx}{dt} = \frac{6Q}{\pi \rho_p d_p^3 H_m}, \; 1000 \le T_p \le 1600 \tag{11}$$

$$\frac{dd_p}{dt} = -\frac{2Q}{\pi \rho_p d_p^2 H_v}, \; 1000 \le T_p \le 1600, \; T_p \ge T_b \tag{12}$$

Where u_p: axial velocity component of particle, v_p: radial velocity component of particle, g: acceleration of gravity,

ρ_p: particle mass density, d_p: particle diameter, Q: the net heat exchange between the particles and its surroundings, T_p and T_b: particle temperature and boiling point temperature, respectively, T: plasma temperature, T_a: ambient temperature, ε: particle surface emissivity; σ_s: Stefan-Boltzmann constant, C_{pp}: particle specific heat, H_m and H_v: latent heat of particle melting and vaporization respectively, and x: the liquid mass fraction of the particle. Drag coefficient C_{Df} is calculated using equation (13) and the property variation at the particle surface layer and the non-continuum effects are taken into account by equations (14) and (15) respectively [5].

$$C_{Df} = \begin{cases} \dfrac{24}{R_e} & R_e \le 0.2 \\ \dfrac{24}{R_e}\left(1+\dfrac{3}{16}R_e\right) & 0.2 < R_e \le 2.0 \\ \dfrac{24}{R_e}\left(1+0.11R_e^{\,0.81}\right) & 2.0 < R_e \le 21.0 \\ \dfrac{24}{R_e}\left(1+0.189R_e^{\,0.62}\right) & 21.0 < R_e \le 200 \end{cases} \quad (13)$$

$$f_1 = \left(\frac{\rho_\infty \mu_\infty}{\rho_s \mu_s}\right)^{-0.45} \quad (14)$$

$$f_2 = \left\{1+\left(\frac{2-\alpha}{\alpha}\right)\left(\frac{\gamma}{1+\gamma}\right)\frac{4}{Pr_s}Kn\right\}^{-0.45}, \; 10^{-2}<Kn<0.1 \quad (15)$$

$$C_D = C_{Df} f_1 f_2 \quad (16)$$

To take into account the steep temperature gradient between plasma and particle surface, the Nusselt correlation can be expressed by equation (17) [6]. The non-continuum effect is taken into account by equation (18) [5].

$$Nu_f = (2.0+0.6R_{ef}^{1/2}Pr_f^{1/3})\left(\frac{\rho_\infty \mu_\infty}{\rho_s \mu_s}\right)^{0.6}\left(\frac{C_{p\infty}}{C_{ps}}\right)^{0.38} \quad (17)$$

$$f_3 = \left\{1+\left(\frac{2-\alpha}{\alpha}\right)\left(\frac{\gamma}{1+\gamma}\right)\frac{4}{Pr_s}Kn\right\}^{-1}, \; 10^{-3}<Kn<0.1 \quad (18)$$

The convective heat transfer coefficient is predicted by equation (19).

$$h_{cf} = \frac{\kappa_f}{d_p} Nu_f f_3 \quad (19)$$

In the above expressions, subscript f, ∞ and s refer to properties corresponding to the film temperature (arithmetic mean of plasma and particle temperatures), plasma temperature and particle temperature respectively, C_p: particle specific heat, h_c: heat transfer coefficient, Nu: Nusselt number, Pr: Prandtl number, Re: Reynold number, $\propto$: thermal accommodation coefficient, γ: specific heat ratio and Kn: Knudsen number. The physical properties of soda-lime glass powders are: mass density 2300 kg-m^{-3}, specific heat 800 J-kg^{-1}K^{-1}, surface emissivity 80%, fusion and boiling temperature 1000~1600 K and 2500 K respectively, heat of fusion, and vaporization 3.69×10^5 J-kg^{-1} and 1.248×10^7 J-kg^{-1}, respectively.

C. Particle Source Terms

To take into account the particle loading effects the particle source terms for the mass, momentum, energy and species conservation equations have been calculated using the PSI-Cell (Particle-Source-In Cell) approach [7], where the particles are regarded as sources of mass, momentum and energy.

Let us assume N_t^0 is the total number of particles injected per unit time, n_d is the particle size distribution, and n_r is the fraction of N_t^0 injected at each point through the injection nozzle. Thus, the total number of particles per unit time traveling along the trajectory (l,k) corresponding to a particle diameter d_l injected at the inlet point r_k can be expressed by equation (20).

$$N^{(l,k)} = n_{d_l} n_{r_k} N_t{}^0 \quad (20)$$

The source terms in the mass and species conservation equation, S_p^C is the net flux rate of particles mass in a computational cell (control volume). Assuming the particles are spherical, the flux rate of particle mass for the particle trajectory (l,k) that traverses a given cell (i, j) can be expressed by equation (21).

$$S_{p,ij}^{C(l,k)} = \frac{1}{6}\pi\rho_p N_{ij}^{(l,k)}(d_{ij,in}^3 - d_{ij,out}^3) \quad (21)$$

The net flux rate of particle mass is obtained by summing over all particles trajectories which traverse a given cell (i, j) by equation (22).

$$S_{p,ij}^C = \sum_l \sum_k S_{p,ij}^{C(l,k)} \quad (22)$$

The source terms for axial and radial momentum conservation equations are evaluated in the same fashion as that of mass conservation equation. In this case, the flux rate of particles momentum for the particle trajectory (l,k) traversing a given cell (i, j) is expressed by equations (23) and (24).

$$S_{p,ij}^{M_z(l,k)} = \frac{1}{6}\pi\rho_p N_{ij}^{(l,k)}(u_{ij,in}d_{ij,in}^3 - u_{ij,out}d_{ij,out}^3) \quad (23)$$

$$S_{p,ij}^{M_r(l,k)} = \frac{1}{6}\pi\rho_p N_{ij}^{(l,k)}(v_{ij,in}d_{ij,in}^3 - v_{ij,out}d_{ij,out}^3) \quad (24)$$

Thus, the corresponding source terms for axial and radial momentum conservation equations can be expressed by equations (25) and (26) respectively.

$$S_{p,ij}^{M_z} = \sum_l \sum_k S_{p,ij}^{M_z(l,k)} \quad (25)$$

$$S_{p,ij}^{M_r} = \sum_l \sum_k S_{p,ij}^{M_r(l,k)} \quad (26)$$

The source term for energy conservation equation $S_{p,ij}^{E}$ consists of the heat given to the particles $Q_{p,ij}^{(l,k)}$, and superheat to bring the particle vapors into thermal equilibrium with the plasma $Q_{v,ij}^{(l,k)}$, and these terms are expressed by equations (27), (28) and (29).

$$Q_{p,ij}^{(l,k)} = \int_{\tau_{in}}^{\tau_{out}} \pi d_p^{\,2} h_c (T_{ij} - T_{p,ij}^{(l,k)}) dt \quad (27)$$

$$Q_{v,ij}^{(l,k)} = \int_{\tau_{in}}^{\tau_{out}} \frac{\pi}{2} \pi d_p^{\,2} \rho_p \left(\frac{dd_p}{dt} \right) C_{pv} (T_{ij} - T_{p,ij}^{(l,k)}) dt \quad (28)$$

$$S_{p,ij}^{E} = \sum_l \sum_k N_{ij}^{(l,k)} (Q_{p,ij}^{(l,k)} + Q_{v,ij}^{(l,k)}) \quad (29)$$

D. Computation Methodology

For the sake of computation, the particle concentration in the inlet is assumed to be uniform and to be separated into five injection points, which are at radial positions of 0.3, 0.45, 0.6, 0.75 and 0.9 mm from the torch centerline. In the present computation the particles diameter distribution is assumed to be Maxwellian, and the powder is assumed to be composed of seven size particles according to its diameter and deviation. The average particle diameter is 58 μm and the maximum deviation is 67%. As a result, there are 35 different possible particle trajectories. The injection velocity of the particles is assumed to be equal to the initial velocity of the carrier gas. The described plasma and particle models are solved by developing a code using control volume algorithm [8]. Solving the plasma temperature and flow fields without injection of any particles starts the computation. Using these conversed temperature and flow fields, particles trajectories together with particle temperature and size histories are computed. The particle source terms for the mass, momentum and energy conservation equations for each control volume throughout the torch are then predicted. The plasma temperature and flow fields are predicted again incorporating these particle source terms. The new plasma temperature and flow fields are used to recalculate the particles trajectories, temperature and size histories. Computing the new source terms and incorporating them into conservation equations constitute the effects of plasma-particle interaction, thereby completing the cycle of mutual interaction. The above computation schemes are repeated until the convergence.

III. Simulated Results

The developed model is used to predict the plasma fields (e.g plasma temperature and flow velocity) and the energy exchange between plasma and particle, particle trajectories, and particle temperature and size histories along the trajectories. The present computation is carried out for soda-lime-silica glass particles. Computation is carried out for a plasma discharge of 10 kW plasma power, 4 MHz induction frequency and 0.1 MPa pressure. Fig.2 shows the effects of carrier gas flow-rate on the spatial plasma temperature. This figure clearly depicts that higher carrier gas flow-rate cools the plasma around the torch centerline and has insignificant effects away from the torch centerline. Also the higher the carrier gas flow-rate the lower the residence time of particles in plasma contact, due to the higher plasma velocity. Fig.3 shows the particle trajectories at 6 L-min^{-1} carrier gas flow-rate and 10 gm-min^{-1} powder feed-rate. It is noticed that particles having larger diameter remain close to the torch centerline and lighter particles move more away from the centerline. Fig.4 shows the particle temperature (for 20, 50 and 90 μm particles) computed from the energy balance equation. It is found that the smaller particle attains the boiling point temperature and larger

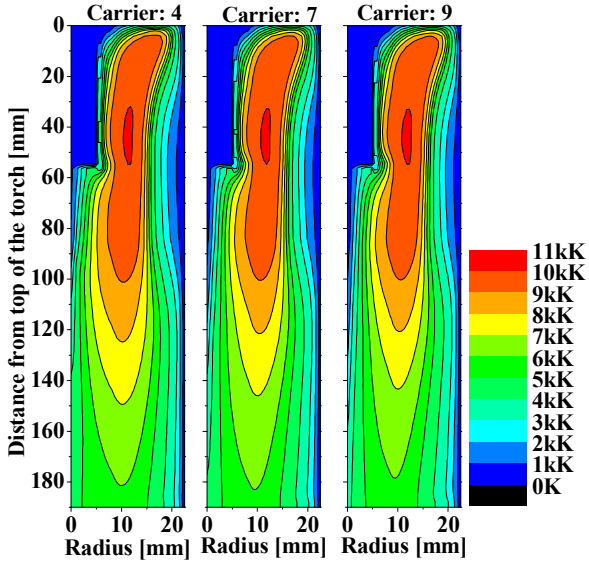


Fig.2 Effects of carrier gas on plasma isotherms

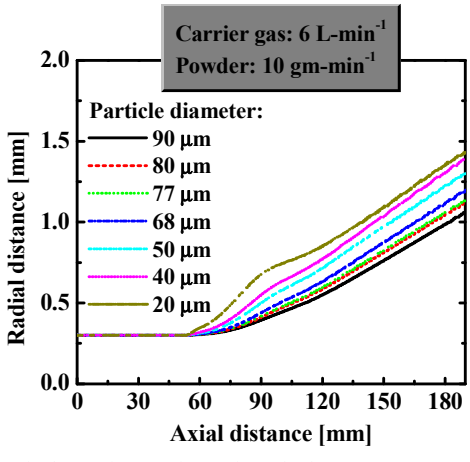


Fig.3 Particle trajectories within the plamsa torch

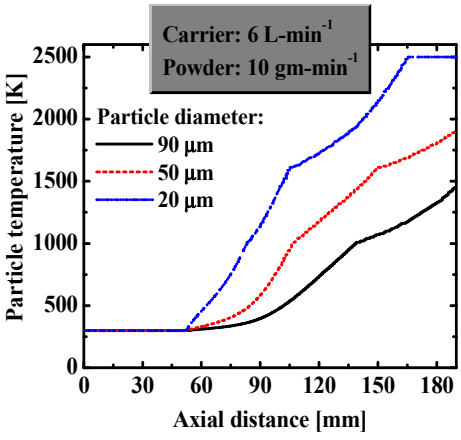


Fig.4 Dependence of particle temperature on diameter

particle attains low temperature. This is because, larger particles fly through the low temperature region (almost along the torch centerline) and smaller particles fly through the high temperature region away from the centerline as described in Fig.3. The porosity of particles is about 80%. The porous particles move through the plasma and gets heated and melted, thus the porosity becomes zero and particles become compact. After melting, particle temperature increases and reaches boiling point temperature and diameter starts to decrease due to vaporization. Thus, it is comprehended that diameter decreases due to the decrease of porosity and vaporization. Fig.5 shows the effects of carrier gas flow-rate on the diameter history within the torch for a particle of 20 μm diameter. For the solid curve diameter remains unchanged up to point A (melting starts), and then from A to B (melting completed) diameter shrinks due to the change of porosity, and from point C diameter changes due to the vaporization. Fig.6 displays the plasma and particle axial velocity. This figure depicts that the lighter particle (20 μm) has similar velocity as that of plasma, and larger particles attain high velocity in the downstream of plasma flow; although the larger particles have smaller velocity in the upstream of plasma flow due to their large inertia. Fig.7 describes the dependence of energy transfer to particles on diameter. It is noticed that energy transfer (per unit mass) to smaller particles is larger than that of larger particles. According to energy balance equations

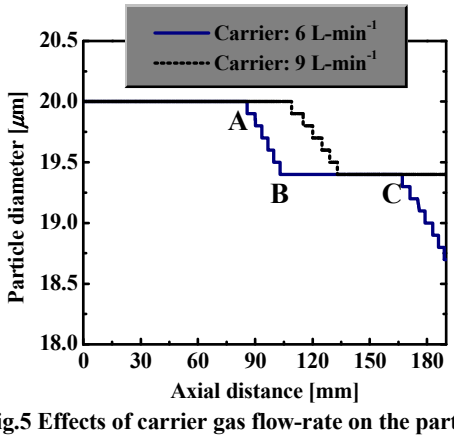


Fig.5 Effects of carrier gas flow-rate on the particle diameter history

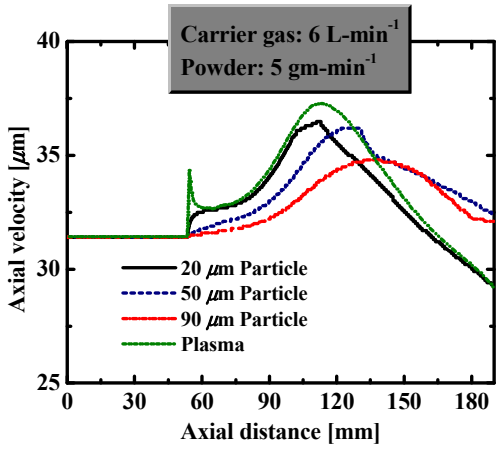


Fig.6 Axial velocity of particle with the plasma

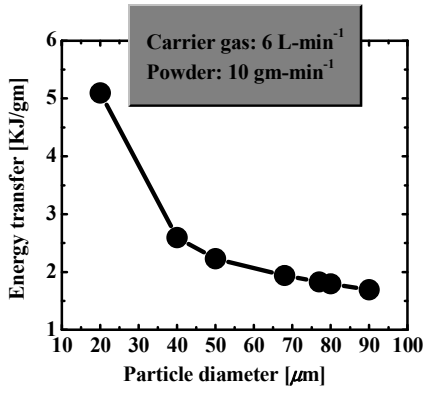


Fig.7 Dependence of energy transfer on particle diameter

(9) and (10), energy transfer to particles is dominated by the surface area of the particles. As the number of smaller particles is much larger than that of larger particles in per unit mass, thus total surface area is larger for smaller particles for per unit mass. These results are also supported by the previous results presented in Fig.4, where smaller particles attain higher temperature, because larger energy transfer to particles means higher particle temperature.

IV. Conclusions

A plasma-particle interactive flow model is described. This model may be used as a tool for the numerical analysis and simulation of any granulated porous micro-particles. The developed model can be used to optimize the carrier gas flow-rate, particle feed-rate, and particle size distribution to achieve the maximum treatment efficiency during thermal treatment of granulated porous particles by single or mixed-gas induction thermal plasmas. Numerically, it is found that the heat transfer to particles decreases at increased carrier gas flow-rate and energy transfer to smaller particles is higher than that of larger particles for per unit mass, and these results well agree with those of experiment [9]. Thus, it may be argued that the efficient thermal treatment of particles depends not only on the physical properties of the

particles, but also on the plasma discharge conditions and particle parameters.

References

[1] T. Watanabe and K. Fujiwara, "Nucleation and growth of oxide nanopartilces prepared by induction thermal plasmas," *Chem. Eng. Comm.,* vol. 191, pp.1343-1361, 2004.

[2] M. I. Boulos, "Heating of Powders in the Fire Ball of an Induction Plasma," *IEEE Tran.Plasma Sci.*, vol. PS-6, pp. 93- , 1978.

[3] J. Mostaghimi, K. C. Paul, and T. Sakuta, "Transient Responses of the Radio Frequency Inductively Coupled Plasma to a Sudden Change in Power", *J. Appl. Phys*., vol. 83, pp. 1898-1908 1998.

[4] M. M. Hossain, Y. Yao, Y. Oyamatsu, T. Watanabe, F. Funabiki and T. Yano, *WSEAS Trans. Heat and Mass Transfer*, vol. 1, 625 (2006).

[5] X. Chen and E. Pfender, "Effect of the Knudsen Number on Heat Transfer to a Particle Immersed into a Thermal Plasma," *Plasma Chem. Plasma Process*. vol. 3, pp. 97- , 1983.

[6] Y. C. Lee, Y. P. Chyou, and E. Pfender, "Particle Dynamics and Particle Heat and Mass Transfer in Thermal Plasmas. Part II. Particle Heat and Mass Transfer in Thermal Plasmas," *Plasma Chem. Plasma Process*., vol. 5, pp. 391-414, 1985.

[7] C. T. Crowe, M. P. Sharma, and D. E. Stock, "The Particle-Source-In Cell (PSI-CELL) Model for Gas-Droplet Flows", *J. Fluids Eng.*, vol. 99, pp.325-332, 1977.

[8] S. V. Patankar, *Numerical Heat Transfer and Fluid Flow*, Hemisphere, New York, 1980.

[9] Y. Yao, M. M. Hossain, Y. Oyamatsu, T. Watanabe, F. Funabiki and T. Yano, "Plasma-particle heat transfer mechanism for in-flight melting of powders in induction thermal plasmas," Proceedings of ASCHT07 1st Asian Symposium on Computational Heat Transfer and Fluid Flow, Xi'an, China, October 18-21, 2007, Paper No. ASCHT2007-069.

.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

1.54 μm Lasing from Silicon in Presence of Erbium Doping

M. Q. Huda, M. Z. Hossain

Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology
Dhaka 1000, Bangladesh
E-mail: mqhuda@eee.buet.ac.bd

Abstract – Lasing at 1.54 μm from Erbium doped silicon has been studied. A model has been developed for the mechanism of energy transfer to erbium by electron-hole recombination through erbium sites. Emission rates of erbium through intra 4f shell transitions by spontaneous and stimulated processes have been equated with the excitation rates. Detailed analysis on rate equations show the feasibility of achieving population inversion and lasing threshold for incorporation of 10^{19} cm^{-3} of optically active erbium sites. Low threshold current densities of the order of A/cm^2 has been estimated for optimized lasing conditions. Linear increase of laser output with excitation current has been simulated. Modulation compatibility of the erbium doped silicon lasing system has been studied by introducing small signal components at various operating conditions. It was found that direct modulation of the 1.54 μm erbium emission with frequencies up to Gega hertz level is feasible. The 3 dB bandwidth of laser response was found to be a strong function of the power output. Rate equations of laser operation were also solved for large signal conditions. Turn-on delays of the order of tens of nanoseconds have been estimated.

I. Introduction

Erbium doped silicon has been extensively studied in recent times for its prospective application in silicon based optoelectronics [1]-[4]. The motivation is to develop a light emitting source in silicon which is compatible with standard processing technology. This would allow the enormous optical bandwidth to be utilized in on-chip distribution of data and clock signals. The bottleneck of present Integrated Circuit technology in realizing Aluminum/Copper data bus networks for high-speed low dimensional chips can thus be eliminated. Silicon is the material for the semiconductor industry at present and in the foreseeable future. It has tremendous advantages regarding the availability, processing techniques, and cost-effectiveness. It also supports optical processes of waveguiding and detection techniques. However, being an in-direct material, silicon is not radiative. As a result, achieving an on-chip light emitter has been the main hurdle towards the realization of silicon based optoelectronics.

Several approaches have been considered in recent years to achieve luminescence in silicon. Among them, the incorporation of Erbium in silicon has attracted lot of interest for its atomically sharp emission at the minimum loss window of fiber optic communication. Different aspects of silicon erbium systems have been studied and light emitting diode operation have been reported several years ago [5]. Si:Er LEDs however, are yet to be applied in practical circuits due to the lack of sufficient emission power. Also, the relatively larger lifetime of erbium luminescence in silicon makes it unsuitable for direct modulation at high frequencies of operation.

Large emphasis has been given in recent times on prospects of light amplification and lasing action on silicon. Most of the work is being carried out on silica based silicon nanocrystals doped with erbium [6]-[8]. Such structures are process compatible with silicon technology. Erbium atoms sensitized by an external source are reported to provide sufficient stimulus for light amplification at 1.54 μm. Laser operation has also been reported. These silica based structures although promising have the inherent problem of requiring an additional exciting source typically with photon energies larger than the silicon bandgap. On-chip photonic application of these structures would most likely involve external modulators. Attachment of on-chip/off-chip modulators along with the lasing device would demand more silicon space inside the chip.

Achieving lasing action from erbium in silicon in the form of a laser diode would be a proper approach for the silicon photonics. However, due to practical limitations of erbium incorporation related issues in silicon and the moderate or low power emission from Si:Er through electroluminescence, such possibilities have not been explored in depth. Xie et al. [9] first discussed the phenomenon of light amplification and laser operation in erbium doped silicon. Not much work, either in theories or experiments have been done on this topic. We have shown in previous publications that sustained stimulated emission in erbium doped silicon is feasible [10]-[11]. In this paper, we show that, laser action in Si:Er is theoretically possible with the present technology of erbium incorporation. We also make analysis on performance and characteristics of prospective erbium doped silicon lasers.

II. Erbium Emission in Silicon

Erbium atoms when incorporated in silicon produce energy states in the bandgap. Electron-hole recombination through these sites produce the energy for pumping erbium atoms from their ground state of $^4I_{15/2}$ to the first excited level of $^4I_{13/2}$. Radiative transition of erbium atoms from their excited state to the ground state produce the luminescence at 1.54 μm. Thus the 1.54 μm emission in erbium incorporated silicon can be described as radiation via a non-radiative route. The process of light amplification and the subsequent lasing action in erbium-doped silicon can be considered to be a quasi-two level laser system, where erbium atoms are pumped electronically and then interact optically in the lasing system. Unlike a two level system where saturation occurs at the onset of population inversion, erbium-doped silicon can be designed to maintain a sustained level of population inversion through proper excitation of erbium atoms. The electron-hole mediated process of erbium excitation can be represented by a Shockley-Rheed-Hall model [3]. In case of a laser diode, erbium atoms would be excited from the ground state $^4I_{15/2}$ to the first excited state of $^4I_{13/2}$ through carrier injection under forward bias. A part of transition of erbium atoms from the excited state to the ground state provide the 1.54 μm emission.

Let N_{Er} be the density of erbium atoms in the active volume, and suppose at a certain condition of electrical excitation, the density of atoms in the $^4I_{13/2}$ state is given as N_{Er}*. Under this condition, the fraction of erbium sites in the bandgap that are occupied by electrons is given as:

$$f_t = \frac{n_{Er}}{N_{Er}} = \frac{e_p + c_n n}{e_n + e_p + c_n n + c_p p}. \quad (1)$$

where, e_n and e_p are the electron and hole emission coefficients from the erbium trap; and c_n, c_p are the capture coefficients. n and p are the carrier densities.

Rate of electrical pumping of erbium atoms is given as:

$$R_{ex} = f_t c_p p\left(N_{Er} - N_{Er}^*\right) \quad (2)$$

Rate equations representing the density of excited atoms, and the optical power density inside the cavity (S) are then given as[12]:

$$\frac{d}{dt}N_{Er}^* = R_{ex} - \frac{N_{Er}^*}{\tau_{Er}} - vkS, \quad (3)$$

$$\frac{dS}{dt} = vkS + \beta\frac{N_{Er}^*}{\tau_{rad}} - \frac{S}{\tau_p}, \quad (4)$$

where, f_t is the fraction of erbium sites occupied by electrons, c_p is the capture coefficients, p is the carrier density, τ_{Er} represents the lifetime of erbium decay, τ_{rad} is the radiative lifetime, τ_p is the lifetime of photon decay inside the cavity, v is the velocity of light, and k is the gain coefficient of light. Three terms on the right hand side of (3) represent the rate of excitation of erbium through electron-hole recombination, overall decay rate of erbium atoms, and the net generation rate of photons through stimulated emission, respectively. The last two terms on (4) represent contribution of light in the laser cavity through spontaneous emission and the rate of photon loss in the cavity. The fraction β has been estimated to be of the order of 10^{-3} [6].

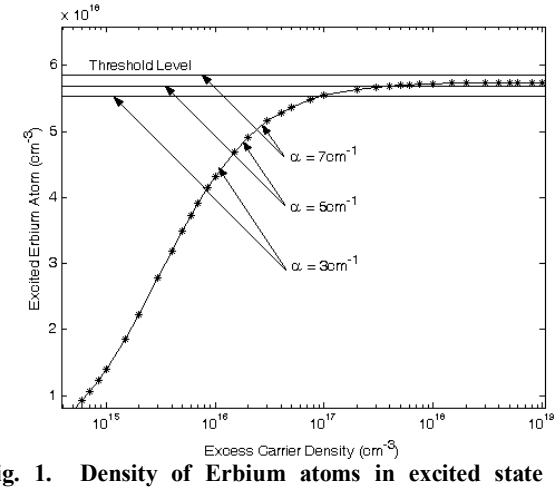


Fig. 1. Density of Erbium atoms in excited state as a function of excess carriers. Erbium density is taken as 10^{19}/cm^3.Threshold level of population inversion is seen to be a function of optical losses in the cavity.

Threshold value of gain coefficient for the laser is given as:

$$k_{th} = \alpha + \alpha_m = \alpha + \frac{1}{2L}\ln\frac{1}{(R_1 R_2)}. \quad (5)$$

where, α represents the internal losses due to scattering, absorption, etc., and α_m is the losses of light through the end mirrors. R_1 and R_2 represent the reflectivities of the two end mirrors.

At threshold, the gain reaches its thresold value of k_{th} and the density of erbium atoms in the excited state gets pinned down at N_{ErTh}*. Density of erbium atoms as a function of excess carrier density in the active medium is shown in Figure 1. Cavity length of 300 μm with mirror reflectivities of 90% have been assumed. It is obvious that the internal loss coefficient is a key parameter in

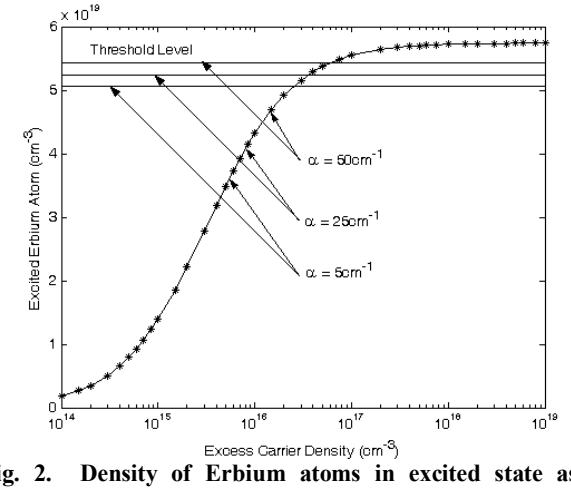


Fig. 2. Density of Erbium atoms in excited state as a function of excess carriers. Erbium density is taken as 10^{20}/cm^3.Threshold level can be achieved for cavity losses well above 50 cm^{-1}.

determining the degree of population inversion necessary for the lasing. Absorption losses of 5 cm^{-1} or smaller is found to be required for the erbium concentration of $10^{19}/cm^3$. To accommodate larger values of loss parameters in the laser cavity higher concentrations of erbium density would be necessary. This is shown in Figure 2 where lasing threshold is seen to be achievable at loss coefficients above 50 cm^{-1} for erbium concentration of $10^{20}/cm^3$. Erbium concentrations of these order has been demonstrated in silica hosts, but remains a challenge to be incorporated in good quality silicon.

III. Lasing Threshold and Output Power

Current density corresponding to the lasing threshold is given as:

$$J_{th} = \frac{n_{th}}{\tau} L_{eq} q, \tag{6}$$

where, n_{th} is the threshold level of excess carrier density, τ is the carrier lifetime in silicon, L_{eq} is the effective depth of erbium incorporated active layer, q is the electronic charge. For current densities above the threshold, photon density builds up inside the cavity due to stimulated emission, and the lasing output is given as[13]:

$$P_{opt} = \frac{hc}{\lambda} SALv\alpha_m. \tag{7}$$

Here, λ is the wavelength, L is the cavity length, α_m represents the coefficient of light output through end mirrors. The photon density S in lasing condition is given by (3) and (4).

Calculated values of laser output as a function of the drive curent is shown in Figure 3. Doping density of $10^{19}/cm^3$ with emission linewidth of 1 Å is assumed[11],[14]. As seen, laser output in mW range is estimated for drive current densities of the order of A/cm^2. It is obvious that larger carrier lifetime of silicon results in smaller lasing

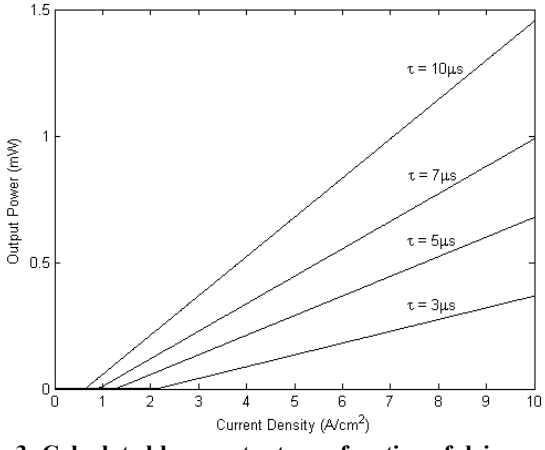


Fig 3 Calculated laser output as a function of drive current density for different silicon lifetime parameters. Lasing threshold of the order of A/cm^2 is estimated for erbium concentration of $10^{19}/cm^3$.

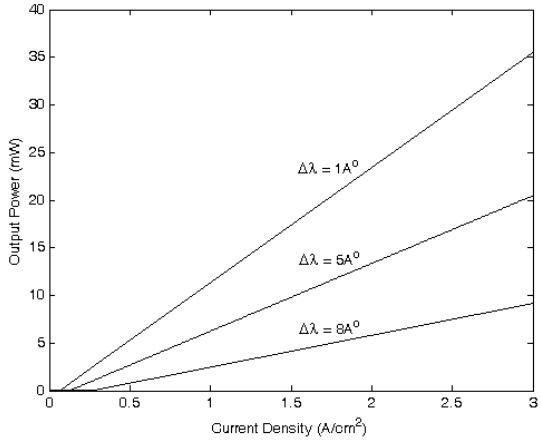


Fig 4 Calculated laser output as a function of drive current density for different emission linewidths for erbium concentration of $10^{20}/cm^3$.

threshold with larger optical output for a specific excitation condition. However, it needs to be mentioned that upper limit of the carrier lifetime in silicon is controlled by the recombination rate of carriers through erbium sites, i.e. the condition when the silicon is of extreme high quality and all recombination routes can be considered negligible in comparison to that through erbium sites. Figure 4 shows the calculated laser output against drive current for different emission linewidth. It is seen that smaller emission linewidth corresponds a smaller lasing threshold with larger optical output. Erbium emission although atomic in nature is typically found to be in the range of 100 Å in photoluminescence measurements [14]. This is due to the coexistence of a number of erbium atomic site symmetry. However, with careful processing, emission linewidth of the order of 1 Å has been reported [15].

IV. Frequency Response

Let us assume that the Si:Er laser is biased above threshold by DC current $J_0 > J_{th}$ and a time-varying current $\Delta J(t)$ is added as:

$$J(t) = J_0 + \Delta J(t) \tag{8}$$

Under steady state conditions, the excited erbium atom and photon density would respond similarly with the drive current and given by:

$$N^*_{Er}(t) = N^*_{Erth} + \Delta N^*_{Er}(t), \quad S(t) = S_0 + \Delta S(t) \tag{9}$$

Putting the time variations in equations (3) and (4), and assuming sinusoidal signals, the small signal modulation response function can be expressed as [12]:

$$M(\omega) = \frac{\Delta S(\omega)}{\Delta J(\omega)} = \frac{A_{14}}{\left(A_{13} - \omega^2\right) + j\omega A_{12}} \tag{10}$$

where, A_{12}, A_{13}, and A_{14} are constants consisting of system parameters. Response of the prospective Si:Er laser normalized to that at zero frequency is given as:

$$H(\omega) = \frac{M(\omega)}{M(0)} = \frac{A_{13}}{\left(A_{13} - \omega^2\right) + j\omega A_{12}}$$

$$= \frac{\omega_r^2}{\left(\omega_r^2 - \omega^2\right) + j\omega\gamma} \tag{11}$$

where

$\omega_r = \sqrt{A_{13}}$ and $\gamma = A_{12}$

ω_r is the relaxation oscillation frequency and γ is the damping constant of the relaxation oscillation respectively. These two terms play an important role in governing the dynamic characteristics of Si:Er laser. The resonance peak f_p is obtained by setting the first derivative of $|H(\omega)|$ to zero. The analytic expression of f_p is

$$f_p = \frac{1}{2\pi}\left(\omega_r^2 - \frac{\gamma^2}{2}\right)^{\frac{1}{2}} \tag{12}$$

The 3-dB modulation bandwidth f_{3dB} is defined as the frequency at which $|H(\omega)|$ is reduced by 3 dB from its DC value. Eq. (10) provides the following analytic expression for f_{3dB}:

$$f_{3dB} = \frac{1}{2\pi}\left(\frac{\left(2\omega_r^2 - \gamma^2\right) + \sqrt{\left(\gamma^2 - 2\omega_r^2\right)^2 + 4\omega_r^4}}{2}\right)^{\frac{1}{2}}$$

or,

$$f_{3dB} = \frac{1}{2\pi}\left(\omega_p^2 + \left(\omega_r^4 + \omega_r^4\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \tag{13}$$

The amplitude of the resonance peak H_p which occurs at frequency $\omega_p = (2\pi f_p)$ is obtained by inserting Eq. (11) into Eq. (10):

$$H_p = \left[\frac{1}{\left(\frac{\gamma}{\omega_r}\right)^2 - \frac{1}{4}\left(\frac{\gamma}{\omega_r}\right)^4}\right]^{\frac{1}{2}} \tag{14}$$

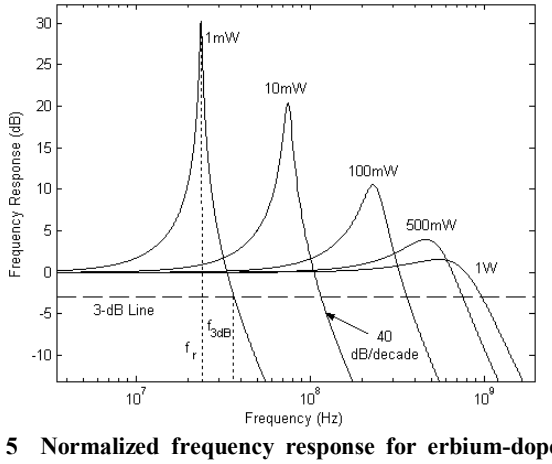


Fig. 5 Normalized frequency response for erbium-doped silicon laser. Direct modulation up to GHz. level is seen feasible for 1 watt power.

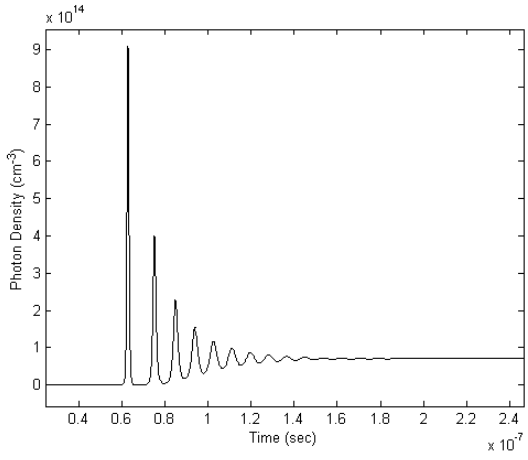


Fig. 6 Profile of the density of photons in the laser cavity against time. Turn on delay around 60 ns with initial spikes are simulated before the continuous mode of lasing.

Normalized frequency response of a laser structure is shown in Fig. 5. It is seen that, direct modulation in Gega hertz. level is feasible for laser outputs in the range of watt.

V. Time Response

At the time of laser switching with a drive current above the threshold, a finite amount of time is spent to pump erbium atoms to the threshold level of population inversion. The onset of lasing is marked by an enhanced decay of erbium atoms due to stimulated emission. This creates momentary depopulation of erbium atoms below the threshold level followed by subsequent pumping. The result is a typical turn-on delay accompanied by laser oscillations. The effect can be calculated by numerical solution of (3) and (4). Photon density inside the laser cavity, representing the time function of laser output is shown in Fig. 6. A drive current of 276 A/cm^2 with erbium concentration of 10^{20}/cm^3 was used. It was found that larger drive currents result in smaller turn-on delays with less effect of initial oscillations.

VI. Conclusions

We have studied the prospects of erbium doped silicon laser for its application in on-chip optical communications. A model based on Shockley-Rheed-Hall recombination kinetics and quasi two level lasing analysis was used. Detailed analysis show that erbium doped silicon lasing system is feasible for certain conditions of erbium incorporation. It has been shown that direct modulation of Si:Er laser with frequencies of the order of Gegahertz is possible. Turn-on delays of the order of tens of ns is estimated.

References

[1] T. Gregorkiewicz, et al. "Energy transfer between shallow centers and rare-earth ion cores: Er^{3+} ion in silicon," Phys. Rev. B, vol. 61, pp. 5369-5375, 2000.

[2] M. J. A. de Dood, J. Knoester, A. Tip, and A. Polman, "Förster transfer and the local optical density of states in erbium-doped silica," Phys. Rev. B, vol. 71, p. 115102, 2005.
[3] M. Q. Huda, , S. A. Siddiqui, , M. S. Islam, "Explaining the luminescence profile of erbium in silicon under short excitation pulses," Solid State Communication., vol. 118, pp. 235-239, 2001.
[4] H. Isshiki, M. J. A. de Dood, A. Polman, and T. Kimura, "Self-assembled infrared- luminescent Er-Si-O single-crystals on silicon," Appl. Phys. Lett., vol. 83, p. 4343, 2004.
[5] B. Zheng, J. Michel, F. Y. G. Ren, L. C. Kimerling, D. C. Jacobson, and J. M. Poate, "Room-temperature sharp line electroluminescence at μ=1.54 μm from an erbium-doped, silicon light-emitting diode," Appl. Phys. Lett. vol. 64, p. 2842, 1994.
[6] T. J. Kippenberg, J. Kalkman and A. Polman, K. J. Vahala, "Demonstration of an erbium-doped microdisk laser on a silicon chip," Physical Review A vol. 74, p. 051802, 2006.
[7] A. J. Kenyon, M. Wojdak, and I. Ahmad, W. H. Loh and C. J. Oton, "Generalized rate-equation analysis of excitation exchange between silicon nanoclusters and erbium ions," Physical Review B vol. 77, p. 035318, 2008.
[8] Yong-Seok Choi,a) Joo Yeon Sung, Se-Heon Kim, Jung H. Shin, and Yong-Hee Lee, "Active silicon-based two-dimensional slab photonic crystal structures based on erbium-doped hydrogenated amorphous silicon alloyed with carbon," Applied Physics Letters Vol. 83, p. 3239-3241, 2003.
[9] Y. H. Xie, E. A. Fitzgerald, Y. J. Mii, "Evaluation of erbium-doped silicon for optoelectronic applications," J. Appl. Phys., vol. 70, 3223, 1991.
[10] M. Q. Huda, , S. I. Ali, "A study on stimulated emission from erbium in silicon," Materials Science and Engineering B, vol. 105, pp. 146-149, 2003.
[11] M. Q. Huda, , S. I. Ali, "Prospects of Laser Operation in Erbium Doped Silicon," Mat. Res. Soc. Symp. Proc., vol. 770, pp. I3.5.1-I3.5.6, 2003.
[12] M. Z. Hossain, "Development of an Analytical Model of Erbium Doped Silicon Laser Diode," M.Sc. Thesis, Bangladesh University of Engineering & Technology, 2006.
[13] Chuang, Shun Lien, Physics of Optoelectronic Devices, Second edition, John Willy & Sons, New York, 1995.
[14] M. Q. Huda, J. H. Evans-Freeman, A. R. Peaker, D. C. Houghton, A. Nejim, "Luminescence from erbium implanted silicon–germanium quantum wells," J. Vac. Sci. Technol. B vol. 16, p. 2928, 1998.
[15] N. Q. Vinh, , H. Przybylinska, Z. F. Krasil'nik, and T. Gregorkiewicz, "Microscopic Structure of Er-Related Optically Active Centers in Crystalline Silicon," Phys. Rev. Lett., vol. 90, p. 066401, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Dielectric study of hafnium oxide thin film annealed in oxygen and deposited using RF sputtering system having MIM configuration

A. Srivastava[1], R.K. Nahar[2], C.K Sarkar[3], Vinay Gupta[4]

[1]Electrical and Electronics Engineering Department, BITS Pilani, Rajasthan333031, India
[2]Sensors and Nanotechnology Group, Central Electronics Engineering Research Institute, Pilani, Rajasthan 333031, India
[4]Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India
[1,2 & 3]Department of Electronics & Telecommunication Engineering, Jadavpur University, Kolkata, India

Corresponding email: ashudel07@gmail.com

***Index Terms*—MIM capacitor; Dielectric loss; Conductivity; high-κ dielectric.**

Abstract-The dielectric study of HfO2 thin films deposited on the platinized silicon substrate using RF-sputtering deposition technique have been carried out in the metal-insulator-metal (MIM) configuration over a wide temperature (300 to 500 K) and frequency (100 Hz to 1 MHZ) ranges. The film were deposited at pre-optimized sputtering voltage of 0.8 kV, substrate bias of 80 volt and annealing temperature of 700^{0}C in oxygen in order to get best results for the oxide charges and the leakage current as a MOS capacitor. The crystallographic structure of the deposited films were investigated using X-ray diffraction and the microstructure of the thin film is examined by SEM.

I Introduction

As MOSFET transistors are scaled-down towards realizing the sub 100 nm node technology, the corresponding SiO_2 gate oxide thickness is going below 2 nm leading to high leakage current and reliability problems due to the direct tunneling through the thin oxide. There are many High-*k* dielectric materials that are investigated to replace SiO_2 namely SiN, TiO_2, Ta_2O_5, Al_2O_3, HfO_2, ZrO_2. Unfortunately, many of these materials are thermodynamically unstable on silicon. HfO_2 has emerged as one the most promising High-*k* gate dielectric material, as it has relatively high dielectric permittivity value (25), the high heat of formation (271 K cal/mol), the high band gap (5.8 eV), compatibility with poly-silicon gate process and the large conduction band offsets with Si, and solid-state thermodynamic stability on Si and SiO_2 [1-5].High-*k* dielectric material are also drawing increasing attention as MIM capacitor for analog/mixed signal and rf integrated circuit application due to their higher capacitance density with the downward scaling of the capacitor area [6–9].

RF sputtering is generally preferred due to the possibility of low temperature processing and the process simplicity. However, the process optimization with energetic ion bombardment as a function of process parameters and their effect on charges is needed. One of the features of deposition of thin film by sputtering is the possibility of tailoring film microstructure employing substrate bias during film deposition. Effect of the substrate bias has been used in the past for thin films for improving film microstructure and polycrystalline film texture. However, little attention has been made towards the detail dielectric studies in HfO_2 thin film especially in the metal-insulator-metal (MIM) configuration grown by RF sputtering under the optimized process parameters with the energetic ion bombardment.

This paper investigates dielectric properties of HfO_2 sputtered thin film deposited on the platinised silicon substrate at pre-optimized sputtering voltage, bias sputtering and annealing temperature in oxygen which were 0.8 kV, 80V and 700^{0}C respectively [5]. The HfO2 film was deposited on platinized silicon substrate as a metal-insulator-metal configuration by RF sputtering at pre-optimized sputtering voltage, bias voltage and annealed in oxygen and the dielectric studies were subsequently carried as a function of frequency and temperature.

II. Experimental details

Five inch sputtering target of HfO_2 with high purity (99.9% purity) supplied by M/s Semiconductor technology was used to deposit the thin films using MRC rf sputtering machine. P-type silicon substrate were used to deposit HfO_2 which had 1–10 ohm-cm resistivity and (100) orientation. The wafers were cleaned using the standard pirana cleaning procedure in order to remove organic and inorganic contaminations. The wafers were etched in dilute HF (1: 20), rinsed in DI water and dried in dry N2 immediately before loading in the vacuum chamber. The background pressure of the vacuum chamber was evacuated to the $1.2x10^{-6}$ torr and the sputtering was done in high purity argon ambient gas. The gas pressure was maintained at about 6 m torr and the film deposited at 0.8 kV sputtering voltage and 80 volt substrate bias for getting optimum results for the oxide charges and the leakage current. The sputtered film thickness of HfO_2 was around 200 nm. The

As- deposited films were further treated at 700^{0}C in Oxygen for 15 minutes each. The top electrodes were Al dots deposited by the thermal evaporator using shadow mask technique having diameter of around 200 □m and film thickness of about 200 nm while back contact was again deposited by thermal evaporator to get planar Al thin film having thickness of about 200 nm. The film thickness was measured using Laser Ellipsometer (Sentech Instruments Laser-Pro) and the microstructure was examined by VEECO-CPII Atomic Force Microscope. The dielectric properties of CCTO films were studied in the MIM configuration Pt/HfO_2/Pt/ Si using an impedance analyzer HP 4294A in the frequency range of 100 Hz–1 MHz.

III. Results and discussions

A. Structural studies

Surface morphology of the thin film effects the electrical properties of the MIM capacitor structure and the improvement in the film properties is attributed to improvement in microstructure of the thin film. Surface topography and 3 D AFM pictures of sputtered 200 nm HfO_2 thin film deposited at sputtering voltage of 0.8 kV, substrate bias of 80V and annealed in oxygen at 700^{0}C is shown in Fig.1. It is observed that the size of nano-structures have been found to be 8.9 nm.

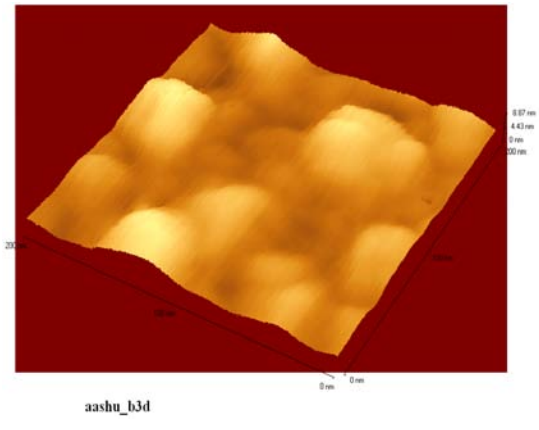


Fig. 1 Surface topography and 3 D AFM pictures of sputtered HfO_2 film which were deposited at a sputtering voltage of 0.8 kV, substrate bias of 80V and thermally annealed in oxygen for 15 minutes at 700^{O}C.

The HfO_2 thin films deposited on the platinized silicon substrate by RF sputtering PLD were adherent strongly to the substrate. The film was deposited at pre-optimized sputtering voltage of 0.8 kV, substrate bias of 80V and thermally annealed in oxygen for 15 minutes at 700^{O}C. X-ray diffraction (XRD) pattern of HfO_2 film of thickness of ~0.2 mm deposited on platinized-silicon (Pt/Si) at 700 °C is shown in Fig. 2. A dominant reflection corresponding to (311) directions was observed along with the weak XRD peaks indicating that the preferential growth of HfO_2 film is along (311) on the Pt/Si substrate.

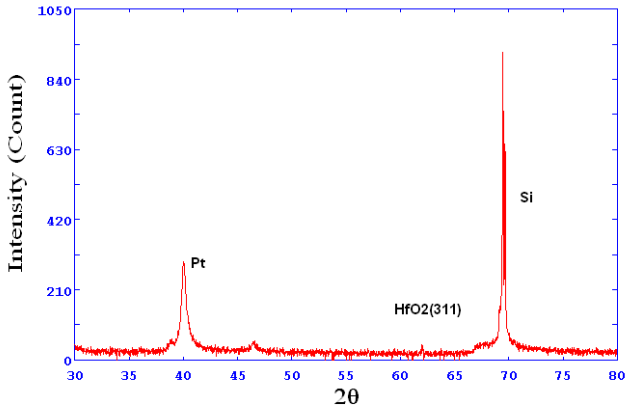


Fig. 2 XRD pictures of sputtered HfO_2 film which were deposited at a sputtering voltage of 0.8 kV, substrate bias of 80V and thermally annealed in oxygen for 15 minutes at 700^{O}C.

B. Electrical studies

Figure 3 shows a plot of ac conductivity as a function of reciprocal temperature (1000/*T*) at five fixed frequencies of 100 Hz, 1 KHz, 10 KHz, 100 KHz, and 1 MHz for HfO2 thin film. It may be noted that the ac conductivity σ(ω) has weak temperature dependence, but increases with increasing frequency (Fig.3).

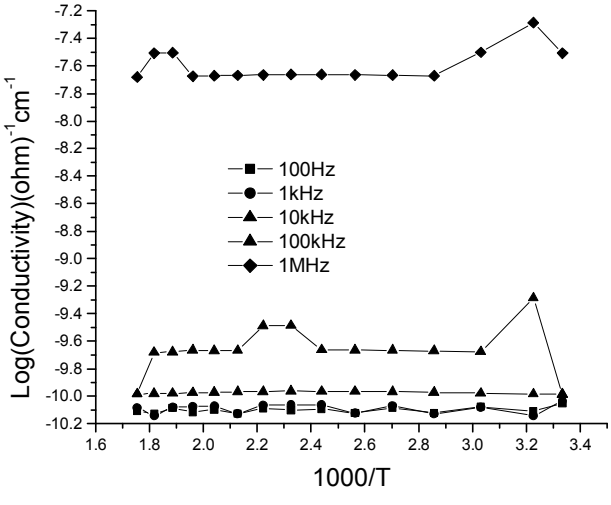


Fig. 3 Variation of ac conductivity of sputtered HfO_2 film as a function of reciprocal temperatures at different frequencies.

The variations of σ(ω) as a function of frequency measured at different temperatures are shown in Fig. 4. The estimated value of σ(ω) was found to increase slightly at higher frequency for all temperature regions.

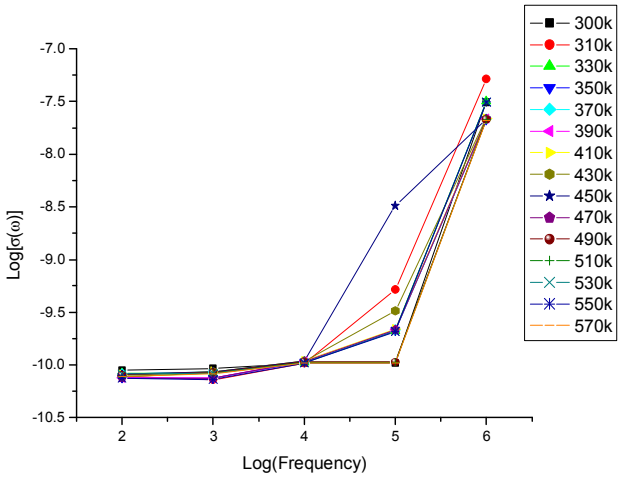


Fig. 4 Log-log plot of σ(ω) of sputtered HfO_2 film as a function of frequency at different temperatures.

The variation of measured dielectric constant ε`(ω) as a function of temperature at the fixed frequencies (1 KHz, 10 KHz, 100 KHz, and 1 MHz) is shown in Fig. 5. We can observe that dielectric permittivity is almost constant for all temperatures and frequencies which is around 19.

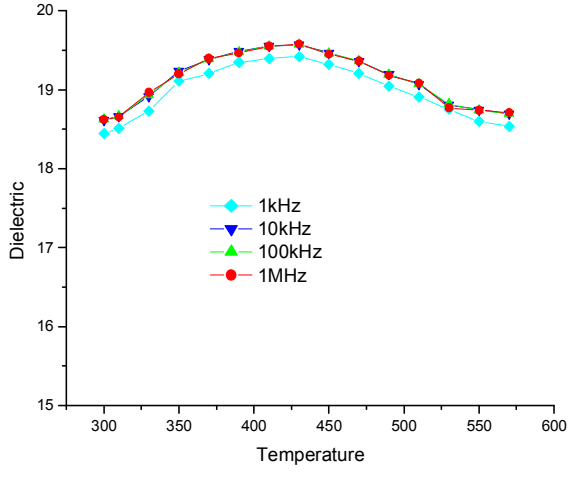


Fig. 5 Dielectric permittivity of sputtered HfO_2 film as a function of temperature at different frequencies.

The variations of dielectric loss tan(δ) as a function of frequency measured at different temperatures are shown in Fig. 6. The value of tan(δ) was found to increase slightly at lower frequency for all temperature regions.

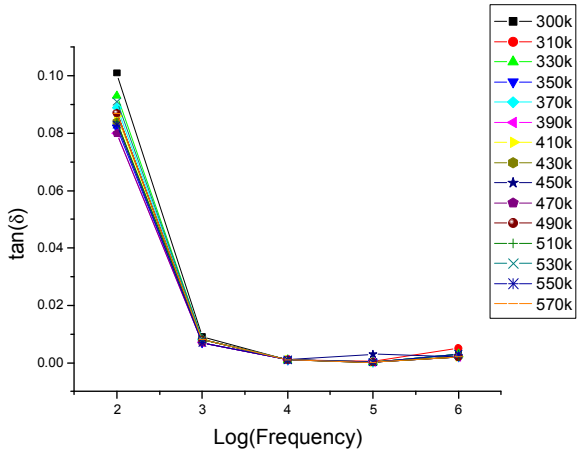


Fig. 6 Dielectric loss (tan δ) of sputtered HfO_2 film as a function of temperature at different frequencies.

IV. Conclusions

Dielectric study of HfO_2 thin films deposited on platinized silicon substrate using RF-sputtering deposition technique have been carried out in the metal-insulator-metal configuration over a wide temperature (300 to 500 K) and frequency (100 Hz to 1 MHZ) range. The film were deposited at pre-optimized sputtering voltage of 0.8 kV, substrate bias of 80 volt and annealing temperature of 700^0C in oxygen in order to get the best results for oxide charges and leakage current as a MOS capacitor. It is observed from 3-D AFM pictures of sputtered 200 nm thin film HfO_2 that the size of nano-structures was around 8.9 nm. XRD peaks indicated that the preferential growth of HfO2 film is along (311) on Pt/Si substrate. The ac conductivity showed weak temperature dependence, but slight increase with increasing frequency. The dielectric permittivity for all temperatures and frequencies was around 19, while the dielectric loss tan(δ) slightly increased at lower frequency for all temperature regions

Acknowledgement

The authors would like to thank CEERI Pilani and Delhi University, for the experimental support to carry out this experimental project. One of the author wishes To thank AICTE, New Delhi for financial support under RPS scheme.

References

[1] H. Wong, H. Iwai, Microelectronic Engineering, 83, (2006), pp.1867–1904
[2] H. Wong, H. Iwai, Physics World, 18 (9) (2005) pp.40.
[[3] D. Misra, H. Iwai, H. Wong, Electrochem. Soc. Interface 14 (2) (2005) pp.30.
[4] B. Sen, C.K. Sarkar, H. Wong, M. Chan, C.W. Kok, Solid State Elect. 50, (2006), pp.237.
[5] R. K. Nahar, Vikram Singh, Aparna Sharma, J Mater Sci:

Mater Electron, 18, (2007) pp.615.
[6] M. Housa, High k gate dielectric, IPO, Bristal (2004) Chapter 1.
[7] G.D. Wilk, R. Wallance, G. Anthony, J. Appl. Phys. 87, 5243 (2001)
[8] Y.H. Kim, J.C. Lee, Microelectronics & Relaib. 44, 183 (2004)
[9] H. Kim, P.C. McLntyre, K.C. Saraswat, Appl. Phys. Letts. 82, 106 (2003)

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Comparison of Photoresponse Characteristics between Nitrogen and Phosphorous Doped n-C/p-Si Heterostructure

Ahmed Tasnim Rasin[1,2] and Sharif Mohammad Mominuzzaman[1]

[1]Department of Electrical an Electronic Engineering, Bangladesh University of Engineering and Technology, BUET, Dhaka 1000, Bangladesh
[2]Department of Electrical an Electronic Engineering, Stamford University Bangladesh, Dhaka-1217, Bangladesh
E-mail: rasin255@yahoo.com

Abstract – Photoresponse of nitrogen and phosphorous doped n-C/p-Si heterostructure have been studied. Camphor ($C_{10}H_{16}O$) was used as starting precursor material in both cases. Phosphorous was doped in varying amounts (1%-7% by mass) and Nitrogen was doped in gas phase with varying partial pressure in the range from 0.3 to 50 mTorr. The doped carbon films were deposited on Silicon substrates by pulsed laser deposition (PLD) technique. Photoresponse of the films varied with increasing nitrogen and phosphorous content. Maximum photoresponse for phosphorous doped carbon films was observed for 5% doping and that for nitrogen doping was observed at 1 mTorr nitrogen partial pressure (NPP). Photoresponse was higher in case of nitrogen doping. Since nitrogen was doped in gas phase it has an advantage over phosphorous of having the better control as dopant. Photoresponse deteriorated in case of both phosphorous and nitrogen doping after the maximum overall photoresponse was observed. The total photoresponse variation was similar to carbon contributions for both cases. The contribution of silicon remained almost constant for P-doping. In case of NPP doping there is possible modification of the structure beyond 10 mTorr.

I. Introduction

Photovoltaic cells are being fabricated at recent times using carbon [1]. The properties of carbon can be tuned over a wide range by effective doping. Undoped amorphous carbon is weakly p-type [2]. It has a mixture of sp^3 and sp^2 bonded carbon of different fractions but has higher sp^3 bonds. But complexity of structure and presence of high density defects in a-C restrict its ability to be doped efficiently. Selection of method of deposition and the optimum precursor material are two most important factors that optical and electrical properties of a film. Graphite was commonly used as the target material for physical vapour deposition of diamond like films. However, recently camphor ($C_{10}H_{16}O$), a natural source is already used to produce fullerenes, micro/nano tubes, and various forms of carbons. Camphor is abundantly available in nature. It is a promising precursor mainly due to its structural advantages. Recent study on camphoric carbon (CC) shows that soot obtained from burning camphor is a better precursor material than graphite because of the presence of high amount of tetrahedral (sp^3) bonding configurations while graphite has only trihedral (sp^2) bonds [3]. Furthermore, the presence of abundant hydrogen in its structure is expected to passivate defects, therefore, doping efficiency would improve. We are working on semiconducting camphoric carbon for its optoelectronic applications. Doped carbon/ silicon heterojunction photovoltaic (PV) is reported, where carbon thin film is deposited from camphor precursor and nitrogen (N) [4] phosphorus (P) [5] and are used as dopants.

In the present work, the photoresponse characteristics of the P doped C/p-Si and N doped C/p-Si heterojunction PV solar cell is studied in detail and is compared.

II. Experimental Details

The energy of carbon species generated by various preparation methods is different and plays an important role to control the sp^3 and sp^2 bonding ratio. The nitrogen doped carbon films were deposited [4] on single crystal Si (100) and fused quartz substrates by xenon chloride (XeCl) excimer pulsed laser deposition (NISSIN 10x, wavelength λ= 308 nm) which is focused on the target by an ultraviolet grade plano-convex lens at an incidence angle of 45° to the target normal, using CC target. In this research CC has been used as a new starting precursor. CC soots are obtained from burning camphor in a quartz cylinder tube. A rotary pump is used to suck the air out. A filter is used to filter the soot and clean the mixed air. The CC soot deposited along the inside walls of the cylinder tube and the filter was collected, dried in the oven and later pressed into pellets to be used as targets. In order to dope, CC target was mixed with varying volumes of N gas (0.3, 0.4, 1, 10, 30 and 50 mTorr). Torr is a non SI unit of pressure and one Torr is equal to 133.32 Pascal. The films were deposited at room temperature of 25 ^{0}C.

In case of phosphorous doping [5] the dried CC soot before being compressed is mixed with varying amounts of red phosphorous powder (1%, 3%, 5% and 7% by mass). Then using this target containing phosphorous, the films are deposited on the quartz substrates by PLD system.

III. Nitrogen vs. Phosphorous as Dopant

Complex structure and high density of defects restrict the ability of amorphous carbon to be doped efficiently. When attempting to dope a-C, control of dopant is very important. Effective doping can modify electronic properties, especially gap states, conductivity, etc in semiconductor materials. Since nitrogen has a smaller radius compared to phosphorus which is closer to carbon, the former is to be preferred. Nitrogen is the most widely studied candidate for n-type dopant in a-C [6]. Phosphorous is widely used as an n-type impurity in silicon [7] and is a possible alternative to nitrogen in carbon [8]. Veerasamy reported n-type doping of highly tetrahedral carbon (ta-C) using solid P and n-type doping of a-C using N gas [9]. Nitrogen being in gas phase has the advantage of better control of dopant concentration over phosphorous in physical deposition systems [4].

IV. Extraction Technique

The quantum efficiency vs. wavelength spectra was plotted in Ultraviolet-Visible-Infrared (300-1200nm) region. When a p-n junction is illuminated by an external source, electron-hole pairs (EHPs) are generated in the depletion region. The high internal electric field causes the EHPs to separate resulting in photocurrent. The ratio of number of generated carriers to the number of incident photons is quantum efficiency. Quantum efficiency is a measure of absorption of photons. If for any solar cell the quantum efficiency is plotted as a function of wavelength, the area under the curve can be related to the overall conversion efficiency.

Curve fitting with Gaussian and Lorentzian distribution were performed on each spectrum. But the fitted curves resulted poor accuracy for Lorentzian functions. Each spectrum was fitted with two, three and four curves. The resultant curve did not match the quantum efficiency vs. wavelength spectra and almost similar results were found in case of three and four Gaussian curves. Each spectrum was finally fitted with three Gaussian curves. The final form of fitted curves and resultant curves of the spectra for nitrogen doping for 50 mTorr is shown in figure 1. Diamond like carbon (DLC) consists of more sp^3 bonds with a mixture of sp^2 bonds. The more the sp^3 bonds the higher the bandgap and hence the occurrence of peak absorption will be at lower wavelengths.

Since energy, $E=h*c/\lambda$ (where h is Planck's constant, c is velocity of light and λ is the wavelength of light) the first Gaussian curve with the lowest peak position (wavelength) corresponds to the contribution of sp^3 bonds in carbon because sp^3 bonds have higher energy band separation.

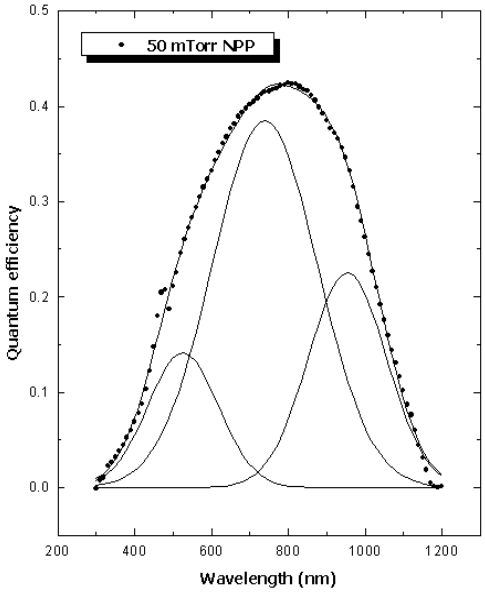


Fig. 1: Quantum efficiency vs. wavelength spectra fitted with Gaussian distribution functions.

The second Gaussian curve with its peak absorption at higher wavelength than the first one is due to sp^2 bonding in carbon. The third Gaussian curve is due to silicon contribution. The photoresponse depends strongly with increasing phosphorous [5] and nitrogen [10] content which has been reported previously.

V. Results and Analysis

The peak positions for both P and N doping are similar. The first Gaussian curve has the peak positions about 550 nm which corresponds to sp^3 bonding variations. The second Gaussian curve has peak around 775 nm and that for the third Gaussian curve is at about 970 nm. The second and the third Gaussian curve correspond to changes in sp^2 and silicon region.

The parameters obtained from the quantum efficiency spectra are given in Table 1 and 2 for the phosphorous and nitrogen doped samples respectively. Summation of area 1 and area 2 curves together give the total contribution of carbon. Silicon contribution is estimated from the area under the curve 3.

Table 1: Total area and C and Si contribution for P-doping

Percentage of phosphorous	Total area	Carbon contribution	Silicon contribution
1%	152.456	114.316	38.14
3%	156.687	120.992	35.695
5%	272.505	211.512	60.993
7%	203.447	158.55	44.897

Table 2: Total area and C and Si contribution for NPP-doping

NPP doping (mTorr)	Total area	Carbon contribution	Silicon contribution
0.3	228.2	166.263	61.912
0.4	179.9	138.824	41.032
1	304.4	255.908	48.581
10	230.6	164.682	65.917
30	216.7	166.054	50.659
50	222.3	164.932	57.388

The summation of area under the three Gaussian curves gives the total photoresponse of the films. From a plot of total area vs. percentage phosphorous doping (Fig. 2) we observe the maximum photoresponse of the films are found for 5% P-doping.

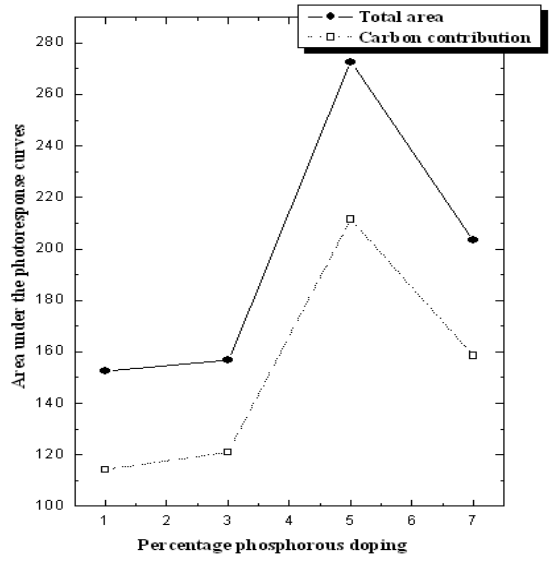


Fig. 2: Total photoresponse variation of phosphorous doped carbon films (with 1%, 3%, 5% and 7% P-doping).

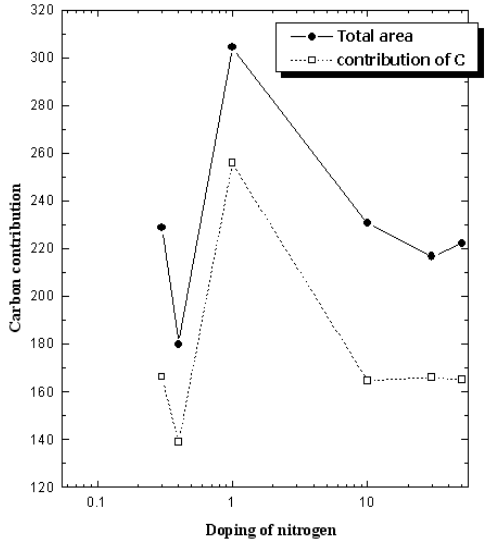


Fig. 3: Total photoresponse variation of nitrogen doped carbon films with NPP doping (0.3, 0.4, 1, 10, 30 and 50 mTorr).

As the summation of area-1 and area-2 curves gives the total contribution of carbon for the overall photoresponse it can be observed that the total photoresponse is dictated by contributions of carbon region since doping was performed only on carbon. In case of nitrogen doping the maximum photoresponse of the films was observed 1 mTorr NPP doping which is shown in figure 3.

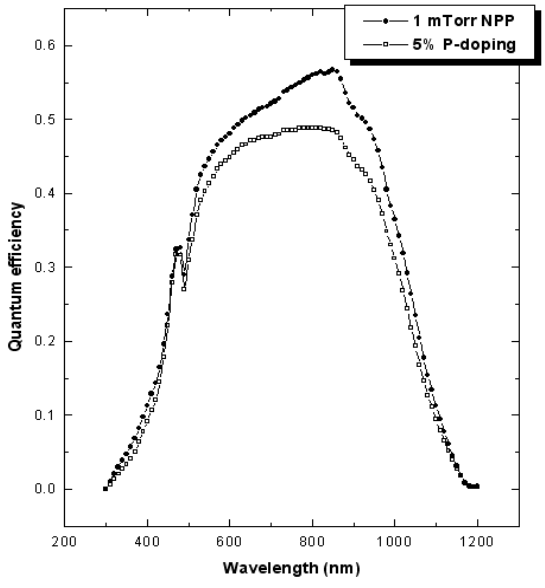


Fig. 4: Quantum efficiency vs. wavelength spectra for 1 mTorr NPP doping and 5% P-doping.

The photoresponse in this case also is observed to be strongly influenced by variation in carbon region. Previous similar study [11] also reveals that the conversion efficiency depends strongly on doping in the carbon layer.

Figure 4 gives a comparison of maximum photoresponse of the PV cells between 5% phosphorous and 1 mTorr NPP doping. The summation of the three area curves for NPP doping was found to be equal to 304.4 and for phosphorous it was 272.505. The maximum photoresponse was observed in case of nitrogen doping. The photoresponse characteristic deteriorates with higher dopant concentration in both cases.

The area curves were divided by the total area to get the normalized value. The percentage normalized values of the area curves gives the relative contributions of carbon and silicon. It is observed that the normalized contribution in the carbon region remain almost unchanged in case of P doped PV cell whereas, for N doped cell the contribution in the carbon layer observed to increase which reveals the different doping mechanism of phosphorous and nitrogen in carbon.

The full width at half maximum, FWHM-1 and 2 curves indicate changes in sp^3 and sp^2 bonding, respectively. FWHM-2 curve for phosphorous doping increases beyond 5% doping, showing trends of increased sp^2 bonding in carbon. The third FWHM curve remains almost constant which relates to silicon contribution. In case of NPP doping, the FWHM changes initially for the small amount

of N content and remain almost constant up to 10mTorr NPP. With further increase of NPP the FWHM observed to change. The variation of the FWHM possibly related to the structural modifications due to defects passivation initially up to 0.4mTorr NPP. With moderate NPP the nitrogen is doped effectively. However, for higher NPP, there is formation of some kind of C-N alloy and structure of the film changes, therefore the FWHM is observed to change. Furthermore as C-N alloy is very hard material, there is also possibility of strain development at the carbon silicon interface. The variation of FWHM with NPP is shown in Fig. 5.

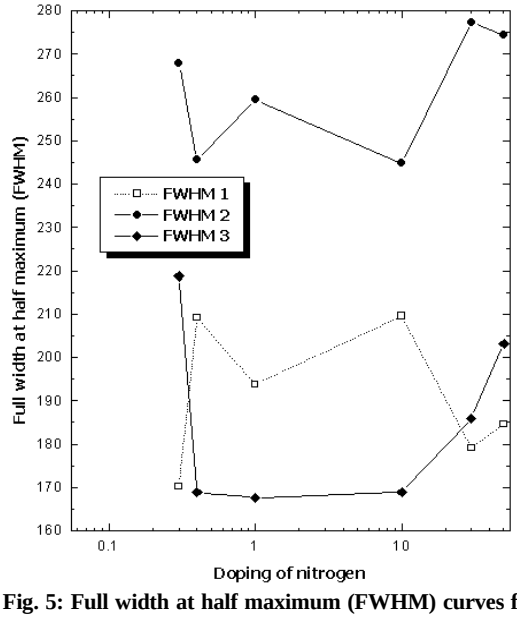


Fig. 5: Full width at half maximum (FWHM) curves for NPP doping.

VI. Conclusion

Phosphorous was doped in varying percentage (1%, 3%, 5% and 7%) on amorphous carbon. Photoresponse for the n-C/p-Si heterostructure varies with increased phosphorous doping. The maximum photoresponse was observed for 5% P-doping. The photoresponse of the film was dominantly dictated by variation in the carbon regions. After maximum photoresponse point there is trend of graphitization of the films. Maximum photoresponse for NPP doping was found for 1 mTorr NPP doping. The photoresponse decreases beyond that point. The overall photoresponse was similar to the variation of carbon in n-C region. With increased nitrogen content there is trend of structural modification. The relative silicon contribution remains almost constant for P doped PV cell whereas for the N doped cell the contribution is observed to vary with NPP which reveals different doping mechanism. The analyses reveal N doping in carbon films are better in terms of photoresponse characteristics.

VII. Acknowledgements

The authors would like to thank Prof. Tetsuo Soga, Nagoya Institute of Technology, for providing the data to make analyses in the present work.

References

[1] J. Robertson, "Diamond-like amorphous carbon", Mater. Sci. Eng. R Vol. 37, pp 129-281, 2002.

[2] V S Veerasamy, G A Amaratunga, C A Davis, A E Timbs, W I Milne, D R Mceknzie, "n-type doping of highly tetrahedral diamond-like amorphous carbon", J. Phys: Condensed matter vol. 5 pp L169-L174, 1993.

[3] S M Mominuzzaman, K M Krishna, T Soga, T Jimbo, M Umeno, "Optical absorption and electrical conductivity of amorphous carbon thin films from camphor: a natural source," Jpn. J. Appl. Phys., vol. 38, no. 2A, pp. 658-663, 1999.

[4]) M Rusop, S M Mominuzzaman, T Soga, T Jimbo and M Umeno, "Nitrogen doped n-type amorphous carbon films obtained by pulsed laser deposition with a natural camphor source target for solar cell application," J. Phys: Condensed matter 17 1929-1946 (2005).

[5] M. Rusop, S. M. Mominuzzaman, T. Soga, T. Jimbo and M. Umeno, "Characterization of phosphorus-doped amorphous carbon and construction of n-carbon/p-silicon heterojunction solar cells," Jpn. J. Appl. Phys., Vol. 42, 2003, pp.2339-2344.

[6] M Kaukonen and R M Nieminen, "Nitrogen doping of amorphous carbon surfaces," Physical review letters (1999).

[7] K. Kadas, G. G. Ferenczy and S. Kugler, "Theory of dopant pairs in four-fold coordinated amorphous semiconductors", J.. Non–Crystal. Solids, vol. 227-230, pp. 367-371, 1998.

[8] M. T. Kuo, P. W. May, A. Gunn, M. N. R. Ashfold and R. K. Wild, "Studies of phosphorus doped diamond-like carbon films", Diamond Related Mater. vol. 9, pg 1222-1227,, 2000.

[9] Zhou Z B, Cui R Q, Pang Q J, Hadi G M, Ding Z M and Li W Y, "Schottky solar cells with amorphous carbon nitride thin films prepared by ion beam sputtering technique", Sol. Energy Cells, vol. 70, pp. 487-493, 2002.

[10] S. M. Mominuzzaman, M. Rusop, T. Soga, T. Jimbo and M. Umeno, "Nitrogen doping in camphoric carbon films and its application to photovoltaic cell", Solar Energy Materials and Solar Cells, vol. 90, pp.3238-3243, 2006.

[11] T. Xuemin, D. K. Mishra, T. Soga, T. Jimbo, M. Umeno, "Amorphous carbon solar cell deposited by pulsed laser deposition", Photovoltaic Energy Conversion, vol. 1, pp 240-243, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Semiconducting Carbon Thin Film Deposition on Silicon by Electroplating

Muhammad Athar Uddin[1,2], Sharif M Mominuzzaman[1]

[1]Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology BUET, Dhaka-1000, Bangladesh
[2]Department of Electrical & Electronic Engineering, International Islamic University Chittagong, Dhaka Campus, Dhanmondi -3, Dhaka 1205, Bangladesh
E-mail: athar_01bd@yahoo.com

Abstract - Carbon thin films were deposited on silicon (Si) substrates by electrolysis of methanol. The effect of camphor ($C_{10}H_{16}O$) — a natural source, incorporation in methanol is investigated. Camphor with varying amount (2%, 4%, 6% and 8%) was mixed in methanol solvent to prepare the electrolytes. Silicon substrates were mounted on the negative electrode. Remarkable change in the variation of current density as a function of applied potential was observed with camphor content. For Si substrates current density was highest for the 6% camphor in methanol solution. P^H of various solutions before and after deposition has been analyzed. Camphor has an influence on P^H. The films was characterized by optical microscopy, scanning electron microscopy (SEM) and fourier transform of infrared (FTIR) spectroscopy. From optical microscopy and SEM micrograph sharp differences between deposited films are observed. FTIR spectroscopy analyses indicate presence of both sp^2 and sp^3 C-H stretch network therefore, reveals semiconducting nature of the deposited carbon films. The ratios of sp^3/sp^2 carbon bonding in deposited films are obsered to vary with camphor in methanol solution.

I. Introduction

There have recently been two important advances in the science of crystalline carbon— the discovery that diamond can be readily grown by vapor deposition and the discovery of a third allotrope of carbon, a molecular crystal of the fullcrene molecule, 'buckyball' C_{60} [1]. There has been a parallel advance in effort in disordered carbons. The range of disordered carbons is wide covering soots, chars, carbon fibres, glassy carbon and evaporated amorphous carbon. These carbons are basically sp^2 bonded. A range of new preparation methods has produced forms of amorphous carbon (a-C) and hydrogenated amorphous carbon (a-C: H), which is mechanically hard, infrared transparent and chemically inert. They are finding immediate applications as hard coating materials for magnetic disc drives or as antireflective coatings for infrared windows. Their beneficial properties arise from the sp^3 component of their bonding and these carbons are frequently called diamond like carbon (DLC). In general, such carbons can be fully amorphous or contain crystalline inclusions. This field of non-crystalline carbons is of interest both technologically to materials scientists' and also at a more fundamental level to solid-state chemists and physicists. Precursors and method of deposition of carbon films are the two dominating factors that dictate the optical and electrical properties of the film. And hence these two factors are strongly considered in order to obtain desired carbon thin film having certain optical and electrical properties required for application in various opto-electronic devices. Therefore, researches on finding alternative precursor materials and simple method of deposition have been getting priority all the time. In connection with this research, camphor ($C_{10}H_{16}O$) has been found as an alternative precursor material because it has some advantages [2].

Interest in depositing of carbon thin film has been motivated by properties of this material and the demand of modern technologies, especially those associated with development in the electronic industry. These properties include extreme hardness, chemical inertness, high electrical resistively, high dielectric strength, optical transparency and high thermal conductivity. These properties are tunable from insulating diamond to semi-metallic graphite and therefore, made these films extremely useful in a variety of applications. Deposition techniques that may provide advantages in these applications are of considerable interest. Many studies have been reported on the preparation of carbon thin films. These include chemical vapor deposition [3]-[4], pulsed laser deposition [2],[5], ion-beam sputtering [6] etc. All the above methods are vapor deposition techniques. However, a deposition of DLC films in the liquid phase is seldom reported. There is experimental evidence that most materials that can be deposited from the vapor phase can also be deposited in the liquid phase using electroplating techniques and vice versa. Enlightened by this conclusion, Namba [7] first attempted to grow diamond phase carbon films in the liquid phase with the aid of an organic solution such as ethanol at a temperature less than 70^0C. DLC films had been obtained there, but the choice of a suitable solution was limited. Then Suzuki *et al.* [8] made an attempt to deposit carbon films by electrolysis of a water-ethylene glycol solution. Graphite carbon had been obtained according to their result. Nevertheless, ethylene glycol is a viscous solution, which will cause some difficulty in cleaning the substrate

after deposition. Hao Wang *et al.* [9] deposited film by using methanol solution. Methanol is selected because its polarizability and conductivity are stronger than those of ethanol and the structure of methanol is even closer to that of diamond. We have reported semiconducting carbon films by ion beam sputtering [6] and pulsed laser ablation [2],[5] obtained from camphor ($C_{10}H_{16}O$), a natural source. In forming carbon films camphor has some advantages over graphite. Graphite is purely sp^3 hybridized whereas camphor consists of both sp^2 and sp^3 hybridized carbon in its structure. Hydrogen in a-C films modifies the properties of the films and introduces many sp^3 sites. Hydrogen passivates the dangling bond in the gap states and also tailors the opto-electronic properties of the film. So while using graphite as the precursor, additional hydrogen gas/ions have to be supplied but camphor has hydrogen abundantly in its structure. Furthermore, the presence of sp^3-hybridized bonds in camphor molecule plays a beneficial role in the deposition of carbon films especially in DLC films. Based on the observations of camphoric carbon films it is suggested that camphor might be suitable candidate as starting materials for semi conducting carbon films in electronic applications.

In the present article, carbon thin film deposition was attempted by the electroplating technique using methanol (CH_3OH), as electrolyte. From our earlier experience with camphor as a starting precursor for the preparation of semiconducting carbon films [2],[5]-[6],[10]-[11] we have mixed camphor with methanol. Camphor is a natural source, abundantly available in Asian countries like Japan, China and India. Varying amount (2%, 4%, 6% and 8%) of camphor was mixed to prepare camphor-methanol solution. Silicon (Si) substrates were used for the deposition. Some characteristics of the methanol and camphor-methanol electrolytes are investigated.

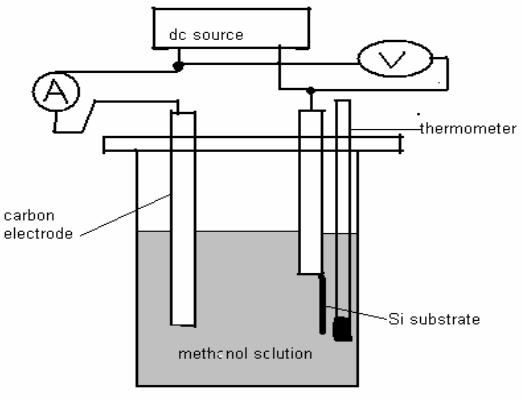


Fig. 1 Schematic diagram of the deposition system.

II. Experimental Details

For the deposition of carbon films on silicon, an experimental set up is prepared. A schematic diagram of the system is shown in figure 1. Silicon with a size of $3.2\times1.5\times0.1$ cm^3, have mounted on the negative electrode. Prior to deposition the substrates were cleaned in electronic grade acetone (CH_3COCH_3) and methanol (CH_3OH) for 5 minuets at 55°C. The distance between the substrate and positive electrode was set to 1.25 cm. The potential applied to the substrate could be changed from 0 V to 2500 V. A thermometer was adjusted to the system to measure the temperature of solution during the deposition. The P^H measurements were done for different electrolytes before and after deposition. The surface morphology of the deposited films were examined by optical microscopy, scanning electron microscopy (SEM) and fourier transform of infrared (FTIR) spectroscopy.

III. Results and Discussions

The substrate current and the P^H play an important role in film formation from an organic solution. Higher current density indicates more polarized charge particles move from solution to electrode, which may have some effect on the growth rate of film. The role of camphor on deposition rate can be understood from the curve of P^H as a function of camphor in methanol (figure 2). P^H of the solution increases with increasing the percentage of camphor in methanol. The P^H of the methanol containing 1% camphor is about 8. The P^H has increased with camphor content and for the methanol containing 20% camphor the P^H is 8.63. The variation of P^H with camphor content indicates the influence of camphor.

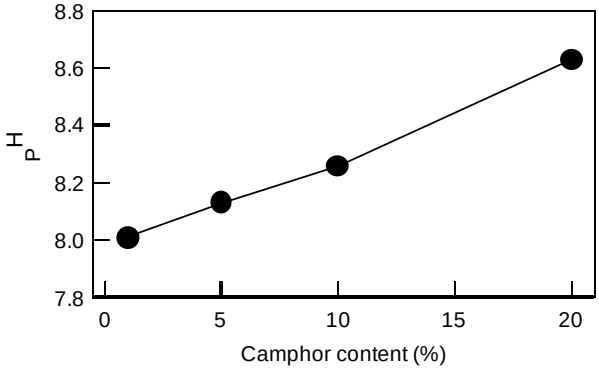


Fig. 2 P^H as a function of camphor in methanol.

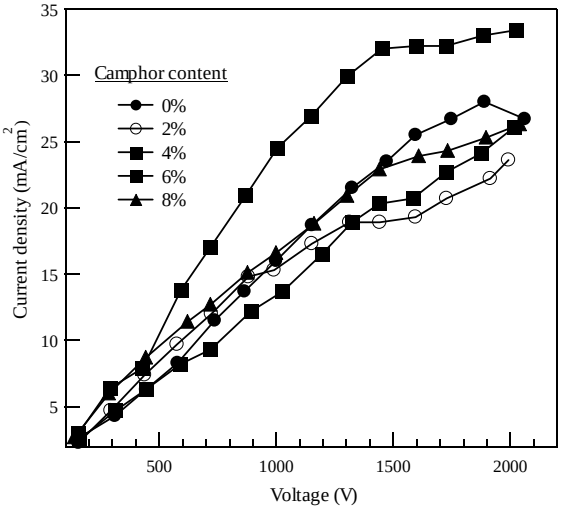


Fig. 3 Current density as a function of applied voltage for different percentages of camphor.

The optimum amount of camphor in methanol solution was measured. The experiment was done for 2%, 4%, 6% and 8% of camphor. Current density was measured for various applied voltages with respect to different electrolytes (camphor content) for Si substrates.

Current density as a function of applied voltage for different percentage of camphor in methanol solution was compared in figure 3. For 0% camphor content (only methanol solution) a moderate current density pattern was found. By adding little camphor (2%) in the solution the current density is decreased from that of only methanol solution and continued decreasing for 4% camphor in the lower voltage region. Then, with further incorporation of camphor (6%), the current density is increased and reaches to a maximum value, but again current density decreases for more camphor added (8% and higher).

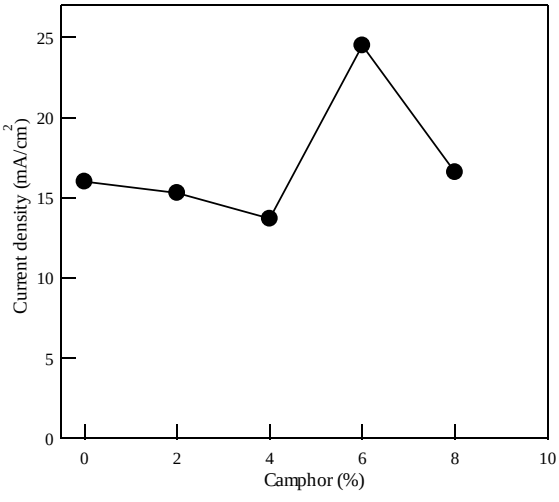


Fig. 4 Current density as a function of % of camphor in methanol at 1000 V

For simplicity and easy assimilation a current density for applied potential of 1000V as a function of camphor solution at room temperature is shown in figure 4. P^H of solution before and after deposition is measured. P^H of same solution is changed with deposition. For example during depositing on a substrate in a methanol solution with 10% camphor the P^H of the solution was 8.26 before deposition and 7.37 after deposition. The difference in P^H before and after deposition indicates that camphor is incorporated in the film.

The surface morphology of the films is observed by optical microscope (400X, Figure 5, 6 and 7). It is found that there are sharp differences between electrodeposited substrates and pure substrates. These differences indicate formation of films. Again film pattern for different percentage of camphor is different. So camphor has a role on film formation.

The surface morphology of the films is also observed by scanning electron microscopy (SEM). Figure 8 shows SEM micrograph of a film deposited on Si substrate (Magnification 1000x) in only methanol solution for 8 hrs. Figure 9 shows SEM micrograph of a film deposited on Si substrate (Magnification 1000x) with 6% camphor in methanol solution for 8 hrs. Small grains here show that carbon films are deposited on Si. Also changes are observed in micrographs of electrodeposited Si substrates when camphor is incorporated in methanol solution. So camphor has influences on carbon thin film formation.

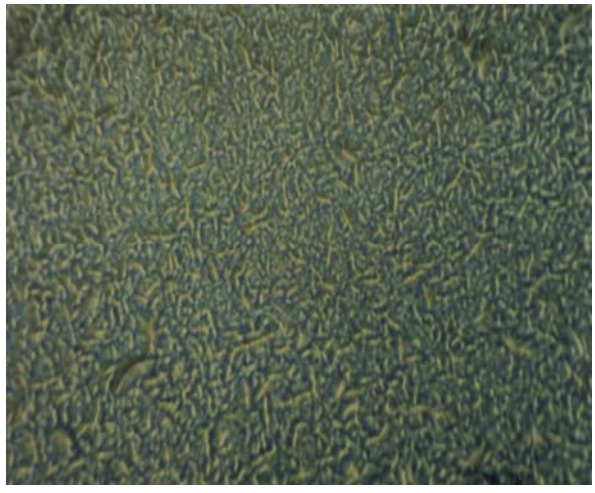


Fig. 5 Pure silicon substrate observed in optical microscope.

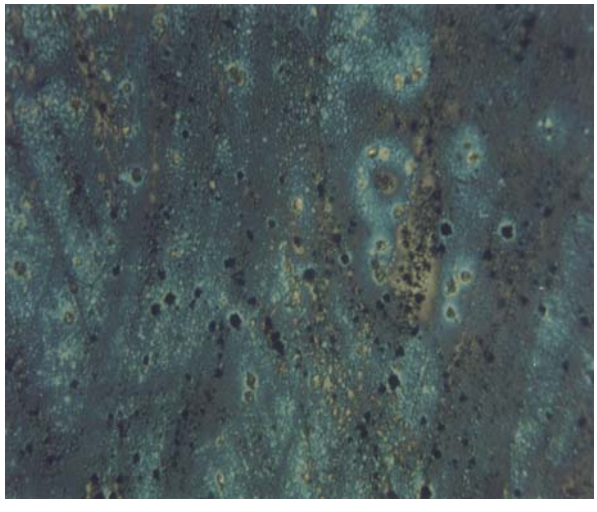


Fig.6 Carbon thin film deposited in 0% methanol solution on Si substrate observed in optical microscope.

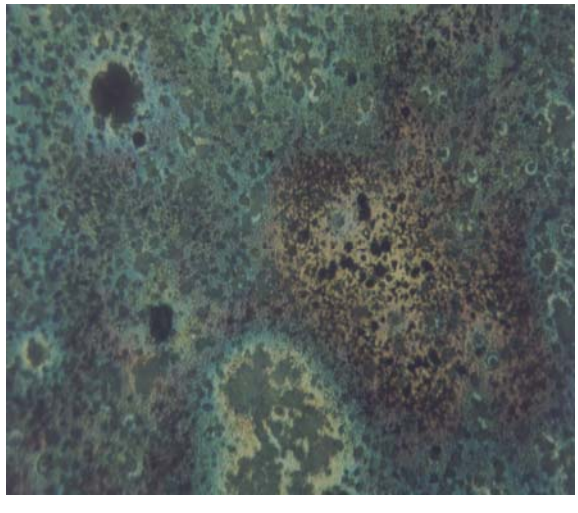


Fig.7 Carbon thin film deposited in 6% methanol solution on Si substrate observed in optical microscope.

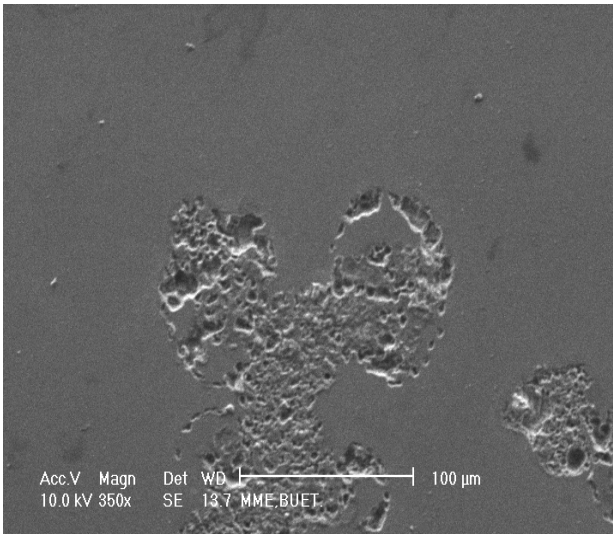


Fig. 8 SEM micrograph of the film deposited in only methanol solution for 8 hrs on Si substrate.

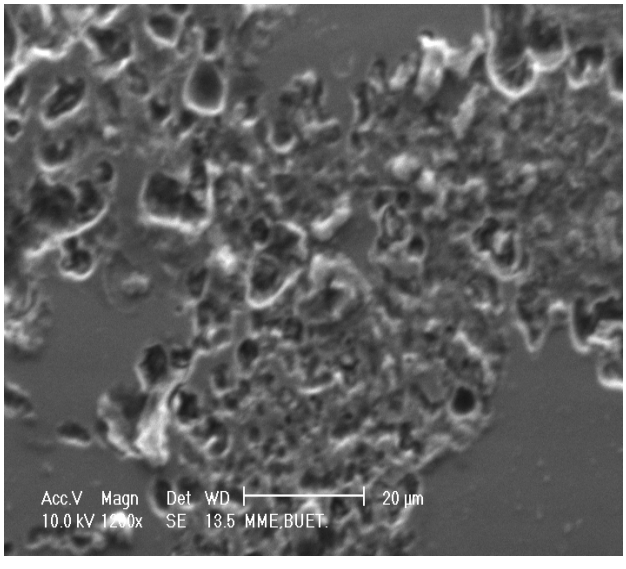


Fig. 9 SEM micrograph of the film deposited in 6% camphor in methanol solution for 8 hrs on Si substrate.

The electrodeposited Si films are also examined by FTIR spectroscopy. FTIR spectra show absorption peaks in between 1250 to 1750 cm^{-1} which is the characteristic of diamond like carbon (DLC)/ amorphous carbon (a-C) films. The peaks observed to vary with camphor in methanol solution. The films deposited from pure methanol and 6% camphor in methanol is shown in Figs. 10 and 11 respectively. The absorption for the film deposited from 6% camphor-methanol solution is very different from the film deposited from pure methanol solution. Absorption peak (minima) around 2850 and 2950 cm^{-1} in FTIR transmittance spectra are related to sp^3 C-H stretch and sp^2 C-H stretch respectively, in carbon network [12].

The FTIR spectra for the films deposited from pure methanol and from 6% camphor in methanol in the range of 2500 to 3000 cm^{-1} are shown in figure 12 and 13 respectively. In our analysis it is observed that there is presence of sp^3 and sp^2 bonded carbon network in the deposited films and the amount of sp^3 and sp^2 varies with camphor content in methanol.

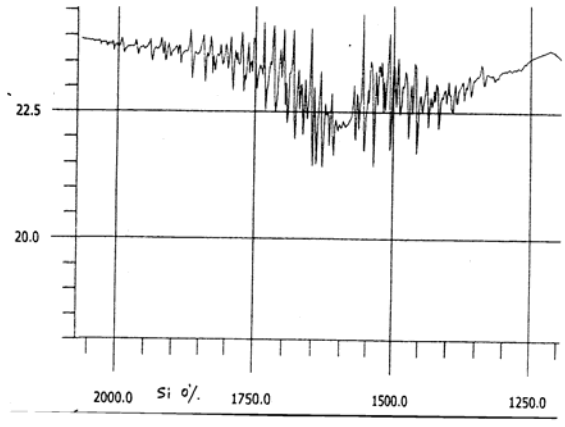


Fig. 10 FTIR response of a Si film of 0% camphor of 4.5 hour (From 1000 cm^{-1} to 2000 cm^{-1})

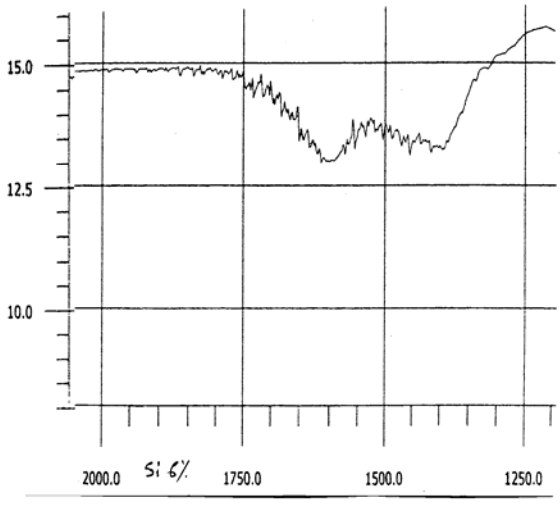


Fig. 11 FTIR response of a Si film of 6% camphor of 4.5 hour (From 1000 cm^{-1} to 2000 cm^{-1})

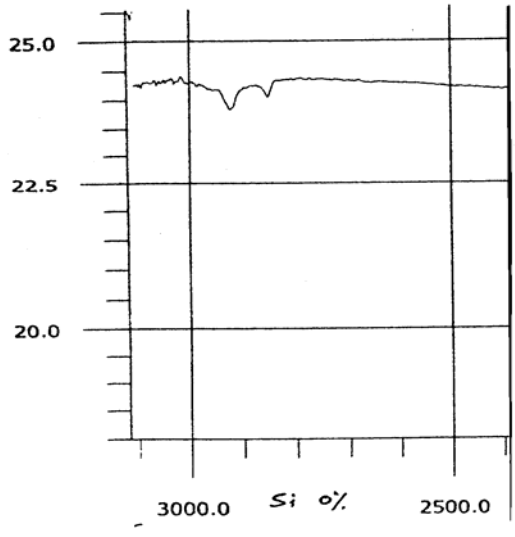


Fig. 12 FTIR response of a Si film of 0% camphor of 4.5 hour (from 2500 cm^{-1} to 3000 cm^{-1})

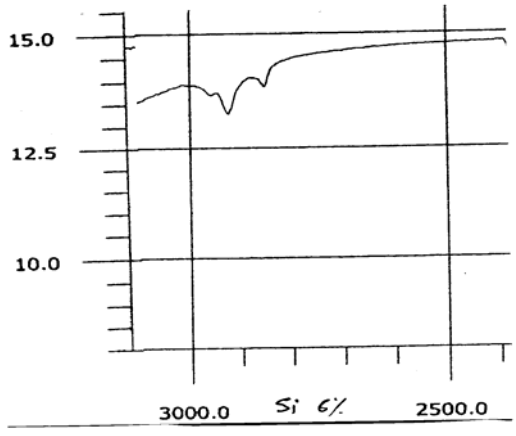


Fig. 13 FTIR response of a Si film of 6% camphor of 4.5 hour (From 2500 cm^{-1} to 3000 cm^{-1})

The ratio of sp^3/sp^2 is found 0.45 for 0% camphor in methanol solution and that for 6% camphor is 0.66. In carbon the band gap is observed to vary with sp^3 and sp^2 ratio. Diamond has only sp^3 bonding configurations while graphite has only sp^2 bonding configurations. In DLC or a-C, the band gap increases with sp^3 bonding. Therefore increase of sp^3 with camphor in methanol solution reveals the increase of the bond gap for the film deposited from camphor.

IV. Conclusions

Organic solution including methanol and different percentage of camphor in methanol has been used as electrolytes to deposit carbon thin films on Si substrates. Conductivity of the electrolytes is studied by measuring the current density as a function of applied voltage. Role of camphor on electro deposition was examined by measuring the P^H of the solution before and after deposition for different percentage of camphor.

The role of camphor on electro deposition is observed. It is found that the current density decreases initially with camphor from that of only methanol solution, and increases thereafter. However, there is a limit for camphor to add in methanol solution for increasing current density. Maximum current density for 6% of camphor in methanol solution is observed. With camphor incorporation in methanol the current density is varied and the deposition rate of the carbon film can be controlled.

P^H of the solution increases by increasing percentage of camphor in methanol. P^H of any solution decreases after deposition from that of before deposition. So camphor has an influence on carbon thin film deposition by electroplating through methanol solution.

FTIR spectroscopic analyses reveal the presence of sp^3 and sp^2 bonded carbon in the deposited films and can be controlled with amount of camphor in methanol solution. Furthermore as the optoelectronic and mechanical properties vary with sp^3/sp^2 ratio, it is also possible to tune the properties of the carbon film with camphor addition by cheap and very simple electroplating technique.

V. Acknowledgements

The authors would like to thank Prof. A. Wahab, Dept of Chemistry, BUET, for providing FTIR facility.

References

[1] M. S. Dresselhaus, G. Dresselhaus and P. C. Eklund, "Science of fullerenes and carrbon," Academic Press. Inc., 1996.

[2] S. M. Mominuzzaman, T. Soga, T. Jimbo, and M. Umeno, "Camphoric carbon soot: a new target for deposition of diamond-like carbon films by pulsed laser ablation," *Thin Solid Films*, vol. 376, pp. 1-4, 2000.

[3] H. A. Yu, Y. Kaneko, S. Yoshimura and s. Otani, "Photovoltaic cell of carbonaceous film/n-type silicon," Appl. Phys. Lett., vol. 68, no. 4, pp. 547-549, 1996.

[4] M. Weiler, S. Sattel, T. Giessen, K. Jung, H. Ehrhardt, V. S. Veerasamy and J. Robertson, "Preparation and properties of highly tetrahedral hydrogenated amorphous carbon," Physical Review B, vol. 53, pp. 1594-1608, 1996.

[5] M.Rusop, S. M. Mominuzzaman, T. Soga, T. Jimbo and M. Umeno, "Characterization of Phosphorus-Doped Amorphous Carbon and Construction of n-Carbon/p-Silicon Heterojunction Solar Cells", Jpn. J. Appl. Phys., vol. 42, 2003, pp.2339-2344.

[6] S. M. Mominuzzaman, K. M. Krishna, T. Soga, T. Jimbo and M. Umeno, "Raman Spectra of Ion Beam Sputtered Amorphous Camphoric Carbon Thin Films", Carbon, vol. 38, 2000, pp. 127-131.

[7] Y. Namba, "Attempt to grow diamond phase carbon films from an organic solution," *J. Vac. Sci. Technol. A*, vol. 10, pp. 3368-3370, 1992.

[8] T. Suzuki, Y. Marita, T. Yamazaki, S. Wada, and T. Noma, "Deposition of carbon films by electrolysis of a water-ethylene glycol solution," *J. Mater. Sci.*, vol. 30, pp. 2067-2069, 1995.

[9] Hao wang, Ming-Rong Shen, Zhao-Yuan Ning, and Chao Ye., "Deposition of diamond-like carbon films by electrolysis of methanol solution," *Appl. Phys. Lett.*, vol. 69, pp. 1074-1076, 1996.

[10] Mominuzzaman, S.M., Krishna, K. M., Soga, T., Jimbo, T.and Umeno, M. "Optical Absorption and Electrical Conductivity of Amorphous Carbon Thin Films from Camphor: a Natural Source", *Jpn J Appl Phys*; 38(2A), 658-663 (1999).

[11] S. M. Mominuzzaman, M. Rusop, T. Soga, T. Jimbo and M. Umeno, "Nitrogen Doping in Camphoric Carbon Films and its Application to Photovoltaic Cell", 14th International Photovoltaic Science and Engineering Conference (PVSEC-14), Bangkok, Thailand, January 26 – 30, 2004.

[12] H.C. Barshilia, Somna Sah, B.R. Mehta, V.D. Vankar, D.K. Avasthi, Jaipal and G.K. Mehta, "Microstructural modification in diamond-like carbon thin films caused by high energy ion irradiation," *Thin Solid Films*, vol. 258, pp. 123-127, 1995.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Three-Dimensional Motion Control using Embedded Controller and FPGA Technology

Satyam, R.D.Kamble, Dhanashri, V.K. Sharma[1]
Maharashtra Academy of Engineering, Pune, 1:JMI, New Delhi

Abstract:- In the paper, controlling of electrodes of a numerically controlled machine is done by a microcontroller, which is to be interfaced with a number of peripherals such as keyboard, EEPROM, display screen and stepper motors using FPGA platform. Interfacing the different modules to the controller through FPGA using VHDL language has been reported. The existing code of microcontroller for one dimension (Z- direction) motion to the three dimensional (X, Y, Z) motion in assembly language has been extended. Successful completion of the work involved motion of 3- motors, each for a particular direction (X, Y, Z). The three motors are responsible for 3 dimensional motion of Electrode. The instantaneous position of the electrode while operation is displayed. The electrode automatically moves to the next position as per the input coordinates. The manual mode of movement of the electrode to the desired position is preserved.

I. INTRODUCTION

A lathe is a machine tool which spins a block of material to perform various operations such as cutting, sanding, knurling, drilling, or deformation with tools that are applied to the work piece to create an object which has symmetry about an axis of rotation. Numerically Controlled machine is an automation of lathe machine, which needs to be programmed for desired functioning. In the paper, Numerically Controlled machine is used as Electric Discharge Machine for making dyes. Dyes are used to cast any raw material into the desired shape. To manufacture dyes, the raw block is exposed to controlled electric spark, shedding the unwanted parts and also moving the block, which is manually done. In present thesis, the block is kept constant and the electrode is moved to the desired coordinates by means of programmable devices along with its drilling action.

Main objective of the paper is to control the motion of NC machine, which will be used for die manufacturing. Its motion is controlled by a microcontroller. This controller is interfaced with different modules through FPGA.

Lathes are used in woodturning, metalworking, metal spinning, and glass working. The material is held in place by either one or two centers, at least one of which can be moved horizontally to accommodate varying material lengths. Examples of objects that can be produced on a lathe include candlestick holders, cue sticks, table legs, bowls, baseball bats, crankshafts and camshafts.

Numerical control or numerically controlled (NC) machine tools are machines that are automatically operated by commands that are received by their processing units. NC machines made it possible for large quantities of the desired components to be very precisely and efficiently produced in a reliable repetitive manner.

Very High Speed Integrated Circuit Hardware Description Language (VHDL) offers common programming language so that designers from different communities share work with ease. Need for circuits having low cost, consuming less power, providing high performance and having smaller dimensions increased. High density programmable logic devices and VHDL are key elements and are well suited for designing with programmable logic devices.

VHDL describes the behavior and structure of electronic systems, but is particularly suited as a language to describe the structure and behavior of digital electronic hardware designs, such as ASICs and FPGAs as well as conventional digital circuits.

Simulation and synthesis are the two main kinds of tools, which operate on the VHDL language. The Language Reference Manual does not define a simulator, but unambiguously defines what each simulator must do with each part of the language.

VHDL does not constrain the user to one style of description. VHDL allows designs to be described using any methodology - top down, bottom up or middle out! VHDL can be used to describe hardware at the gate level or in a more abstract way. Successful high-level design requires a language, a tool set and a suitable methodology.

1. VHDL does not restrict the user to one type of description only as it can be used as behavioral, structural, data flow and mixed language.
2. Although VHDL was intended for electronics it has become a universal modeling language. It is used for modeling and simulation of electromechanical, hydraulically chemical and other systems.
3. The language support flexible design methodologies like top down bottom up or mixed.
4. The language has elements that make large-scale design modeling easier, example component, functions, procedures and packages.
5. The language supports hierarchy that is digital system can be modeled as a set of interconnected components or subcomponents.
6. It supports both synchronous and asynchronous timing models.
7. Test benches can be written using the same language to test other VHDL models.

II. PROGRAMMABLE LOGIC DEVICES

A programmable logic device or PLD is an electronic component used to build digital circuits. Unlike a logic gate, which has a fixed function, a PLD has an undefined function at the time of manufacture. Before the PLD can be used in a circuit it must be programmed. Other names for this class of device are Programmable Logic Array (PLA), Programmable Array Logic (PAL), and Generic Array Logic (GAL).

With programmable logic devices, designers use inexpensive software tools to quickly develop, simulate, and test their designs. Then, a design can be quickly programmed into a device, and immediately tested in a live circuit. The PLD that is used for this prototyping is the exact same PLD that will be used in the final production of a piece of end equipment. There are no NRE costs and the final design is completed much faster than that of a custom, fixed logic device.

Another key benefit of using PLDs is that during the design phase customers can change the circuitry as often as they want until the design operates to their satisfaction. That's because PLDs are based on re-writable memory technology - to change the design, the device is simply reprogrammed. Once the design is final, customers can go into immediate production by simply programming as many PLDs as they need with the final software design file.

Inside each PLD is a set of fully connected macrocells. These macrocells are typically comprised of some amount of combinatorial logic (AND and OR gates, for example) and a flip-flop. In other words, a small Boolean logic equation can be built within each macrocell. This equation will combine the state of some number of binary inputs into a binary output and, if necessary, store that output in the flip-flop until the next clock edge. Of course, the particulars of the available logic gates and flip-flops are specific to each manufacturer and product family. But the general idea is always the same.

The two major types of programmable logic devices are field programmable gate arrays (FPGAs) and complex programmable logic devices (CPLDs).

Complex Programmable Logic Devices (CPLDs)

As chip density is increased, it is natural for the PLD manufacturers to evolve their products into larger (logically, but not necessarily physically) parts called Complex Programmable Logic Devices (CPLDs). For most practical purposes, CPLDs can be thought of as multiple PLDs (plus some programmable interconnect) in a single chip. The larger size of a CPLD allows you to implement either more logic equations or a more complicated design. In fact, these chips are large enough to replace dozens of those pesky 7400-series parts.

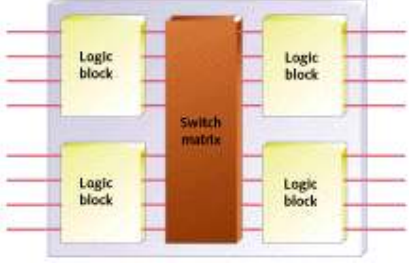


Fig. 1. Block diagram of a typical CPLD.

Block diagram of a hypothetical CPLD is shown in Fig. 1. Each of the four logic blocks shown there is the equivalent of one PLD. However, in an actual CPLD there may be more (or less) than four logic blocks. These logic blocks are themselves comprised of macrocells and interconnect wiring, just like an ordinary PLD.

III. DESIGN DESCRIPTION

Interfacing of FPGA with peripherals is shown in Fig. 2, and is described here.

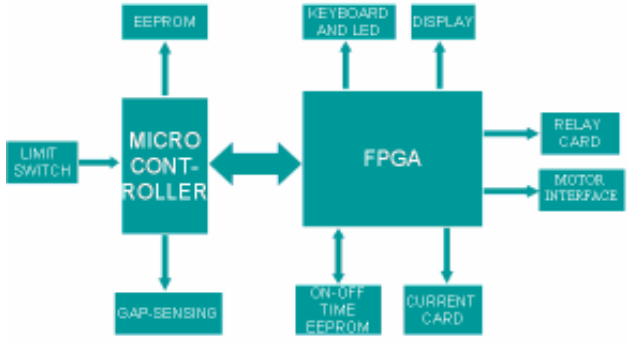


Fig. 1. Interfacing of FPGA with microcontroller.

FPGA: Since, microcontroller has limited number of ports, FPGA is used to interface different modules with microcontroller.
Display: It shows parameters of different directions in which motion is to take place.
Controller: It is used to control the action of 3-drives which rotate the motors.
EEPROM: Programmed to control ON/OFF time of square pulses, which are input to current card, is stored in this IC.
Relay Card: It is used to connect electrodes with three phase power supply, in case of sparking.
Scale Interface: Scales movement is controlled by programmed FPGA.
Keyboard and LED: Input of the dimensions for scale movement is given through keyboard. LED's used to indicate whether relay is on or not.
Current Card: Square pulses are applied to it, through FPGA, which is used to produce spark via electrodes.

The paper deals with the interfacing of controller, keyboard, and stepper motor drive.

IV. INTERFACING WITH MICROCONTROLLER

The system developed makes use of PIC18F442 microcontroller. Algorithm for Controller and FPGA interfacing is described here.

1) If RESET=1 then all the variables (address, data, interrupt, read, write, load) are assigned the value '0' and the state is IDLE.
2) On rising edge clock, if state is IDLE then,
   if cs = '1' and rs = '0' and r_wx = '0' then
   state = WAIT_FOR_EN_ADDR
   if cs = '1' and rs = '0' and r_wx = '1' then
   state = ENABLE_SEND_ADDR
   if cs = '1' and rs = '1' and r_wx = '0' then
   state = WAIT_FOR_EN_DATA
   if cs = '1' and rs = '1' and r_wx = '1' then
   state = CHECK_ENABLE
   else
   state = IDLE
3) When state = WAIT_FOR_EN_ADDR then,
   if enable signal is high then
   state = INPUT_ADDR
   else it will remain in same state.
4) When state = INPUT_ADDR then,
   if enable signal=0 then,

```
        according to the signal dro_edm, address of DRO/EDM will
be stored in the address buffer register.
        state = WAIT_FOR_END
 5)     When state = WAIT_FOR_END then
          if cs = '0' and rs = '1' and r_wx = '1' then,
          state = IDLE
          else it will remain in the same state.
 6)     When state =  WAIT_FOR_EN_DATA then,
          if  enable signal is high  then,
          state = INPUT_DATA and reset the following flags
          wait_flag = 0
          initialize = 0
          busy = 0
          else it will remain in same state
 7)     When state =  INPUT_DATA then,
        in the first clock if enable is low then
        according to  the dro_edm signal,
        DRO/EDM data will get stored in the data buffer register.
        In the next clock, check address and accordingly write data or
        store in temporary register for conversion.
 8)   When state = WAIT_FOR_DATA_END then,
        if cs = '0' and rs = '1' and r_wx = '1' then
        end of write data
        else it will remain in same state.
 9)  When state = ENABLE_SEND_ADDR then,
        if en = '1' then
        state = GIVE_ADDR
        else it will remain in same state.
10)  When state = GIVE_ADDR  then,
        if en = '0' then  it will take the address
        else it will remain in same state.
10)  When state = READ_INDEX_REG_END then,
        if cs = '0' and rs = '1' and r_wx = '1' then
        reading of address is completed
        else remain in same state.
10)  When state = CHECK_ENABLE then,
       if en = '1' then
       state =  GIVE_DATA
       else remain in same state.
10)  When state = GIVE_DATA then,
       if en = '0' then  it will take the data from the given address
       else remain in same state.
```

V. INTERFACING WITH KEYBOARD

Algorithm for Key Board and FPGA interfacing is presented here.

```
 1)    if  RESET =1  then
       state = START
       signals strobe=0
       row reg.=0
       column reg.=0
       code=0
 2)    if  there occurs a rising edge event in clock and state = START
       stb=0
       row reg.=0
       column reg.=0
       state = DISCHARGE
 3)    When state = DISCHARGE
       wait until all the columns discharge to 1's.
       After discharging all columns to 1,
       state = DETECT
 4)    When state = DETECT
       Check for key press
       If (key pressed) then
       State = DEBOUNCE
       Else
       State = DETECT
 5)    When state = DEBOUNCE
       Wait for debounce time   50ms
       State = ENCODE 0
       row_reg = "111111111110"
 6)    When state = ENCODE 0
       if key press
       column_reg = column
       else
       state = ENCODE 1
       wait for releasing of key
       row_reg = "111111111101"
 7)    When state = ENCODE1
       if key press
       column_reg = column
       else
       state = ENCODE 2
       wait for releasing of key
       row_reg = "111111111011"
 8)    When state = ENCODE 2
       if key press
       column_reg = column
       else
       state = ENCODE  3
       wait for releasing of key
       row_reg = "111111110111"
 9)    When state = ENCODE3
       if  key press
       column_reg = column
       else
       state = ENCODE 4
       wait for releasing of key
       row_reg = "111111101111"
10)    When state = ENCODE4
       if key press
       column_reg = column
       else
       state = ENCODE 5
       wait for releasing of key
       row_reg = "111111011111"
11)    When state = ENCODE5
       if key press
          column_reg = column
       else
       state = ENCODE 6
       wait for releasing of key
          row_reg = "111110111111"
12)    When state = ENCODE6
         if  key press
         column_reg = column
       else
       state = ENCODE 7
       wait for releasing of key
       row_reg = "111101111111"
13)    When state = ENCODE7
         if  key press
         column_reg = column
       else
       state = ENCODE 8
       wait for releasing of key
       row_reg = "111011111111"
14)    When state = ENCODE8
       if key press
       column_reg = column
       else
       state = ENCODE 9
       wait for releasing of key
       row_reg = "110111111111"
15)    When state = ENCODE9
       if key press
       column_reg = column
       else
       state = ENCODE 10
       wait for releasing of key
       row_reg = "111111101111"
16)    When state = ENCODE10
       if key press
       column_reg = column
       else
       state = ENCODE 11
       wait for releasing of key
       row_reg = "101111111111"
17)    When state = ENCODE11
       if key press
       column_reg = column
       if  (DRO_EDM='0' or column =  "01111" )    then
       state = INTERRUPT
       else
       state= INT_ON_PRESS
18)    When state = INT_ON_PRESS
         strobe = 1
         code = table_out;
```

```
        state = INTERRUPT
19)   When state = INTERRUPT:
        strobe = 0
20)   When state = INT_ON_RELEASE
        code = table_out
```

VI. INTERFACING WITH MOTOR DRIVE

Algorithm for Motor Drive and FPGA interfacing is described below.

```
1) If reset=0
        enable signals for all three motors are set to 1
        direction and clock signals of all three signals
        are set to 0
3)  On the basis of select lines a particular motor gets
    selected such as

    sel0 = 0 & sel1 = 0 => MOTOR_X is selected
        enbx = 0
        clkx = clock
        dirx = not dir.

    sel0 = 1 & sel1 = 0 => MOTOR_Y is selected
        enby = 0
        clky = clock
        diry = not dir.

    sel0 = 0 & sel1 = 1 => MOTOR_Z is selected
        enbz = 0
        clkz = clock
        dirz = not dir.
For others,
        enbx = 1
        enby = 1
        enbz = 1
        dirx = 0
        diry = 0
        dirz = 0
        clkx = 0
        clky = 0
        ckz = 0
```

VII. PROGRAMMING MICROCONTROLLER

In this section, microcontroller PIC18F442 is programmed to communicate with different modules via FPGA. The program is developed for a microcontroller to respond when FPGA asks the controller for its services. Further the program for motion of electrode in single dimension is extended to a 3- Dimensional motion by programming the microcontroller.

Communication between FPGA and Microcontroller

Communication between FPGA and microcontroller is obtained through keyboard. Response of system on pressing key is follows:

- Key pressed.
- Rows and columns are detected.
- Code generated and stored in FPGA.
- FPGA controller sends strobe signal to microcontroller.
- Microcontroller gets interrupted.
- Micro controller leaves behind current task, sets flag for new key pressed and disable the interrupt.
- Micro controller returns to the left task, after completing which it reads code from FPGA.
- According to the code read it moves to its respective routine in Check Switch And Take Action loop.

Extending the code for motion in one dimension (Z-direction) to a three dimensional motion (X, Y, Z) is detailed below.

Algorithm For Motion

```
Auto mode given (spark button pressed).
        Flag for motion in Z- direction get set.
        If all ten blocks gets executed then
        If direction is upwards then
        If (num position > 3)
        Spark off (motion stops)
Else
        Reset the flag for motion in Z- direction
        Set the flag for motion in X-direction.
        If (display data < target position)
        Motor up.
Else
        Motor down.
        When (display data = target position)
        Reset flag for motion in X- direction.
        Set the flag for motion in Y- direction
If (display data < target position)
        Motor up.
Else
        Motor down.
When (display data = target position)
        Reset flag for motion in Y- direction.
        Set the flag for motion in Z- direction.
 If motor direction is downwards and block number =10
        then motion direction = UP .
 If (block number< 10) then increase the block
   hence the z co-ordinate gets increase and drilling action take place.
```

VIII. TESTING

In order to test the validity of code generated, 3-Stepper motors are interfaced along with the drives to FPGA and following testing are carried out:

1) Manual Mode

a) Z-Direction motion Testing: Z-Motor moves in anti-clockwise direction only till (Z+) Button is pressed and in clockwise direction when (Z-) Button gets pressed.
b) X-Direction motion Testing: X-Motor moves in anti-clockwise direction only till (X+) Button is pressed and in clockwise direction when (X-) Button gets pressed.
c) Y-Direction motion Testing: Y-Motor Motor moves in anti-clockwise direction only till (Y+) Button is pressed and in clockwise direction when (Y-) Button gets pressed.

2) Automatic Mode

In this mode Co-ordinate entry is first made and then following action took place when we pressed auto mode:

a) Z-Direction motion Testing: Z-Motor moves in anti-clockwise direction and in clockwise direction when moving for upward and downward directions, respectively.
b) X-Direction motion Testing: X-Motor moves in anti-clockwise direction and in clockwise direction when moving for upward and downward directions, respectively.
c) Y-Direction motion Testing: Y-Motor moves in anti-clockwise direction and in clockwise direction when

moving for upward and downward directions, respectively.

d) Automation

After checking the motion in each direction separately, code incorporating all three motions was tested in which Z-direction motor first moved DOWN (clockwise) as per the co-ordinate entered, moved UP (anti-clockwise) to neutral position after reaching the down limit. Then, started the X-direction motor after whose limit, Y-direction motor got started. After moving to next co-ordinate positions, Z-motor again started its motion and the same process was repeated for two more co-ordinates.

IX. RESULTS

The software is developed and interfaced with the hardware. The complete setup is shown in Fig. 3. The step signals driving the stepper motor in X-Y-Z- direction is shown in Fig. 4-5-6 respectively. Experimental results of the scheme are shown in Fig. 7-14, along with their description.

Fig. 3. Complete setup.

Fig. 4. Signal for driving the stepper motor- X-direction.

Fig. 5. Signal for driving the stepper motor- Y-direction.

Fig. 6. Signal for driving the stepper motor- Z-direction.

Fig. 7. Microcontroller Kit for interfacing.

Fig. 8. Power supply for system.

Fig. 9. Signal for driving the stepper motor- Y-direction.

Fig. 10. Keyboard interfacing with microcontroller and drill motors.

Fig. 11. Detail of keyboard for inputting the co-ordinates of drill bit.

Fig. 12. Reading showing the depth in z-direction.

Fig. 13. Shafts of stepper motor for-X-Y-Z co-ordinate.

Fig. 14. Micro-controller mounted on the back of the keyboard.

X. CONCLUSION

Motion of NC machine in single dimension is successfully implemented and enhanced for other two directions by interfacing the FPGA with Microcontroller. Other modules like extending the program for 3-dimension etc is carried out successfully. Preserving the manual mode for moving the electrodes, automation of the NC machine is achieved and the electrodes move to the next co-ordinate positions on their own, as entered by the user. Other than the application of NC machine in making dyes, the system can be extended to other job works such as drilling, grinding, cutting, shaping, threading etc.

REFERENCES

[1] Fujii, W., Yokoyama, T., 'Construction of FPGA Based Hardware Controller for Autonomous Decentralized Control for UPS Application,'12th International Power Electronics and Motion Control Conference, Aug. 2006, pp. 846–851.

[2] Ide, T., Yokoyama, T., 'A study of deadbeat control for three phase PWM inverter using FPGA based hardware controller,' IEEE 35th Annual Power Electronics Specialists Conference, PESC 2004, Vol. 1, pp. 50 – 53.

[3] Chakravarthy, N., Jizhong Xiao, 'FPGA-based Control System for Miniature Robots,' IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 2006, pp. 3399–3404.

[4] Ngoc Quy Le, Jung Uk Cho, Jae Wook Jeon, 'Application of Velocity Profile Generation and Closed-Loop Control in Step Motor Control System,' International Joint Conference SICE-ICASE, 2006, pp. 3593–3598.

[5] Krishnamurthy, P., Khorrami, F., 'Robust adaptive voltage-fed permanent magnet step motor control without current measurements,' IEEE Transactions on Control Systems Technology, Vol. 11, Issue 3, May 2003, pp. 415–425.

[6] Wei Zhenzhong, Zhang Guangjun, Li Xin, 'The application of machine vision in inspecting position-control accuracy of motor control systems,' Proceedings of the Fifth International Conference on Electrical Machines and Systems, ICEMS 2001, Vol. 2, pp. 787–790.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Situational Awareness Based on Neural Control of an Autonomous Helicopter During Hovering Manoeuvres

Igor Astrov and Andrus Pedai

Department of Computer Control, Tallinn University of Technology
Ehitajate tee 5, 19086 Tallinn, Estonia
E-mail: igor.astrov@dcc.ttu.ee

Abstract - This paper focuses on a critical component of the situational awareness, the neural network control of autonomous vertical flight for an unmanned aerial vehicle. Application of the proposed two stage flight strategy which uses two autonomous adaptive neural dynamical feedback controllers was carried out for a nontrivial small-scale helicopter model comprising five states, two inputs and two outputs. This control strategy for chosen helicopter model has been verified by simulation of hovering manoeuvres using software package Simulink and demonstrated good performance for fast situational awareness in real-time search-and-rescue operations.

I. Introduction

Situation awareness has been formally defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [1]. As the term implies, situation awareness refers to awareness of the situation. Grammatically, situational awareness (SA) refers to awareness that only happens sometimes in certain situations.

SA has been recognized as a critical, yet often elusive, foundation for successful decision-making across a broad range of complex and dynamic systems, including emergency response and military command and control operations [2].

The term SA have become commonplace for the doctrine and tactics, and techniques in the U.S. Army [3]. SA is defined as "the ability to maintain a constant, clear mental picture of relevant information and the tactical situation including friendly and threat situations as well as terrain". SA allows leaders to avoid surprise, make rapid decisions, and choose when and where to conduct engagements, and achieve decisive outcomes.

The tactical unmanned aerial vehicle (TUAV) is one of the key tools to gather the information to build SA for all leaders. The TUAV is the ground maneuver commander's primary day and night system. The TUAV provides the commander with a number of capabilities including:

- Enhanced situational awareness.
- Target acquisition.
- Battle damage assessment.
- Enhanced battle management capabilities (friendly situation and battlefield visualization).

The combination of these benefits contributes to the commander's dominant SA allowing him to shape the battlefield to ensure mission success and to maneuver to points of positional advantage with speed and precision to conduct decisive operations. Some conditions for conducting aerial reconnaissance with TUAVs are as follows.

- Time is limited or information is required quickly.
- Detailed reconnaissance is not required.
- Extended duration surveillance is not required.
- Objective is at extended range.
- Verification of a target is needed.
- Threat conditions are known and risk to ground assets is high.
- Terrain restricts approach by ground units.

A small-scale unmanned helicopter offers many advantages, including low weight and cost, the ability to fly within a narrow space and the unique hovering and vertical take-off and landing (VTOL) flying characteristics.

Autonomous vertical flight is a challenging but important task for TUAVs to achieve high level of autonomy under adverse conditions. The fundamental requirement for vertical flight is the knowledge of the height above the ground, and a properly designed controller to govern the process.

This paper presents our research results in the study of vertical flight (take-off and hovering cases) neural control of autonomous unmanned small-scale helicopters which make such SA task scenario as "go-search-find-return" possible.

With the SA strategy, we proposed a two stage flight control procedure using two ADAptive LInear NEuron neural networks (ADALINE NNs) to address the dynamics variation and performance requirement difference in initial and final stages of flight trajectory for an unmanned small-scale helicopter.

The contribution of the paper is twofold: to develop new schemes appropriate for SA enhancement using TUAVs by neural network control of vertical flight of autonomous unmanned small-scale helicopters in real-time search-and-rescue operations, and to present the results of hovering manoeuvre for chosen model of the helicopter for fast SA in simulation form using the MATLAB/Simulink environment.

II. ADALINE

ADALINE was developed in [4]. It consists of a weight, a bias and a summation function.
Operation:

$$y_i = wx_i + b \tag{1}$$

Its adaptation is defined through a cost function (error metric) of the residual

$$e = d_i - (wx_i + b) \tag{2}$$

where d_i is the desired signal.
With the mean squared error metric

$$E = \frac{1}{2N}\sum_i^N e_i^2 \tag{3}$$

the bias and adapted weight become:

$$b = \frac{\sum_i x_i^2 \sum_i d_i - \sum_i x_i \sum_i x_i d_i}{N\sum_i (x_i - \bar{x})^2} \tag{4}$$

$$w = \frac{\sum_i (x_i - \bar{x})(d_i - \bar{d})}{\sum_i (x_i - \bar{x})^2} \tag{5}$$

III. Helicopter Model

Consider the nonlinear helicopter model [5] in terms of a state variable representation as follows:

$$\dot{x} = f(x) + g_1 u_1 + g_2 u_2 \tag{6}$$

$$y = Cx \tag{7}$$

where

$$f(x) = \begin{bmatrix} x_2 \\ a_0 + a_1 x_2 + a_2 x_2^2 + \left(a_3 + a_4 x_4 - \sqrt{a_5 + a_6 x_4}\right)x_3^2 \\ a_7 + a_8 x_3 + (a_9 \sin(x_4) + a_{10})x_3^2 \\ x_5 \\ a_{11} + a_{12} x_4 + a_{13} x_3^2 \sin x_4 + a_{14} x_5 \end{bmatrix},$$

$$g_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, g_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$u = [u_1 \;\; u_2]^T, x = [h \;\; \dot{h} \;\; \omega \;\; \theta \;\; \dot{\theta}]^T, y = [y_1 \;\; y_2]^T,$$

h = height above ground (m),
ω = rotational speed of the rotor blades (radn/s),
θ = collective pitch angle of rotor blades (radn),
u_1 = input to the throttle,
u_2 = input to collective servomechanisms.

The parameters a_0 through a_{14} are given by:
$a_0 = -17.67 m/s^2$, $a_1 = a_2 = -0.1 s^{-1}$, $a_3 = 5.31 \times 10^{-4}$,
$a_4 = 1.5364 \times 10^{-2}$, $a_5 = 2.82 \times 10^{-7}$, $a_6 = 1.632 \times 10^{-5}$,
$a_7 = -13.92 s^{-2}$, $a_8 = -0.7 s^{-1}$, $a_9 = a_{10} = -0.0028$,
$a_{11} = 434.88 s^{-2}$, $a_{12} = -800 s^{-2}$, $a_{13} = -0.1$ and
$a_{14} = -65 s^{-1}$.

IV. Simulation

Since the height is most critical for take-off and hovering manoeuvre, the control mechanization of the vertical trajectory profile will be demonstrated.
To illustrate the performance of the neural control design procedure for the helicopter model given by (6)-(7), we present two simulation examples: the first one addresses to control the take-off and hovering trajectory by one neural adaptive controller, while the second example presents the case of two neural adaptive controllers using to control the take-off and hovering trajectory.
The goal of the following simulations is twofold. First, we verify that these neural adaptive controllers are able to control the take-off and hovering trajectory. Second, we observed the effect of enhancing SA because the variety of such trajectory parameters as maximal height of flight and heights of hovering above points of hovering easily can be changed the possible take-off and hovering trajectory of helicopter.

A. Example 1

Consider the control of helicopter model (6)-(7) by one neural adaptive controller.
Initial conditions are chosen to be:

$$x_1(0) = 0m,\ x_2(0) = 0m/s,\ x_3(0) = 200 radn/s,$$

$$x_4(0) = 0.15 radn,\ x_5(0) = 0 radn/s.$$

Simulation results of the offered block scheme with one neural adaptive controller (see Fig. 1) are given in Fig. 6.

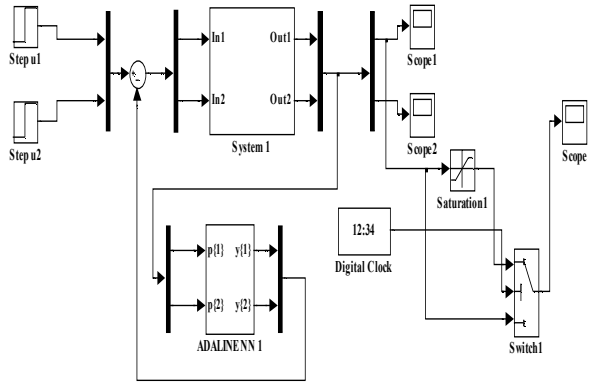


Fig. 1. Block diagram of control system with one neural adaptive controller.

A structure of the helicopter model (6)-(7) is illustrated in Fig. 2.
Detailed block diagrams of the adaptive NN from Fig. 1 are given in Figs. 3-5.

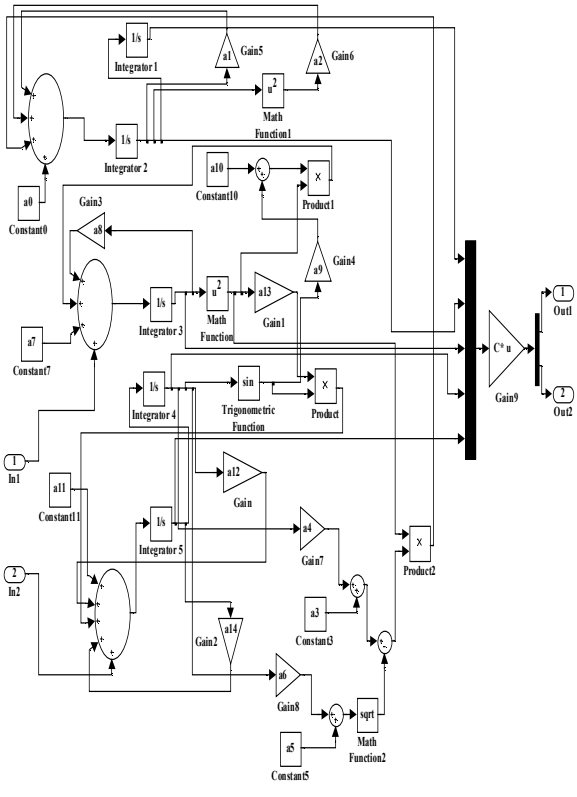


Fig. 2. The internal structure of the System 1 from Fig. 1.

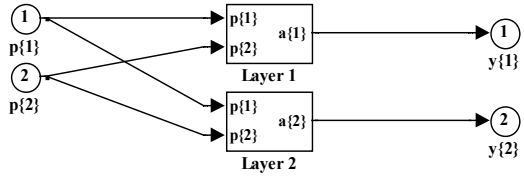


Fig. 3. Block diagram of ADALINE NN from Fig. 1.

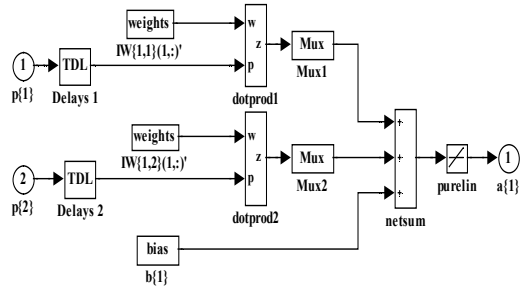


Fig. 4. Block diagram of Layer 1 from Fig. 3.

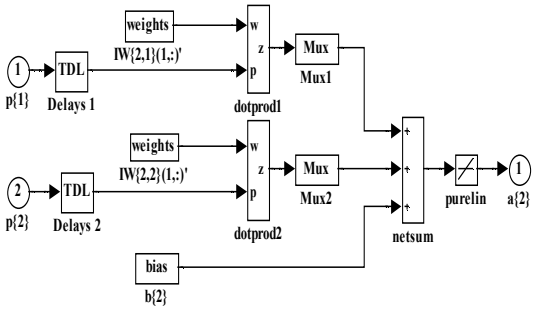


Fig. 5. Block diagram of Layer 2 from Fig. 3.

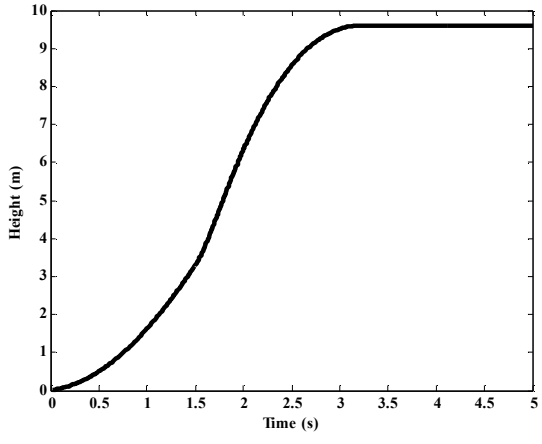


Fig. 6. Height trajectory of flight control using one neural adaptive controller.

Some disadvantages of this example are as follows.

- Possibility of lag in only one selected height position.
- Impossibility to consider a terrain restriction in a place of hovering can lead to accident and loss of this helicopter.

B. Example 2

Consider the control of helicopter model (6)-(7) for the case of take-off and hovering trajectory by hybrid system of two neural adaptive controllers.
Simulation results of the offered block scheme with two neural adaptive controllers (see Fig. 7) are shown in Fig. 8.

We simulated the block diagrams of Systems and ADALINE NNs with the same structural units given in Example 1 except that the full take-off and hovering trajectory was separated into initial and final phases with boundary point in the first lag position.

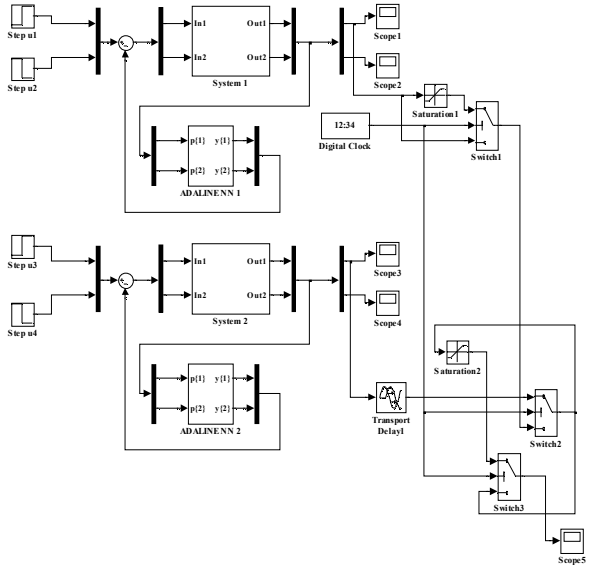


Fig. 7. Block diagram of control system with two neural adaptive controllers.

Some advantages of this example are as follows.

- Possibility to consider a terrain restriction in a place of a hovering in comparison with Example 1.
- Smoother trajectory of flight in comparison with Example 1 and smooth switching of regulation from one controller to another.
- Possibility of lag in two different selected height positions in comparison with Example 1.
- Using of two simple adaptive NNs to control the take-off and hovering trajectory of flight.

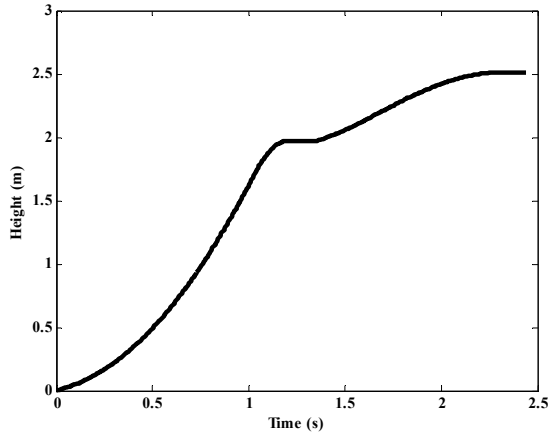


Fig. 8. Height trajectory of flight control using two neural adaptive controllers.

These results support the theoretical predictions well and demonstrate that this research technique would work in real-time flight conditions.

V. Conclusions

The need for highly reliable and stable hovering for helicopters and VTOL class autonomous vehicles has increased morbidly for critical situations in real-time search-and-rescue operations for fast SA.

For fast, stable and smooth hovering manoeuvre, we proposed a two stage flight strategy, which separates the flight process into initial and final phases. Two controllers on the base of ADALINE NNs as feedback gain controllers are designed for the initial phase and final phase respectively. The effectiveness of the proposed two stage flight strategy has been verified in field of flight simulation tests for chosen model of the helicopter using software package Simulink.

From the simulation studies of flight tests, the following can be observed:

- The block diagram of flight neural control is very useful for graphic representation of the flight trajectory.
- The received controllers for various flight phases are autonomous and completely shared in time.
- The trajectory tracking display forms give a researcher an immediate view of a helicopter motion with a range of such trajectory parameters as maximal height of flight and heights of hovering above places of hovering. This allows us to investigate the sensitivity of the control system, providing a medium for such development and evaluation and enhancing the researcher's understanding of hovering manoeuvres.
- The neural control using two stage flight strategies works more quickly and qualitatively than the neural control using one stage flight strategy.

From the applications viewpoint, we believe that this two stage flight strategy using flexible and effective neural control furnish a powerful approach for enhancing SA in applications to VTOL class autonomous vehicles.

References

[1] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems", Human Factors, vol. 37, no. 1, pp. 32-64, March 1995.

[2] J. Gorman, N. Cooke, and J. Winner, "Measuring team situation awareness in decentralized command and control environments", Ergonomics, vol. 49, nos. 12-13, pp. 1312–1325, October 2006.

[3] Interim Brigade Combat Team Newsletter. [Online]. Available: http://www.globalsecurity.org/military/library/report/call/call_01-18_toc.htm

[4] B. Widrow and S.D. Sterns, *Adaptive Signal Processing*, New York: Prentice-Hall, 1985.

[5] J. Kaloust, C. Ham, and Z. Qu, "Nonlinear autopilot control design for a 2-DOF helicopter model", IEEE Proc.-Control Theory Appl., vol. 144, no. 6, pp. 612-616, November 1997.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Set of Stabilizing PD Controllers For Multi-Input-Multi-Output Systems

Leena G[+] , K.B.Dutta[*] , and G Ray[*]

+ Electronics and Communication Engineering, C.I.T.M, Faridabad, Haryana 121001.
*Department of Electrical Engineering, I.I.T Kharagpur-721302, West Bengal, India.
Email: gray@ee.iitkgp.ernet.in

Abstract: For a multi-input multi-output (MIMO) system a set of stabilizing Proportional Derivative (PD) controllers was designed for each decoupled subsystem. The design approach is based on generalized result of the classical Hermite Biehler theorem. Stability analysis of the composite system with the designed set of decoupled controller parameter is studied via LMI framework. A genetic algorithm based search technique is adopted to select an optimal PD controller gain from a search space of PD stabilizing controllers. A MIMO system is considered to show the effectiveness of the design procedure.

Index Terms- Proportion-Derivative controller, stabilization, linear matrix inequality, interval matrix, Lyapunov function.

I. INTRODUCTION

This paper considers the problem of designing a set of stabilizing proportional-derivative (PD) controllers for a multivariable system. A major obstacle of designing the best controller has been the difficulty in characterizing the entire set of stabilizing controllers. Very few papers are published [1]-[3] where a set of stabilizing controllers is designed. An effective solution to this problem was obtained in [1]. This is accomplished by generalizing a classical stability result developed in the last century, the Hermite Biehler theorem. The present paper deals with the design of a set of decoupled PD controllers based on generalized Hermite Biehler theorem for each subsystem of MIMO linear systems. Subsequently, the significant results of Siljak et al. [4] based on Linear Matrix Inequalities (LMIs) framework is employed to show how the designed set of decoupled PD controllers is effective to quadratically stabilize the interconnected dynamic system.

This paper is organized as follows. Section II provides a brief review of a set of PD controllers based on Hermite Biehler theorem [1]. Section III considers a class of linear system that arises out of interconnection among the subsystems and subsequently, a set of stabilizing PD controller was designed for each decoupled subsystems neglecting the interaction terms.. The stability analysis of the composite linear system with the designed set of PD controllers via an LMI framework is also included in the same section. Simulation result presented in section-IV, illustrates the effectiveness of the proposed control strategies for a class of MIMO system. Finally, conclusions are given in Section V.

II. A SET OF STABILIZING PD CONTRO-LLER AND PROBLEM FORMULATION

Consider the feedback control system shown in Fig 1. Here r is the command signal, y is the output, $G(s) = N(s)/D(s)$ is the plant to be controlled; $N(s)$ and $D(s)$ are coprime polynomials. The controller $C(s)$ is to be designed which is given by

$$C(s) = K_p + sK_d. \tag{1}$$

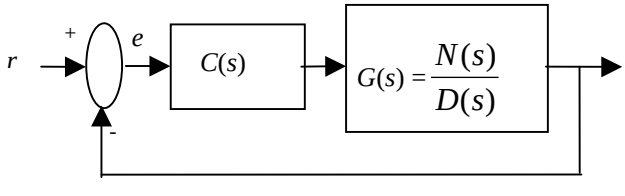


Fig. 1 Feedback control system

The closed-loop characteristic polynomial is

$$\delta(s, K_p, K_d) = D(s) + (K_p + sK_d)N(s), \tag{2}$$

The problem of stabilization using a PD controller is to determine the value of K_p, and K_d for which the closed-loop characteristic polynomial $\delta(s, K_p, K_d)$ is Hurwitz. A new polynomial is constructed by multiplying $\delta(s, K_p, K_d)$ with $N^*(s)(= N(-s)) = N_e(s^2) - sN_0(s^2)$, where N_e and N_o are the even and odd parts of $N(s)$ and examining the resulting polynomial, one can write $\sigma_r(\delta(s, K_p, K_d)N^*(s)) = l(\delta(s, K_p, K_d)N^*(s)) - r(\delta(s, K_p,$

$K_d)N^*(s)) = l(\delta(s,K_p,K_d)) - r(\delta(s,K_p,K_d)) - (l(N(s))$
$-r(N(s)))$. (3)

The polynomials $l(\delta(s,K_p,K_d))$ *and* $r(\delta(s,K_p,K_d))$ indicate the number of roots in $\mathbb{C}^-$ and $\mathbb{C}^+$ respectively. Now, the closed-loop characteristic polynomial $\delta(s,K_p,K_d)$, of degree n is Hurwitz if and only if $l(\delta(s,K_p,K_d)) = n$ and $r(\delta(s,K_p,K_d)) = 0$. Equation (3) can be restated as

$$\sigma_r(\delta(s,K_p,K_d)N^*(s)) = n - (l(N(s)) - r(N(s))). \quad (4)$$

We have to determine those values of K_p, K_d for which (4) holds and $\delta(s,K_p,K_d)N^*(s)$ has the following expression:

$$\begin{aligned}\delta(s,K_p,K_d)N^*(s) &= [-s^2D_o(s^2)N_o(s^2) + D_e(s^2)N_e(s^2)\\ &+K_p(N_e(s^2)N_e(s^2) - s^2N_o(s^2)N_o(s^2))]\\ &+s[D_o(s^2)N_e(s^2) - N_o(s^2)D_e(s^2)\\ &+K_d(N_e(s^2)N_e(s^2) - s^2N_o(s^2)N_o(s^2))]. \quad (5)\end{aligned}$$

The new polynomial with s = $j\omega$ is given by

$$\delta(j\omega,K_p,K_d)N^*(j\omega) = p(\omega,K_p) + jq(\omega,K_d), \quad (6)$$

where, $p(\omega,K_p) = p_1(\omega) + K_p\, p_2(\omega)$,

$$q(\omega,K_p) = q_1(\omega) + K_d\, q_2(\omega),$$

$$\begin{aligned}p_1(\omega) &= [D_e(-\omega^2)N_e(-\omega^2) + \omega^2D_o(-\omega^2)N_o(-\omega^2)],\\ p_2(\omega) &= [N_e(-\omega^2)N_e(-\omega^2) + \omega^2N_o(-\omega^2)N_o(-\omega^2)],\\ q_1(\omega) &= \omega[D_o(-\omega^2)N_e(-\omega^2) - D_e(-\omega^2)N_o(-\omega^2)],\\ q_2(\omega) &= \omega[N_e(-\omega^2)N_e(-\omega^2) + \omega^2N_o(-\omega^2)N_o(-\omega^2)].\end{aligned}$$

The new polynomial described by (6) is normalized in the following manner.

$$p_f(\omega,K_p) = \frac{p(\omega,K_p)}{(1+\omega^2)^{(m+n)/2}},$$

$$q_f(\omega,K_d) = \frac{q(\omega,K_d)}{(1+\omega^2)^{(m+n)/2}}.$$

It can be seen from (6), that K_d appears in the odd part whereas K_p appears in the even part. Furthermore for every fixed K_p, the zeros of $p(\omega,K_p)$ will not depend on K_d. The range of K_d for a fixed K_p can be obtained by following the procedure as discussed above. Our main aim is to develop a set of stabilizing PD controllers for a class of MIMO system neglecting interaction terms and subsequently the stability analysis of the composite system is investigate with the designed PD controllers based on LMI formulation [5].

III. STABILITY ANALYSIS OF INTERCONNECTED SYSTEM WITH DESIGNED SET OF PD CONTROLLERS

Let us consider a class of MIMO system, which arises out of *n* interconnected subsystems, is described by

$$\begin{aligned}\dot{x}_i &= A_{ii}x_i + B_iu_i + h_i(t,x), \qquad i = 1,2,\ldots,N\\ y_i &= C_{ni}x_i,\end{aligned} \quad (7)$$

where x_i is the state vector, u_i is the input vector, $h_i(t,x) = \sum_{j=i,\, j\neq i}^{N} A_{ij}x_j$ are the state interconnection terms. The only information about the interaction term is that it satisfies the quadratic constraint. The transfer function of the i^{th} decoupled system (without interaction terms, i.e. $h_i(t,x) = 0$) is described by

$$C_{ni}(sI - A_{ii})^{-1}B_i, \qquad i = 1,2,\ldots,N. \quad (8)$$

and a set of stabilizing PD controllers for each decoupled system is designed using the method [1] briefly described in Section II.

A. Stability Analysis

The linear system (7) can be rewritten as

$$\begin{aligned}\dot{x}_i &= A_{ii}x_i + B_iu_i + w_i(t,x)\\ y_i &= C_{ni}x_i \qquad i = 1,2,\ldots,N,\end{aligned} \quad (9)$$

where $w_i = h_i$. It is assumed that the pair (A_{ii}, B_i) is stabilizable and the i^{th} subsystem interaction term w_i satisfies the quadratic constraints

$$w_i^T(t,x)w_i(t,x) \le \alpha_i^2 x^T W_i^T W_i x, \quad \text{for, } i = 1,2,\ldots,N, \quad (10)$$

where $\alpha_i > 0$ are bounding parameters and W_i are constant matrices of appropriate dimensions. Fig. 2 shows that the i^{th} subsystem with the PD controller. The input u_i to the i^{th} subsystem is

$$u_i = K_{pi}e_i(t) + K_{di}\dot{e}_i(t) \quad (11)$$

where $e_i = ref_i - y_i$ is the error of i^{th} subsystem, K_{pi}, and K_{di} are respectively, the proportional and derivative (PD) controller gains of the i^{th} subsystem.

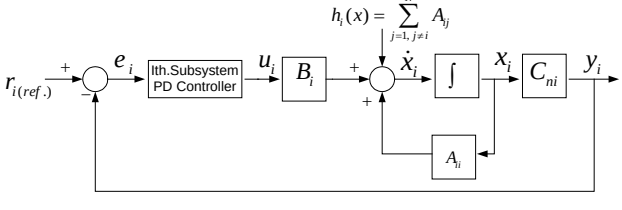


Fig. 2 i^{th} subsystem with the PD controller

Assume, $ref_i = 0$, then $u_i = K_{pi}(-y_i) + K_{di}(-\dot{y}_i)$, (12)

To study the stability of the interconnected system substitute for u_i in equation (9) to get

$$E_{ni}\dot{x}_i = A_{ni}x_i + w_i(x), \quad (13)$$

where, $E_{ni} = I + B_i K_{di} C_{ni}$, $A_{ni} = A_{ii} - B_i K_{pi} C_{ni}$, for $i = 1, 2, ...N$. In a compact form the equation (13) can be rewritten as

$$E_n \dot{x} = A_n x + w(x), \tag{14}$$

where $A_n = dia\{A_{n1}, A_{n2}, ...A_{nN}\}$ and $E_n = dia\{E_{n1}, E_{n2}, ...E_{nN}\}$ are interval matrices of appropriate dimensions. In the compact notation, $w = (w_1^T, w_2^T, ...w_N^T)^T$ and $x = (x_1^T, x_2^T, ...x_N^T)^T$ are the interconnection vector and the state vector respectively. The stability analysis of the composite system while employing a set of decoupled PD controllers is discussed below.

Theorem: The interconnected system (9) is robustly stabilizable with degree α_i by the control law (11) if for matrices P_1, P_2, P_3 of compatible dimensions, and $\gamma_1, \gamma_2, \ldots, \gamma_N > 0$ there exists a feasible solution for the following LMI problem for all the corner matrices of A_n and E_n.

Minimize $\sum_{i=1}^{N} \gamma_i$,

subject to $P_1 > 0$, and

$$\begin{bmatrix} A_n^{r_1 T} P_2 + P_2^T A_n^{r_1} & A_n^{r_1 T} P_3 + P_1 - P_2^T E_n^{r_2} & P_2^T & W_1^T & \cdots & W_N^T \\ P_3^T A_n^{r_1} + P_1 - E_n^{r_2 T} P_2 & -E_n^{r_2 T} P_3 - P_3^T E_n^{r_2} & P_3^T & 0 & \cdots & 0 \\ P_2 & P_3 & -I & 0 & \cdots & 0 \\ W_1 & 0 & 0 & -\gamma_1 I & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ W_N & 0 & 0 & 0 & \cdots & -\gamma_N I \end{bmatrix} < 0, \tag{15}$$

where $r_1, r_2 = 1, 2, \ldots, 2^{n^2}$, $\gamma_i = 1/\alpha_i^2$, n is the size of the matrices A_n and E_n, for $i = 1, 2, \ldots, N$.

Proof: The constraint (10) is equivalent to the quadratic inequality [4]

$$\begin{bmatrix} x^T & w^T(x) \end{bmatrix} \begin{bmatrix} -\sum_{i=1}^{N} \alpha_i^2 W_i^T W_i & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x \\ w(x) \end{bmatrix} \le 0. \tag{16}$$

For the descriptor system (14), we introduce an augmented system to get the following equation.

$$F\,\dot{z} = \begin{bmatrix} 0 & I \\ A_n & -E_n \end{bmatrix} z + \overline{w}(x(t)). \tag{17}$$

where $F = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$, $\overline{A} = \begin{bmatrix} 0 & I \\ A_n & -E_n \end{bmatrix}$, $z = \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$ and $\overline{w}(x) = \begin{bmatrix} 0 \\ w(x) \end{bmatrix}$.

Let us choose a Lyapunov function candidate for the descriptor system (17) as discussed in [6]

$$V = z^T F P z, \tag{18}$$

where $P = \begin{bmatrix} P_1 & 0 \\ P_2 & P_3 \end{bmatrix}$ is non singular with $P_1 = P_1^T > 0$, and $FP = (FP)^T$ due to special structures of F and P. We compute

$$\dot{V} = z^T (\overline{A}^T P + P^T \overline{A}) z + \overline{w}^T(x) P z + z^T P^T \overline{w}(x).$$

The descriptor system (17) is stable, provided the following conditions hold.

$$P_1 > 0, \qquad z^T (\overline{A}^T P + P^T \overline{A}) z + \overline{w}^T(x) P z + z^T P^T \overline{w}(x) < 0. \tag{19}$$

Equation (19) is equivalently can be written as

$$\begin{aligned} &P_1 > 0, \\ &x^T (A_n^T P_2 + P_2^T A_n) x + \dot{x}^T (-E_n^T P_3 - P_3^T E_n) \dot{x} + \\ &\dot{x}^T (P_1 - E_n^T P_2 + P_3^T A_n) x + x^T (A_n^T P_3 + P_1 - P_2^T E_n) \dot{x} + \\ &w^T(x) P_2 x + w^T(x) P_3 \dot{x} + x^T P_2^T w(x) + \dot{x}^T P_3^T w(x) < 0. \end{aligned} \tag{20}$$

These inequalities can be rewritten as,

$$P_1 > 0,$$

$$\begin{bmatrix} x^T & \dot{x}^T & w^T(x) \end{bmatrix} \begin{bmatrix} A_n^T P_2 + P_2^T A_n & A_n^T P_3 + P_1 - P_2^T E_n & P_2^T \\ P_3^T A_n + P_1 - E_n^T P_2 & -E_n^T P_3 - P_3^T E_n & P_3^T \\ P_2 & P_3 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ w(x) \end{bmatrix} < 0. \tag{21}$$

By using S-procedure [7], it is possible to combine quadratic inequalities (16) and (21) into one single LMI such that

$$\begin{bmatrix} A_n^T P_2 + P_2^T A_n + \beta \sum_{i=1}^{N} \alpha_i^2 W_i^T W_i & A_n^T P_3 + P_1 - P_2^T E_n & P_2^T \\ P_3^T A_n + P_1 - E_n^T P_2 & -E_n^T P_3 - P_3^T E_n & P_3^T \\ P_2 & P_3 & -\beta I \end{bmatrix} < 0, \tag{22}$$

where $P_1 > 0$ and a number $\beta > 0$. By repeatedly applying the Schur-complement formula to equation (22) and $\beta = 1$, the equation (22) can be rewritten as

$$\begin{bmatrix} A_n^T P_2 + P_2^T A_n & A_n^T P_3 + P_1 - P_2^T E_n & P_2^T & W_1^T & \cdots & W_N^T \\ P_3^T A_n + P_1 - E_n^T P_2 & -E_n^T P_3 - P_3^T E_n & P_3^T & 0 & \cdots & 0 \\ P_2 & P_3 & -I & 0 & \cdots & 0 \\ W_1 & 0 & 0 & -\gamma_1 I & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ W_N & 0 & 0 & 0 & \cdots & -\gamma_N I \end{bmatrix} < 0, \tag{23}$$

where $\gamma_i = 1/\alpha_i^2$.

The matrices A_n and E_n of equation (23) are interval matrices (obtained from equation (13)). A sufficient condition for the stability robustness of interval matrices, i.e., matrices having the elements varying within given bounds, requires that the Lyapunov equation be negative definite when evaluated at the so-called corner matrices [8]-[9]. The corner matrices of an $n \times n$ interval matrix A is defined as $A^r = \{a_{ij}\}^r$, $r = 1, 2, \ldots, 2^{n^2}$ with

$a_{ij}^{r} = alij\ or\ auij$, $i, j = 1,2,...n$, where *alij and auij* are minimum and maximum values respectively of ij^{th} element of interval matrix. Hence equation (23) should be satisfied for all the corner matrices of A_n and E_n to ensure that the composite system (17) to be asymptotically stable. The matrix W_i is computed so that constraint (10) is satisfied and the bounding parameter α_i is to be maximized. Hence equation (23) can be reformulated as an LMI optimization problem as stated in equation (15). In other words, system (7) is robustly stabilized by the set of decoupled PD controllers provided the LMI problem (15) has a feasible solution for all corner matrices. This completes the proof of the theorem.

IV. Simulation Results

A MIMO system is considered to illustrate the results of this paper. Each subsystem is described by the equation (7), where N = 2. The following are the numerical values for the two interconnected systems and each subsystem (A_{ii}, $i = 1,2$) is considered as a forth order system.

Subsystem-1: $$A_{11} = \begin{bmatrix} -0.922 & 1 & -0.266 & -0.009 \\ -2.75 & -2.78 & -1.36 & -0.37 \\ 0 & 0 & 0 & 1 \\ -4.95 & 0 & -55.5 & -0.39 \end{bmatrix},$$

$$B_1 = \begin{bmatrix} 0 & 36.1 & 0 & 0 \end{bmatrix}^T,\ C_{n1} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$$

Subsystem-2: $$A_{22} = \begin{bmatrix} -0.21 & 1 & -1.6 & -0.005 \\ -1.9 & -1.8 & 9.3 & -0.12 \\ 0 & 0 & 0 & 1 \\ -3.1 & 0 & -56 & 0.032 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 0 & 78.9 & 0 & 0 \end{bmatrix}^T, C_{n2} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$$

State interconnection Matrices:

$$A_{12} = \begin{bmatrix} 0.024 & 0 & -0.087 & -0.002 \\ -0.158 & 0 & 1.11 & -0.011 \\ 0 & 0 & 0 & 0 \\ 0.222 & 0 & 8.17 & 0.004 \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} 0.021 & 0 & 0.121 & 0.003 \\ -1.1 & 0 & -1.62 & -0.015 \\ 0 & 0 & 0 & 0 \\ -2.43 & 0 & 1.37 & -0.034 \end{bmatrix}, \quad (24)$$

A. A Set of PD Controllers for MIMO Linear System

Following the method discussed in Section-II, a set of PD controllers was designed for each of the two subsystems from its transfer function as $C_{ni}(sI - A_{ii})^{-1}B_i$, i = 1, 2

The set of stabilizing PD controllers obtained for subsystem 1 and subsystem 2 without considering the interaction terms are presented in Fig. 3 and Fig. 4 respectively. A rectangular area inside the shaded regions of Figs. 3 and 4 is considered as

$$K_{p1} \in [-2.01 \quad 0.91],\ K_{d1} \in [0.08 \quad 0.81],$$

$$K_{p2} \in [-1.59 \quad 0.205],\ K_{d2} \in [0.09 \quad 0.31]. \quad (25)$$

From the specified range of stabilizing controllers given in (25), an optimal controller parameter (K_{pi}^{*}, K_{di}^{*}) for each subsystem ($i = 1,2$) was obtained maximizing the fitness function J_f, where J_f is defined as

$$J_f = \frac{1}{1+J}, \quad J = \int_0^t \sum_{i=1}^{2} e_i^{\,2} dt, \quad (26)$$

and e_1, e_2 are output errors of the composite system. This problem is solved using genetic operations like arithmetic crossover, uniform mutation and ranking selection [10]. The population size of 50 was taken and GA was run for 25 generations. The optimal gains are obtained by maximizing the fitness function (26) while the first subsystem is subjected to 0.05-unit step disturbance with zero initial states and subsequently for each subsystem optimal controller gains are obtained as:

$$K_{p1}^{*} = -0.9537,\ K_{d1}^{*} = 0.1263,\ K_{p2}^{*} = -0.9564,\ K_{d2}^{*} = 0.0935. \quad (27)$$

The output and state responses of the interconnected system are obtained with the above optimal controller parameters and they are presented in Figs. 5-7. The simulation results shown in Figs. 5-7 reveal that the interconnected system is stable even though the PD controllers were designed based on the information of each decoupled subsystem dynamics.

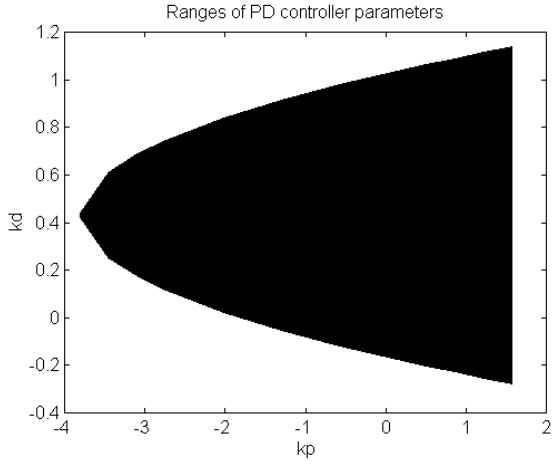


Fig. 3 The stabilizing set of (K_p, K_d) values

for the decoupled subsystem 1.

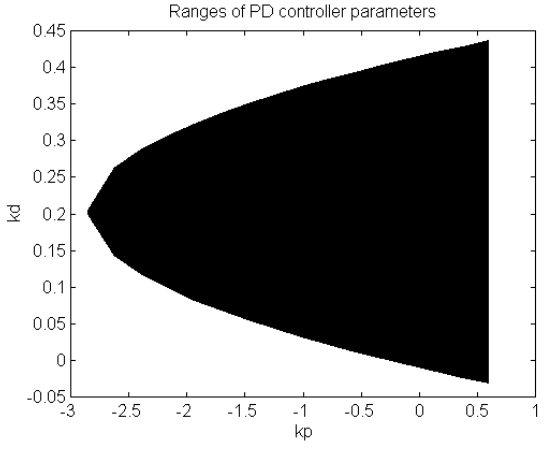


Fig. 4 The stabilizing set of (K_p, K_d) values for the decoupled subsystem 2.

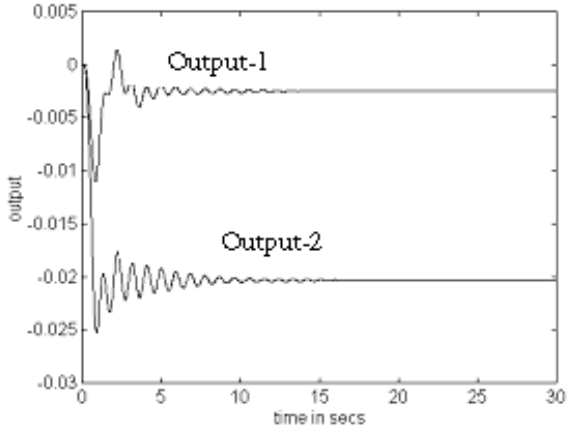


Fig. 5 Output Responses of subsystems 1 and 2.

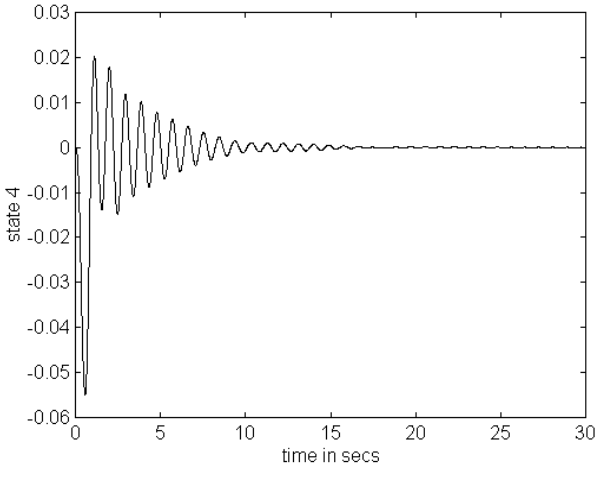


Fig. 6 State response of subsystem-1

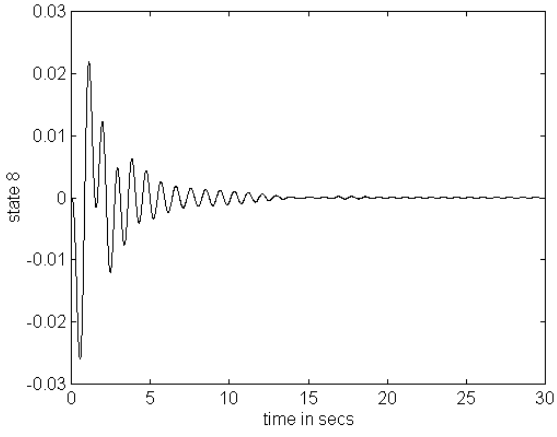


Fig. 7 State response of subsystem-2

B. Stability Analysis of MIMO Linear System With Designed PD Controller: Numerical Results

Theoretically the stability was studied by solving the LMI optimization problem (15) for all the corner matrices of A_n and E_n. For the designed ranges of controller gains (25), the matrices A_n and E_n are found out using (14) as

$$A_n = diag\{A_{n1}, A_{n2}\}, \quad E_n = diag\{E_{n1}, E_{n2}\}, \qquad (28)$$

where

$$A_{n1} = \begin{bmatrix} -0.922 & 1 & -0.266 & -0.009 \\ -2.75 & -2.78 & [-34.2 & 71.2] & -0.37 \\ 0 & 0 & 0 & 1 \\ -4.95 & 0 & -55.5 & -0.39 \end{bmatrix},$$

$$E_{n1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & [2.89 & 29.2] & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A_{n2} = \begin{bmatrix} -0.21 & 1 & -1.6 & -0.005 \\ -1.9 & -1.8 & [-6.87 & 134.8] & -0.012 \\ 0 & 0 & 0 & 1 \\ -3.1 & 0 & -56 & 0.032 \end{bmatrix},$$

$$E_{n2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & [7.1 & 24.6] & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The interconnection terms are $h_1 = A_{12}x_2$ and $h_2 = A_{21}x_1$ from which W_1 and W_2 are obtained such that equation (10) is satisfied as

$$W_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.024 & 0 & -0.087 & -0.002 \\ 0 & 0 & 0 & 0 & -0.158 & 0 & 1.11 & -0.011 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.222 & 0 & 8.17 & 0.004 \end{bmatrix},$$

$$W_2 = \begin{bmatrix} 0.021 & 0 & 0.121 & 0.003 & 0 & 0 & 0 & 0 \\ -1.1 & 0 & -1.62 & -0.015 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2.43 & 0 & 1.37 & -0.034 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (29)$$

Two elements in each of A_n and E_n are of interval form, i.e., $r_1 = 1, 2^2$ and $r_2 = 1, 2^2$, thus eight corner matrices are obtained for A_n and E_n. Table 1 shows the corner matrices of A_n and E_n. Each combination of corner matrices of A_n and E_n are taken, thus sixteen combinations are obtained for which LMI problem (15) is solved using LMI control toolbox [11] with the numerical values W_1, W_2 as obtained in (29). Table 2 shows the α_1, α_2 values obtained by solving the LMI problem (15) with the designed ranges given by (25). There exist a feasible solution for all the sixteen combinations of corner matrices of A_n and E_n and hence it is concluded that the set of PD controllers designed for the linear decoupled subsystems (8) stabilizes the interconnected system (7).

Table 1 Corner matrices of A_n and E_n

$A_n^1 = \{A_{n1}\underline{(2,3)}, A_{n2}\underline{(2,3)}\}$	$E^1 = \{E_1\underline{(2,3)}, E_2\underline{(2,3)}\}$
$A_n^2 = \{A_{n1}\underline{(2,3)}, A_{n2}\overline{(2,3)}\}$	$E^2 = \{E_1\underline{(2,3)}, E_2\overline{(2,3)}\}$
$A_n^3 = \{A_{n1}\overline{(2,3)}, A_{n2}\underline{(2,3)}\}$	$E^3 = \{E_1\overline{(2,3)}, E_2\underline{(2,3)}\}$
$A_n^4 = \{A_{n1}\overline{(2,3)}, A_{n2}\underline{(2,3)}\}$	$E^4 = \{E_1\overline{(2,3)}, E_2\overline{(2,3)}\}$

where $A_n\underline{(2,3)}$ and $A_{n1}\overline{(2,3)}$ denote the lower and upper limits of $(2,3)^{th}$ element of matrix A_n.

Table 2 α_1, α_2 values obtained solving LMI problem with PD controller

	E^1		E^2		E^3		E^4	
	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
A_n^1	0.06	0.27	0.17	0.27	0.06	0.13	0.12	0.13
A_n^2	0.02	0.23	0.10	0.27	0.02	0.13	0.10	0.13
A_n^3	0.06	0.03	0.17	0.03	0.06	0.01	0.16	0.01
A_n^4	0.02	0.03	0.10	0.03	0.02	0.01	0.10	0.01

V. Concluding Remarks

A set of stabilizing decoupled PD controllers for MIMO system (neglecting state interaction terms) is designed based on generalized Hermite Biehler theorem. It has been proved that the designed set of PD controllers stabilizes the composite system provided the state interaction term satisfies certain quadratic constraints and subsequently, the feasible solution of bounding parameters (α) for the interaction terms exits. This, in turn, requires satisfying a set of sufficient condition that is framed in LMI formulation (15). The genetic algorithm based optimization technique is employed to get an optimal controller gain from the range of stabilizing controller parameters. Simulation results illustrate the applicability of the proposed set of decoupled PD controller for a class of interconnected system.

References

[1] G. J. Silva, A. Datta, and S. P. Bhattacharyya, "PID Controllers for Time Delay Systems," Birkhauser, Control Engineering Series, 2005.

[2] M.T. Ho and C.Y. Lin, "PID controller design for robust performance," *IEEE Trans. on Automat. Control*, vol. 48, no. 8, pp. 1404-1409, 2003.

[3] M. T. Ho and S. T Huang, "On the synthesis of robust PID controllers for plants with structured and unstructured uncertainty," *Int. J. of Robust and Nonlinear Control*, vol. 15, no. 6, pp. 269-85, 2005.

[4] D. D. Siljak and D. M. Stipanovic, "Robust stabilization of nonlinear systems: The LMI approach," *Math. Prob. Engg.*, vol. 6, pp. 461-493, 2000.

[5] S. Boyd, E. Feron, L. E. Ghaoui, and V. Balakrishnan "*Linear Matrix Inequalities in System and Control Theory,*" Siam Philadelphia 1994.

[6] Y. Y. Cao, and Z. Lin., "A descriptor system approach to robust stability analysis and controller synthesis," *IEEE Trans. on Automat. Control*, vol. 49, no. 11, pp. 2081-2084, 2004.

[7] V. A. Yakubovich, "'The S-procedure in nonlinear control theory," English translation in *Vestnik Leningrad Univ. Math.*, vol. 4, pp. 73-93, 1977.

[8] M. Mansour, ''Sufficient condition for the asymptotic stability of interval matrices," *Int. J. of Control*, vol. 47, no. 6, pp. 1973-1974, 1988.

[9] F. Garofalo, G. Celentano, and L. Glielmo, "Stability robustness of interval matrices via Lyapunov quadratic forms," *IEEE Trans. on Automat. Control*, vol. 38, no. 2, pp. 281-284, 1993.

[10] D. E. Goldberg, "*Genetic Algorithms in Search, Optimization, and Machine Learning*," Addison Wesley, 1999.

[11] P. Gahinet, A. Nemirovski, A. J. Laub, and M. Chilali, "*LMI Control Toolbox for use with Matlab*," The Math works Inc, 1995.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Novel Approach for Complete Identification of Dynamic Fractional Order Systems Using Stochastic Optimization Algorithms and Fractional Calculus

Deepyaman Maiti, Mithun Chakraborty, and Amit Konar

Department of Electronics and Telecommunication Engineering, Jadavpur University
Kolkata - 700032, West Bengal, India
E-mail: deepyamanmaiti@gmail.com, mithun.chakra108@gmail.com, and konaramit@yahoo.co.in

Abstract - This contribution deals with identification of fractional-order dynamical systems. System identification, which refers to estimation of process parameters, is a necessity in control theory. Real processes are usually of fractional order as opposed to the ideal integral order models. A simple and elegant scheme of estimating the parameters for such a fractional order process is proposed. This method employs fractional calculus theory to find equations relating the parameters that are to be estimated, and then estimates the process parameters after solving the simultaneous equations. The said simultaneous equations are generated and updated using particle swarm optimization (PSO) technique, the fitness function being the sum of squared deviations from the actual set of observations. The data used for the calculations are intentionally corrupted to simulate real-life conditions. Results show that the proposed scheme offers a very high degree of accuracy even for erroneous data.

I. Introduction

Proper estimation of the parameters of a real process, fractional or otherwise, is a challenge to be encountered in the context of system identification [1] - [3]. Accurate knowledge of the parameters of a system is often the first step in designing controllers. Many statistical and geometric methods such as least square and regression models are widely used for real-time parameter estimation. The problem of parameter estimation becomes more difficult for a fractional order system compared to an integral order one. The real world objects or processes that we want to estimate are generally of fractional order [4]. A typical example of a non-integer (fractional) order system is the voltage-current relation of a semi-infinite lossy RC line or diffusion of heat into a semi-infinite solid, where heat flow q(t) is equal to the half-derivative of temperature T(t): $\frac{d^{0.5}T(t)}{dt^{0.5}} = q(t)$.

So far, however, the usual practice when dealing with a fractional order process has been to use an integer order approximation. In general, this approximation can cause significant differences between a real system and its mathematical model. Disregarding the fractional order of the system was caused mainly by the non-existence of simple mathematical tools for the description of such systems. Since major advances have been made in this area recently, it is possible to consider also the real order of the dynamical systems. Such models are more adequate for the description of dynamical systems with distributed parameters than integer-order models with concentrated parameters. With regard to this, in the task of identification, it is necessary to consider also the fractional-order of the dynamical system.

Most classical identification methods cannot cope with fractional order transfer functions. Yet, this challenge must be overcome if we want to design a proper adaptive or self-tuning fractional order controller. Need for design of adaptive controllers gives an impetus to finding accurate schemes for system identification.

Computation of transfer characteristics of the fractional order dynamic systems has been the subject of several publications [5] – [8], e.g. by numerical methods [7], as well as by analytical methods [8]. In this paper we propose a method for parameter identification of a fractional order system for a chosen structure of the model using fractional calculus theory to obtain simultaneous equations relating the unknown parameters and then solving these equations to obtain accurate estimates. This method enables us to work with the actual fractional order process rather than an integer order approximation. Using it in a system with known parameters will do the verification of the correctness of the identification.

We first consider that the fractional powers are constant and display the accuracy of the proposed method both when random corruptions are absent and present. Then we remove this limitation and propose a scheme using the particle swarm optimization (PSO) algorithm by which a fractional order system can be completely identified with a high degree of accuracy even in presence of significant quantities of error in the readings. PSO, a stochastic optimization strategy from the family of evolutionary computation, is a biologically-inspired technique originally proposed by Kennedy and Eberhart in [12]. The PSO algorithm will generate process models guided by the fractional differintegral definitions and optimize the models after comparing the simulated outputs from such models with the set of outputs obtained from the actual fractional order system (for the same input).

Section II discusses the basics of fractional calculus and PSO.

II. Fractional Calculus and PSO Algorithm

Sub-section A gives the necessary theory and formulae of fractional differintegral. The relevant facts about the PSO algorithm are mentioned in sub-section B.

A. Theory of Fractional Calculus

The fractional calculus is a generalization of integration and derivation to non-integer order operators. At first, we generalize the differential and integral operators into one fundamental operator ${}_aD_t^\alpha$ where: ${}_aD_t^\alpha = \frac{d^\alpha}{dt^\alpha}$ for $\Re(\alpha) > 0$; 1 for $\Re(\alpha) = 0$; $\int_a^t (d\tau)^{-\alpha}$ for $\Re(\alpha) > 0$.

The two definitions used for fractional differintegral are the Riemann-Liouville definition [9], [10] and the Grunwald-Letnikov definition [9], [11].

The Grunwald-Letnikov definition is

$${}_aD_t^\alpha f(t) = \lim_{h\to 0} \frac{1}{h^\alpha} \sum_{j=0}^{\left[\frac{t-a}{h}\right]} (-1)^j \binom{\alpha}{j} f(t-jh) \quad (1)$$

($[y]$ means the greatest integer not exceeding y).

Derived from the Grunwald-Letnikov definition, the numerical calculation formula of fractional derivative can be achieved as

$${}_{t-L}D_t^\alpha f(t) \approx h^{-\alpha} \sum_{j=0}^{[L/T]} b_j f(t-jh) \quad (2)$$

L is the length of memory. T, the sampling time always replaces the time increment h during approximation. The weighting coefficients b_j can be calculated recursively by:

$$b_0 = 1, b_j = \left(1 - \frac{1+\alpha}{j}\right) b_{j-1}, (j \ge 1). \quad (3)$$

B. Particle Swarm Optimization (PSO)

The PSO algorithm [12] - [14] attempts to mimic the natural process of group communication of individual knowledge, which occurs when a social swarm elements flock, migrate, forage, etc. in order to achieve some optimum property such as configuration or location.

The 'swarm' is initialized with a population of random solutions. Each particle in the swarm is a different possible set of the unknown parameters to be optimized. Representing a point in the solution space, each particle adjusts its flying toward a potential area according to its own flying experience and shares social information among particles. The goal is to efficiently search the solution space by swarming the particles toward the best fitting solution encountered in previous iterations with the intent of encountering better solutions through the course of the process and eventually converging on a single minimum error solution.

Let the swarm consist of N particles moving around in a D-dimensional search space. Each particle is initialized with a random position and a random velocity. Each particle modifies its flying based on its own and companions' experience at every iteration. The i^{th} particle is denoted by X_i, where $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$. Its best previous solution (pbest) is represented as $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$. Current velocity (position changing rate) is described by V_i, where $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$. Finally, the best solution achieved so far by the whole swarm (gbest) is represented as $P_g = (p_{g1}, p_{g2}, \ldots\ldots, p_{gD})$.

At each time step, each particle moves towards pbest and gbest locations. The fitness function evaluates the performance of particles to determine whether the best fitting solution is achieved. The particles are manipulated according to the following equations:

$$v_{id}(t+1) = \omega v_{id}(t) + c_1.\varphi_1.(p_{id}(t) - x_{id}(t)) + c_2.\varphi_2.(p_{gd}(t) - x_{id}(t))$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1).$$

(The equations are presented for the d^{th} dimension of the position and velocity of the i^{th} particle.)

Here, c_1 and c_2 are two positive constants, called cognitive learning rate and social learning rate respectively, φ_1 and φ_2 are two random functions in the range [0,1], ω is the time-decreasing inertia factor designed by Eberhart and Shi [9]. The inertia factor balances the global wide-range exploitation and the nearby exploration abilities of the swarm.

III. Process of Identification for Constant Fractional Powers

We have considered a fractional order process whose transfer function is of the form $\frac{1}{a_1 s^\alpha + a_2 s^\beta + a_3}$. The orders of fractionality α and β are known and the coefficients a_1, a_2 and a_3 are to be estimated. One important advantage of the proposed scheme is that we do not require to know the ranges of variation of a_1, a_2 and a_3. It should be noted that without loss of generality, we may presume the dc gain to be unity so that the dc gain and its possible fluctuations are included in the coefficients a_1, a_2 and a_3. If C(s) is the output and R(s) the input,

$$\frac{C(s)}{R(s)} = \frac{1}{a_1 s^\alpha + a_2 s^\beta + a_3},$$

$$\Rightarrow R(s) = a_1 s^\alpha C(s) + a_2 s^\beta C(s) + a_3 C(s)$$

In time domain,

$$r(t) = a_1 D^\alpha c(t) + a_2 D^\beta c(t) + a_3 c(t) \quad (4)$$

$$\Rightarrow r(t) \approx a_1 T^{-\alpha} \sum_{j=0}^{[L/T]} b_j c(t-jT) + a_2 T^{-\beta} \sum_{j=0}^{[L/T]} b_j c(t-jT) + a_3 c(t)$$

The proposed scheme requires sampled input at time instant t and sampled outputs at time instants t, t – T, t – 2T, t – 3T, ... Sampled outputs are required for a time length L previous to t, T being the sampling time. Calculation of fractional derivatives and integrals requires the past history of the process to be remembered. So more the value of L, the better.

Thus the values of $D^\alpha c(t)$ and $D^\beta c(t)$ can be calculated using (1), (2) and (3) so that (4) reduces to the form $a_1 p + a_2 q + a_3 r = s$, where p, q, r, s are constants whose values have been determined.

Let us assume that we have a set of sampled outputs c(t) from the system for unit step test signal.
That is, we have

$$u(t) = a_1 D^{\alpha} c(t) + a_2 D^{\beta} c(t) + a_3 c(t). \quad (5)$$

Now there are three unknown parameters, namely a_1, a_2 and a_3. So we need three simultaneous equations to solve for them. One equation is (5). We will integrate both sides of (5) to obtain

$$\int u(t)dt = \int [a_1 D^{\alpha} c(t) + a_2 D^{\beta} c(t) + a_3 c(t)]dt .$$

This gives us

$$r(t) = a_1 D^{\alpha-1} c(t) + a_2 D^{\beta-1} c(t) + a_3 D^{-1} c(t) \quad (6)$$

Here r(t) signifies unit ramp input and c(t) is the output due to unit step input. Thus we have derived a second equation relating a_1, a_2 and a_3.
The third equation will be obtained by integrating both sides of (6). This gives us

$$p(t) = a_1 D^{\alpha-2} c(t) + a_2 D^{\beta-2} c(t) + a_3 D^{-2} c(t) \quad (7)$$

Here p(t) signifies parabolic input and c(t) is the output due to unit step input.
It can be seen that (5), (6), (7) are distinct equations in a_1, a_2 and a_3. So we can solve them simultaneously to identify the three unknown parameters a_1, a_2 and a_3.
As we have displayed elsewhere, direct application of the above scheme gives very satisfactory results when the readings c(t) are accurate. If we now add an error component e(t) to c(t) to have a distorted output waveform $c(t) \equiv c(t) + e(t)$ from which we want to make our identification, (5) will be transformed to

$$u(t) = a_1 D^{\alpha}[c(t) + e(t)] + a_2 D^{\beta}[c(t) + e(t)] + a_3[c(t) + e(t)] \quad (8)$$

So (8) will not give an accurate relation between a_1, a_2 and a_3 due to the presence of the terms $a_1 D^{\alpha} e(t)$, $a_2 D^{\beta} e(t)$ and $a_3 e(t)$. Likewise, equations obtained by applying the transformation $c(t) \equiv c(t) + e(t)$ on (6) and (7) will also be inaccurate. Our aim will be to minimize this inaccuracy.
One significant fact we observed is that for the same random error waveform e(t), $D^{\alpha_1} e(t) << D^{\alpha_2} e(t)$ if $\alpha_1 < 0$ and $\alpha_2 > 0$ when in effect, $D^{\alpha_1} e(t)$ becomes an integration.
A rigorous mathematical proof explaining this observation cannot be presented here for constraint of space. But for now, a philosophical explanation may be put forward as follows. The physical significance of differentiation is the slope of the function we want to differentiate (although this is not strictly the case for fractional differentiation), whereas integration deals with the area under the curve. The random error component we considered consists of fluctuations having both positive and negative values. Thus an integration operation over this error waveform can be expected to yield quite a low value, since the areas with opposing signs should nullify the effects of each other. On the other hand, integration of the output waveform will give a high positive result, since the output waveform can assume non-negative values only. So the effect due to the error component is minimized.
However, we cannot say anything definite about the results of differentiation operations, in fact no pattern can be ascribed to the result obtained by differentiating either the output or the error waveforms.
To support our contention, in table 1, we tabulate the values of $D^{\alpha} e(t)$ for 10 different sets of randomly generated e(t) with $\alpha = 1.5, 1.2, 0.9, 0.6, 0.3, -0.3, -0.6, -0.9, -1.2, -1.5$. The amplitude of e(t) varies between –0.01 and 0.01. Length of memory = 10 seconds, i.e. the fractional derivatives are calculated at time t = 10 seconds. Sampling rate is once in 0.001 seconds.
The transfer function of our system is $\frac{1}{a_1 s^{\alpha} + a_2 s^{\beta} + a_3}$, and as we are well aware, $\alpha, \beta > 0$ for a practical system, so that in (8), the orders of derivation α, β are positive. To remedy this, we can perform a simple transformation on the transfer function of the system, which we can write as $\frac{s^{-n}}{a_1 s^{\alpha-n} + a_2 s^{\beta-n} + a_3 s^{-n}}$, where $(n-1) < \alpha < n$ and $\alpha > \beta$.

Table 1. Variation of $D^{\alpha}e(t)$ with α. (The 10 sequences e(t) are consecutive and independent.)

e(t)	$D^{\alpha} e(t)$ for derivation order α									
	$\alpha =$ 1.5	$\alpha =$ 1.2	$\alpha =$ 0.9	$\alpha =$ 0.6	$\alpha =$ 0.3	$\alpha =$ -0.3	$\alpha =$ -0.6	$\alpha =$ -0.9	$\alpha =$ -1.2	$\alpha =$ -1.5
1	-435.7842	-50.7575	-5.7583	-0.6287	-0.0661	-0.0008	0.0001	0.0006	0.0013	0.0025
2	-603.6659	-59.5517	-5.7933	-0.5742	-0.0617	-0.0013	-0.0005	-0.0004	-0.0002	0.0002
3	424.4136	44.8209	4.7948	0.5242	0.0581	0.0002	-0.0002	-0.0001	-0.0001	-0.0003
4	-256.3730	-26.5634	-3.0495	-0.3928	-0.0549	-0.0011	-0.0002	-0.0001	-0.0002	-0.0005
5	-107.8138	-12.0636	-1.4119	-0.1631	-0.0164	0.0004	0.0004	0.0005	0.0008	0.0013
6	642.4164	78.0679	9.1798	1.0311	0.1069	0.0006	-0.0001	-0.0002	-0.0006	-0.0012
7	184.7026	22.6486	2.6896	0.3051	0.0321	0.0000	-0.0001	-0.0001	0.0002	0.0003
8	-393.9215	-45.1151	-5.0296	-0.5467	-0.0562	0.0001	0.0002	0.0000	-0.0001	-0.0002
9	-109.5421	-4.2756	0.4383	0.1517	0.0272	0.0006	0.0001	0.0000	-0.0001	0.0000
10	-32.4628	-7.6152	-1.4990	-0.2670	-0.0439	-0.0008	0.0002	0.0007	0.0014	0.0024

Proceeding as before we can now obtain our three simultaneous equations as:

$$D^{-n}u(t) = (a_1 D^{\alpha-n} + a_2 D^{\beta-n} + a_3 D^{-n})[c(t) + e(t)] \quad (9)$$

$$D^{-n-1}u(t) = (a_1 D^{\alpha-n-1} + a_2 D^{\beta-n-1} + a_3 D^{-n-1})[c(t) + e(t)] \quad (10)$$

$$D^{-n-2}u(t) = (a_1 D^{\alpha-n-2} + a_2 D^{\beta-n-2} + a_3 D^{-n-2})[c(t) + e(t)] \quad (11)$$

It can now be checked that all orders of derivation are now negative so that we will actually be performing fractional order integrations rather than fractional order differentiations.

IV. Illustration

Let the process whose parameters are to be estimated is $\frac{1}{a_1 s^{2.23} + a_2 s^{0.88} + a_3}$.

The input considered is $r(t) = 1$ i.e. unit step.

Synthetic data for $c(t)$ is created using $a_1 = 0.8$, $a_2 = 0.5$ and $a_3 = 1$, i.e. the values of c(t) are obtained at different time instants (using a MATLAB program) assuming a process with transfer function $\frac{1}{0.8s^{2.23} + 0.5s^{0.88} + 1}$. The simultaneous equations corresponding to (9), (10) and (11) are

$$D^{-3}u(t) = (a_1 D^{-0.77} + a_2 D^{-2.12} + a_3 D^{-3})[c(t) + e(t)]$$

$$D^{-4}u(t) = (a_1 D^{-1.77} + a_2 D^{-3.12} + a_3 D^{-4})[c(t) + e(t)]$$

$$D^{-5}u(t) = (a_1 D^{-2.77} + a_2 D^{-4.12} + a_3 D^{-5})[c(t) + e(t)]$$

Length of memory L = 10 seconds and T = 0.001 seconds is used to calculate the fractional derivatives.

We will display the accuracy of identification when the output readings used to calculate the fractional derivatives are ideal and also when they are erroneous to the extent of a random error component in the range $[-0.05, 0.05]$ in each reading. This error component is quite large since the output response is often below unity. The output response of the system for unit step input is shown both in presence and absence of the error component.

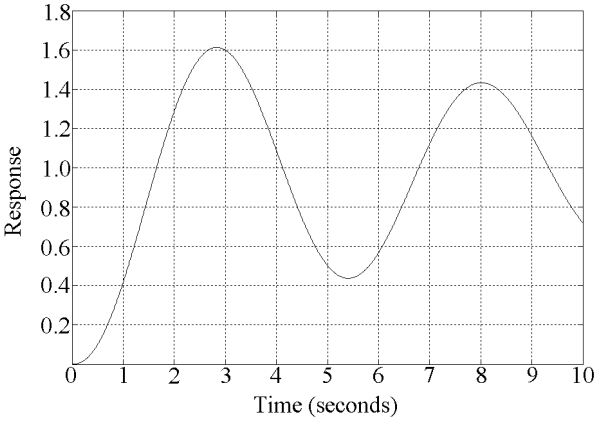


Fig. 1 Actual unit step response.

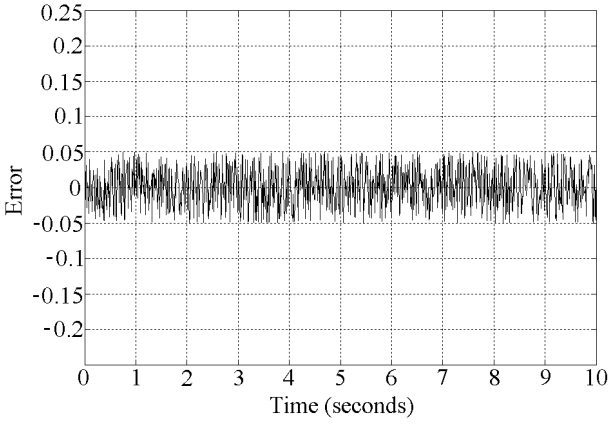


Fig. 2 Random error component added.

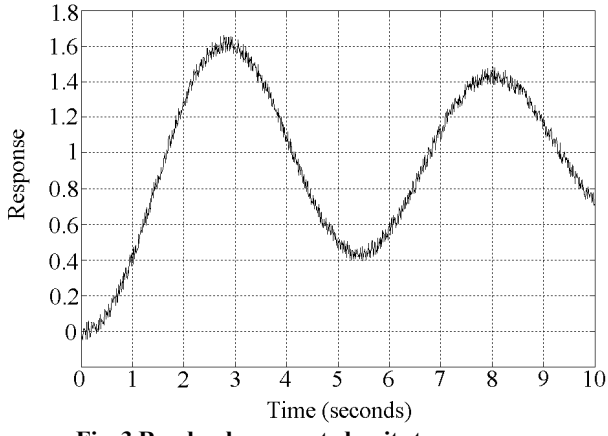


Fig. 3 Randomly corrupted unit step response.

A. Ideal Case: e(t) = 0 for all t

The following derivatives are then calculated numerically using (1), (2) and (3):

$D^{-0.77}c(t) = 6.1777$, $D^{-2.12}c(t) = 51.3011$,

$D^{-3}c(t) = 136.1477$, $D^{-1.77}c(t) = 32.2818$,

$D^{-3.12}c(t) = 152.6826$, $D^{-4}c(t) = 314.8183$,

$D^{-2.77}c(t) = 108.0207$, $D^{-4.12}c(t) = 342.4005$,

$D^{-5}c(t) = 576.6986$.

The set of simultaneous equations is

$$\begin{bmatrix} 6.1777 & 51.3011 & 136.1477 \\ 32.2818 & 152.6826 & 314.8183 \\ 108.0207 & 342.4005 & 576.6986 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 166.7167 \\ 416.9167 \\ 834.1670 \end{bmatrix}$$

After solving, we have $a_1 = 0.8001, a_2 = 0.4996, a_3 = 1.0000$ as the unknown parameters. The errors in estimating them are respectively 0.0125%, 0.0800% and 0%.

Summation of square errors of this process model outputs relative to the output data set for unit step input is 0.0030.

B. Non-ideal Case: Each Element in e(t) is Between –0.05 and 0.05

To each reading c(t) is added a random error component varying between –0.05 and 0.05.

We proceed as before to obtain the set of simultaneous equations as

$$\begin{bmatrix} 6.1798 & 51.3179 & 136.1948 \\ 32.2919 & 152.7357 & 314.9314 \\ 108.0577 & 342.5242 & 576.9207 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 166.7167 \\ 416.9167 \\ 834.1670 \end{bmatrix}$$

After solving, we have $a_1 = 0.7992, a_2 = 0.4996, a_3 = 0.9996$ as the unknown parameters. The errors in estimating them are respectively 0.1000%, 0.0800% and 0.0400%.

Summation of square errors of this process model outputs relative to the actual output data set for unit step input is 0.0062.

So, we can conclude that the proposed identification scheme is highly accurate even in presence of large quantities of random error.

V. Complete Identification by Applying PSO

From sections III and IV, we understand that if we can find the fractional powers α and β first, we can then accurately identify the other parameters (coefficients). We will employ a standard two-parameter optimization PSO algorithm where the two parameters to be optimized are the fractional powers of the system.

Let the range of variation of α is α_{min} to α_{max} and the range of variation of β is β_{min} to β_{max}. Then, initially a random solution set of (α,β) is generated with the position vector limits $\{(\alpha_{min},\alpha_{max}),(\beta_{min}\beta_{,max})\}$. The position vector (α,β) is updated using the PSO dynamics [13]. So the solution space is two-dimensional, as are the position and velocity vectors.

The fitness function $F(\alpha,\beta)$ is derived as follows. We apply a test signal R(s)=1/s (unit step) to the actual system and obtain sampled values of output c(t). For $\alpha = \alpha$ and $\beta = \beta$, the coefficient terms a_1, a_2, a_3 is generated. Let, for the process model, the output response for unit step input (obtained by simulation) is p(t). We will define a parameter $F = \sum_{t=i}^{f} [c(t) - p(t)]^2$, which gives a measure of the deviation of the output of the trial process model from the output of the actual process. F is the fitness function that the PSO algorithm will try to minimize. At F = 0, the unknown parameters (α,β) are optimized. The corresponding a_1, a_2, a_3 are the identified coefficients. The process model corresponding to the optimized solution set should provide output identical to c(t) for unit step input.

Clearly, our only source of information about the actual process is the set of output readings from it. The PSO algorithm will try to find a process model that matches these readings. So we have to perform one transformation from s-domain to discrete time domain, since the actual readings will obviously be in time domain. An alternative approach is to convert all data into z-domain. But then we would have been required to perform two transformations: discrete time domain to z-domain and s-domain to z-domain. Our approach is thus simpler.

Admittedly, the same problem could have been solved by a straightforward application of a five-parameter estimation PSO with the five parameters as $\alpha, \beta, a_1, a_2, a_3$. But by the direct use of fractional calculus definitions we have simplified the problem to the optimization of just two parameters (the fractional powers). The other three parameters can be derived from these two and are not independent. So we have presented a more efficient approach.

VI. Illustration of Complete Identification

We have used a sampling frequency of 20 hertz, i.e. the output waveform is sampled once every 0.05 seconds.

The range of variation α is 2.0 to 2.4, that of β is 0.7 to 1.1.

The PSO parameters used are: the inertia factor ω decreases linearly from 0.9 to 0.4, the cognitive learning rate c_1=1.4, the social learning rate c_2=1.4.

Number of PSO particles in the population is 10. The PSO algorithm is run for 40 iterations, and this is kept as the stop criterion.

The position vectors of the particles are randomly initialized in the range [(2.0, 2.4), (0.7, 1.1)]. The velocity vectors are randomly initialized in the range [(-0.2, 0.2), (-0.2, 0.2)]. The limits on the position vectors are [(2.0, 2.4), (0.7, 1.1)]. No limit is kept on the velocity vectors.

Now the PSO algorithm is run. After 40 iterations, the best particle has the position (2.2301, 0.8808), which is the identification of the fractional powers. The identification of the three coefficients (a_1, a_2, a_3) is (0.7996, 0.4998, 1.0001). The value of the best fitness after 40 iterations is 4.7388×10^{-6}.

So, the identified system is $\frac{1}{07996s^{2.2301} + 0.4998s^{0.88} + 1.0001}$. The percentage errors in identification for $\alpha, \beta, a_1, a_2, a_3$ are respectively 0.0045, 0.0909, 0.0500, 0.0400, 0.0100.

Some relevant graphs are presented below.

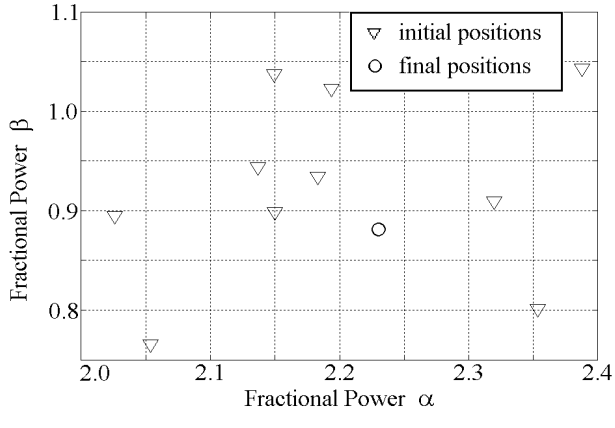


Fig. 4 Convergence of the PSO particles.

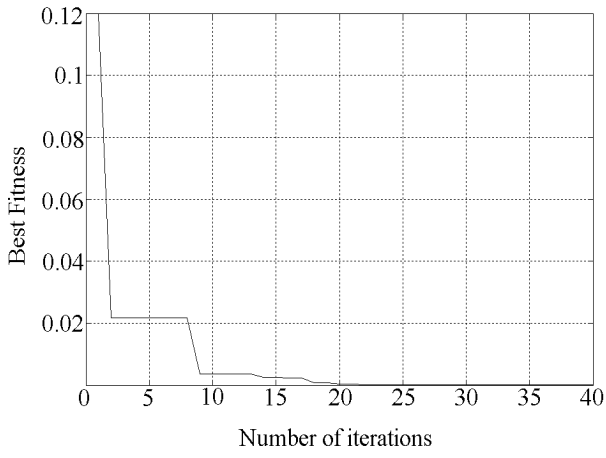


Fig. 5 Variation of best fitness with number of iterations for one run.

Results from 10 other sample runs (using identical conditions) are shown in table 2.

Table 2. Ten sample runs of the identification algorithm.

α	β	a_1	a_2	a_3	Fitness X 10^6
2.2301	0.8810	0.7996	0.4998	1.0002	3.8420
2.2304	0.8814	0.7993	0.5000	1.0002	5.2414
2.2302	0.8812	0.7994	0.4999	1.0002	3.9793
2.2302	0.8811	0.7995	0.4999	1.0002	3.5780
2.2301	0.8810	0.7995	0.4998	1.0002	3.6754
2.2301	0.8808	0.7997	0.4998	1.0001	4.7388
2.2302	0.8812	0.7994	0.4999	1.0002	3.6306
2.2301	0.8810	0.7995	0.4998	1.0002	3.8420
2.2302	0.8811	0.7995	0.4999	1.0002	3.5800
2.2303	0.8812	0.7994	0.4999	1.0002	3.8707

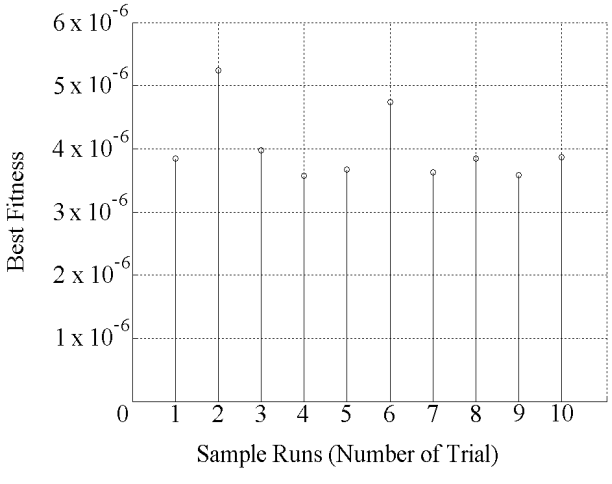


Fig. 6 Variation of best fitness with sample runs (using data from table 2).

VII. Comparison, Comments and Conclusion

An elegant method for the identification of parameters of a fractional order system is proposed. The process of identification can actually be implemented by a simple computer program in C or MATLAB. Of course, the same method can easily be employed to estimate the parameters of an integer order process model as well.

The challenge in fractional order system identification is that the fractional powers are not restricted to assume only discrete integral values, but are distributed in a continuous interval. For two integral order systems, identical time domain responses mean identical transfer functions. But for fractional order systems, we often find that a better identification of the actual process has actually a lower fitness than a worse model.

The method of finding a relation between the coefficients by use of fractional calculus simplifies a complex problem. The subsequent application of an evolutionary algorithm thus provides very accurate estimations. So the proposed scheme is both simple and efficient. Therein lies its merit.

References

[1] J. P. Norton, *An Introduction to Identification,* Academic Press, London, 1986.

[2] K. J. Astrom, B. Wittenmark, *Adaptive Control, Second Edition*, Pearson Education, Inc., 1995.

[3] L. Ljung, *System Identification- Theory for the User,* Prentice-Hall, Englewood Cliffs, N.J., 1987.

[4] P. J. Torvik, R. L. Bagley, "On the appearance of the fractional derivative in the behaviour of real materials", Transactions of the ASME, vol. 51, June 1984, pp. 294-298.

[5] L. Dorcak, V. Lesko, I. Kostial, "Identification of fractional-order dynamical systems", Proceedings of the 12th International Conference on Process Control and Simulation – ASRTP'96, no. 1, Kosice, Slovak Republic, pp. 62-68.

[6] I. Podlubny, *Fractional Differential Equations,* Academic Press, San Diego, 1999.

[7] L. Dorcak, *Numerical Methods for Simulation the Fractional-Order Control Systems*, UEF SAV, The Academy of Sciences Institute of Experimental Physics, Kosice, Slovak Republic, 1994.

[8] I. Podlubny, *The Laplace Transform Method for Linear Differential Equations of the Fractional Order*, UEF-02-94, The Academy of Sciences Institute of Experimental Physics, Kosice, Slovak Republic, 1994.

[9] K. B. Oldham, J. Spanier, *The Fractional Calculus*, Academic Press, New York, 1974.

[10] en.wikipedia.org/wiki/Riemann_Liouville_differintegral.

[11] en.wikipedia.org/wiki/Fractional_calculus.

[12] J. Kennedy, R. C. Eberhart, "Particle swarm optimization", Proc. of the IEEE International Conference on Neural Networks, pp.1942-1948, 1995.

[13] Y. Shi, R. C. Eberhart, "A modified particle swarm optimizer", Proc. of the IEEE Congress on Evolutionary Computation, pp. 69-73, 1998.

[14] M. P. Song, G. H. Gu, "Research on particle swarm optimization: A review", Proc. of the Third International Conf. On Machine Learning and Cybernetics, pp. 2236-2241,2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Adaptive Template Based Object Tracking with Particle Filter

Md. Zahidul Islam[1] and Chil-Woo Lee[2]

Dept. of Electronics & Computer Engineering, Chonnam National University
Gwangju, South Korea
E-mail: zahid@image.chonnam.ac.kr

Abstract - In this paper, we describe a new approach to improve the video based object tracking system with particle filter using shape similarity. It deals with single object tracking whose dynamics age highly non-linear. The shape similarity between a template and estimated regions in the video sequences can be measured by their normalized cross-correlation of distance transformation. Here within this present job, observation model of the particle filter is based on shape from distance transformed edge features. Template is created instantly by selecting any object in a video scene and updated in every frame. Experimental results have been offered to show the effectiveness of the proposed method.

I. Introduction

We have proposed a particle-filter based algorithm. It deals with a single object tracking whose dynamics are highly non-linear. Consistent object tracking in a cluttered visual environment is of great significance. Object tracking is one of the innermost and challenging tasks in computer vision problem such as visual surveillance, human computer interaction etc. Tracking objects is performed in a sequence of video frames and it consists of two main stages: isolation of objects from background in each frames and association of objects in successive frames in to trace them. Object tracking in image processing is usually based on reference image of the object, or properties of the objects. To start object tracking, generally the trackers need to be initialized by an outer component [1]. But, unfortunately robust and efficient object tracking is still an open research issue.

In our present proposed method, to measure the similarity between the template and estimated regions from particle, we use distance transform (DT) edge template matching. That is, we apply distance transform to edge images for the template and estimated particle regions, and then we apply normalized cross-correlation on them. To track efficiently we need a good observation model and we demonstrate normalized cross-correlation based observation model doing well for object tracking. Usually, in many systems, motion, colour, contour and shape are used as features for tracking. Some one used skin colour but this is used only for hand and face tracking and it has some limitations like unique definition of skin colour, shadows, occlusions and changing illuminations [6]. According to [7] assumption, the only moving objects in video scene is person. This postulation does not hold for many applications. But in this paper we try to make a general system for tracking any indoor outdoor object like person, face or hand by instant template matching. First we select any object form video scene manually, and then it is considered instantly as a template. Our system works in adaptive manner by updating template in every frame. The key mechanism of our developed observation model is based on similarity measures by normalized cross-correlation between DT based template and tracking object in a video scene. The more details we can find in overall proposed system model section.

The rest of the paper is organized as follows: section 2 describes related and existing work in this field. Section 3 introduces basic particle filter for object tracking. The proposed system models which unified distance transform, normalized cross-correlation with particle filter, are discussed in section 4. Section 5 verify about the proposed system with some experimental results on various real video data in different environments. Conclusive remarks are addressed at the end of the paper in section 6.

II. Previous Work

In this section we focus on various models and techniques for video based object tracking. Traditionally, the tracking problem is formulated as a sequential recursive estimation [8] having on an estimate of the probability distribution of the target in the previous frame, the problem is to estimate the target distribution in the new frame using all available prior knowledge and new information brought by the new frame. The state-space formalism, where the current tracked object properties are described in an unknown state vector updated by noisy measurements, is very well adapted to model tracking. Unfortunately the sequential estimation has an analytical solution under very restrictive hypothesis. The well known Kalman filter [17] is such a solution and is optimal for the class of linear Gaussian estimation problems. The Particle filter, a numerical method that allows finding an approximate solution to the sequential estimation, has been successfully used in many target tracking problems and visual tracking problems. But Kalman filter has limitations for multidimensional tracking. Particle filter success, in comparison with Kalman filter, can be explained by its capability to cope with multi-modality of the measurement densities and

non-linear observation models. In visual tracking, multi-modality of the measurement density is very frequent due to the presence of multifaceted scene elements which has a similar appearance to the target. The observation model, which relates the state vector to the measurements, is non-linear because image data endures feature extraction, a highly non-linear operation.

For hand tracking according to Israd et al. [9], they apply CONDENSATION algorithm to combine skin colour and hand contour. But that is not a general system for moving object tracking. There is an extension of CONDENSATION algorithm in [16]. But this algorithm can not be applicable for any all-purpose system and skin colour has some limitations. For object tracking, a colour based particle filter is proposed in many works [4,5]. In their case, target model of the particle filter is defined by the colour information of the tracked object. According to Lehuger et al. [4], an adaptive mixture colour model is used for updating reference colour model to make more robust visual tracking. In [10], they try to integrate, multiple features such as colour, shape, motion, edge. But finally they apply only colour and texture cue in particle filter based implementation. Moving edge features is used in [3]. So, from the various literature reviews we observe that, it is still a tricky task for more robust real time object tracking.

III. Particle-Filtering Background

Tracking objects in video involves the modelling of non-linear and non-gaussian systems. In order to model accurately the underlying dynamics of a physical system, it is important to include elements of non-linearity and non-gaussianity in many application areas. Particle Filters can be used to achieve this.

Particle filter is a sequential Monte Carlo methodology where the basic idea is the recursive computation of relevant probability distributions using the concepts of importance sampling and approximation of probability distributions with discrete random measures. The fundamental idea of Particle filter approximates the filtered posterior (next) distribution (density) by a set of random particles (sampling) with associated weights. It weights particles based on a likelihood score and then propagates these particles according to a motion model. Particle filtering assumes a Markov Model for system state estimation. Markov model states that past and future states are conditionally independent of a given current state. Thus, observations are dependent only on current state. A block diagram of particle filtering is shown in Fig. 1.

Particle filter consists of essentailly two steps: prediction and update. Given all available observations $y_{1:t-1} = \{y_1, \ldots, y_{t-1}\}$ up to time $t-1$, the prediction stage uses the probabilistic system transition model $p(x_t \mid x_{t-1})$ to predict the posterior at time t as

$$p(x_t \mid y_{t-1}) = \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1} \quad (1)$$

At time t, the observation y_t is available, the state can be updated using Bay's rule

$$p(x_t \mid y_{1:t}) = \frac{p(y_t \mid x_t) p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \quad (2)$$

where $p(y_t \mid x_t)$ is described by the observation equation. In the particle filter, the posterior $p(x_t \mid y_{1:t})$ is approximated by a finite set of N samples $\{x_t^i\}_{i=1,\cdots,N}$ with importance weights w_t^i. The candidate samples $\tilde{x}_t^i$ are drawn from an importance distribution $q(x_t \mid x_{1:t-1}, y_{1:t})$ and the weight of the samples are –

$$w_t^i = w_{t-1}^i \frac{p(y_t \mid \tilde{x}_t^i) p(\tilde{x}_t^i) \mid x_{t-1}^i}{q(\tilde{x}_t \mid x_{1:t-1}, y_{1:t})} \quad (3)$$

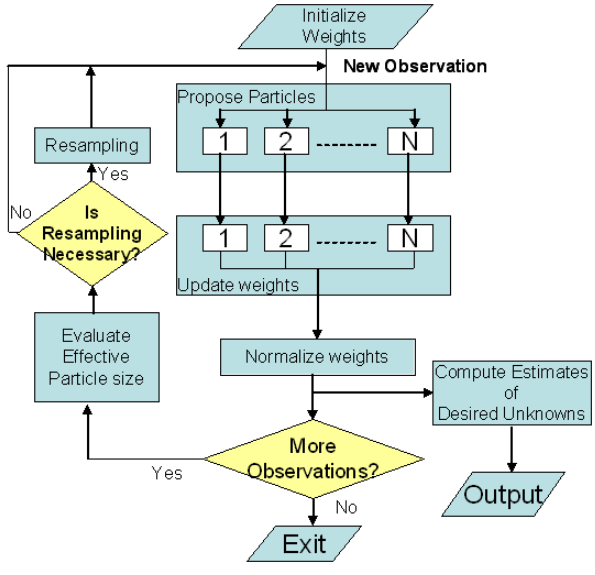


Fig. 1 Block diagram of particle filter. It shows the basic working flow and relational steps

The samples are resampled to generate an unweighted particle set according to their importance weights to avoid degeneracy. In the case of the bootstrap filter [13], $q(x_t \mid x_{1:t-1}, y_{1:t}) = p(x_t \mid x_{t-1})$ and the weights become the observation likelihood $p(y_t \mid x_t)$.

For implementation of particle filter we need the following mathematical model:

- Transition model / state motion model $P(x_t \mid x_{t-1})$: this specifies how objects move between frames.
- Observation model $p(y_t \mid x_t)$: this specifies the likelihood of an object being in a specific state (i.e. at the specific location).
- Initial state $Est(1)$ / prior distribution model $p(x_0)$: describes initial distribution of object states.

IV. Towards Robust Tracking

The task of robust tracking demands a robust observation model. To make template based observation model, in this paper, first we select and extract object shape information

from a video scene by edge detection (canny edge detector) and then employ robust shape matching based on distance transformation [12] and compares between first initialized object and tracked object by normalized cross-correlation [13]. Matching involves correlating the templates with the distance transformed scene and determining the locations where the mismatch is below a certain user defined threshold. Template is updated in adaptive fashion in tracking sequence.

A. Proposed Observation Model

Basically the observation model is used to measure the observation likelihood of the samples and this is an important concern for object tracking. In the last few years, many observation models have been developed for particle filter tracking. In [2,9], a contour based appearance template is chosen to model the target. The tracker based on a contour template gives an accurate description of the targets but poorly in clutter, non rigid object and is generally time consuming. Also, the initialization is not easy and tricky in this method. On the other hand, the colour based trackers are faster and more robust against contour based tracker. In this case, the colour histogram is typically used to model the targets to combat the partial occlusion, and non-rigidity. The draw back of the colour histogram is that spatial layout is ignored, and the trackers on it are easily confused by a background with similar colours.

We use to develop our observation model by distance transform (DT) based shape information which has so many positive features. To take better effort form DT based image, we use normalized cross correlation (NCC) based matching score, which is more invariant than ordinary DT matching score. Such a combination significantly improves the robustness and discriminative power of the tracking features.

The cross correlation based template matching is motivated by the distance measure (squared Euclidean distance). There are several disadvantages using this approach, for example, if the image energy varies with position, matching can fail. Moreover, it is not invariant to changes in image amplitude such as those caused by changing lighting conditions across the image sequences. The correlation coefficient overcomes these difficulties by normalizing the image and feature vectors to unit lengths, yielding a cosine-like correlation coefficient. So we are intended in our present work for DT image map based matching with normalized cross-correlation to develop observation model to make more robust particle filter based tracker. We try to take both advantages from DT image map and normalized cross correlation regarding to develop our particle filter based tracking system.

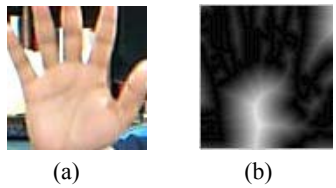
B. Likelihood Function

The likelihood function gives the probability density function of image features given the state. The selected features are pixel-oriented. Thus the appearance will be given by a matrix whose elements are the pixel intensity values. Therefore, it can be assumed that the appearance is independent from the speed component.

Since our likelihood function depends on the sample position but does not depend on its speed, propagated samples could have a small position error, but their speed values could become completely different from the true one in a few frame.

C. Distance Transformation (DT)

The typical matching [14] with DT is shown schematically in Figure 1 which involves two binary images, a segmented template and a segmented image which we call feature template and feature image. To formalize the idea of DT matching similar to chamfer matching [15], the shape of an object is represented by a set of point. The image map is represented as a set of feature points. In our present work we always update our reference template in every frame. So, it is more robust for matching in each changing of tracked object. DT based matching has several advantages such as, in order to be tolerant to small shape variations, any similarity function between two shapes should vary smoothly when the feature point locations change by small amounts. The typical original image and DT image map is shown in Figure 2. In the original image if we select the hand as template for matching to find it, followed by DT we can match it very smoothly and robustly by means of normalized cross-correlation which is discussed in next section.



(a) (b)

Fig. 2 (a) Original Image. (b) Corresponding DT image map.

D. Matching by Normalized Cross-correlation

Traditional correlation based matching methods are limited to the short baseline case [13]. According to Zhao *et al.* [13], normalized cross-correlation (NCC) is proposed for matching two images with large camera motion. Their method is based on the rotation and scale invariant normalized cross-correlation. In our present case, we use normalized cross-correlation for image processing applications in which the brightness of the image and template can vary due to lighting and exposure conditions, the image can be first normalized. This is done at every step by subtracting the mean and dividing by the standard deviation. That is, the cross-correlation of template *t(x,y)* with a subimage *f(x,y)* is given by the following equation

$$N_{f,t} = \sum_{x,y} \frac{(f(x,y) - \bar{f})(t(x,y) - \bar{t})}{\sigma_f \sigma_t} \quad (4)$$

E. Particle Filter Based Implementation

In our present work we integrate the template matching by normalized cross-correlation with particle filter for robust

object tracking. In our implementation we flow some steps which is given below:

- **Adaptive template based probabilistic tracking:** These trackers rely on the deterministic search of a window, whose shape information matches a reference template. In this matching case, first, instantly selected template image is converted to DT image and then normalized cross-correlation principle between reference template and tracked moving object has been applied.
- **State Space:** We have modeled the states, as its location in each frame of the video. The state space is represented in the spatial domain as X = (x,y). We have initialized state space for the first frame manually by selecting the object of interest in the video scene by rectangle.
- **System Dynamics:** A second-order auto regressive dynamics is chosen on the parameters used to represent our state space i.e. (x,y). The dynamics is given as: $X_{t+1} = Ax_t + Bx_{t-1}$. Matrices A and B could be learned from a set of sequences where correct tracks have been obtained.
- **Observation y_t** The observation y_t is proportional to the normalized cross-correlation matching score between tracked window of the predicted location in the frame and the reference template. In every frame the reference DT template is updated according to object motion in the video scene. So, y_t α NCC Score (q,q_x), where, q is the reference template and q_x is the predicted location window.

The particle filter iteration using our proposed method is as follows:

- Initialize x_t for first frame. The selected object become reference template for the first frame and it is updated in every frame.
- Generate a particle set of N particles $\{x^m_t\}_{m=1..N}$
- Prediction for each particle using second order auto-regressive dynamics.
- Compute normalized cross-correlation matching score.
- Weight each particle based on matching score.
- Select the location of target as a particle with best matching score.
- Re-Sampling the particles for next iteration.

V. Experimental Results

In this section we will demonstrate some experimental results on several real-world video sequences captured by pan/tilt/zoom video camera in indoor, outdoor and office environment. The captured sequences simulate various tracking conditions including quick movement, shape deformation, background clutter, appearance changes, camera pan/tilt/zoom, and partial occlusion. For all testing sequences, we use the same algorithm configuration, e.g. state-space, system dynamics, observation model and template based probabilistic tracking. In the first experiment, resulting sequence is taken from indoor environment for tracking a single moving person under static background environment. These sequences are shown in Figure 3.

For each experiment, initialization is done by manually which make template instantly by selecting the region of interest (ROI) and we use 120 number particles.

The second and third resulting sequences are also showing the effectiveness of our proposed system model. These sequences are taken from outdoor environment for tracking single moving person among multi persons. The first outdoor sequence is shown in Figure 4 and second one is shown in Figure 5.

Fig. 3 Single moving person tracking in indoor static environment.

This tracking algorithm is implemented in C++ and OpenCV library on Windows XP platform with standard Pentium 4 (with 1.5 GB RAM) machine. So, it can be concluded from all experimental results, that our proposed system can be very efficient general system as instant / adaptive template based object tracker. These all results show the algorithm performance under different scenarios

Fig. 4 Moving person tracking in outdoor environment (first outdoor sequence).

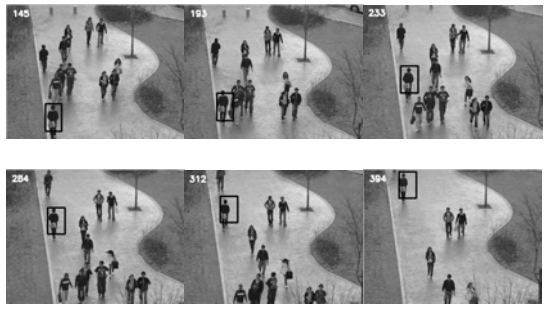


Fig. 5 Moving person tracking in different outdoor environment (second outdoor sequence).

VI. Conclusion

In this paper, a new technique is presented for tracking a moving object in a video sequence. In the unified framework of particle filtering, we presented a tracking algorithm based on instant / adaptive template matching by normalized cross-correlation. The template image is converted as DT image map before measuring similarity by normalized cross-correlation. The observation model is updated by the best matching score in every frame. Our proposed system can be initialized by rectangle indicating the object position and then adapt itself to the changing appearance or environments and thus make more robust tracking. The system has been tested on a variety of video data and some satisfactory results have been obtained. Our system can be more upgraded in various ways. The observation model can be developed with the both colour and shape information. Instead of distance transform we can also use shape description method and for real-time processing we can improve the system performance with multi-core processing. And finally, the whole system will be applied for multiple targets tracking whose dynamics are highly non-linear based on shape and colour combination likelihood.

ACKNOWLEDGMENT

This research work was supported by MIC & IITA through IT leading R&D support project [2006-S-028-01].

References

[1] C. Yunqiang and R. Yong, "Real time object tracking in video sequences," Signals and Communications Technologies, Interactive Video, Part II: pp. 67 – 88, 2006.

[2] P. Li, T. Zhang, and A. E. C. Pece, "Visual contour tracking based on particle filters," Image Vision Comput., 21(1):111 – 123, 2003.

[3] V. T. Krishna and R. N. Kamesh, "Object tracking in video using particle filtering," In Proc. IEEE Intl. Conf. on Accou., Speech, and Signal Proces,. Volume 2, pp. 657 – 660, March 2005.

[4] A. Lehuger, P. Lechat and P. Perez, "An adaptive mixture color model for robust visual tracking," In Proc. IEEE Int. Conf. on Image Process, pp. 573 – 576. October, 2006.

[5] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "A color-based particle filter," In Proc. First International Workshop on Generative-Model-Based Vision, pp. 53 – 60, 2002.

[6] A. Koschan, S. K. Kang, J. K. Paik, B. R. Abidi, and M. A. Abidi, "Video Object Tracking Based on Extended Active Shape Models with Color Information," In Proc. 1st European Conference on Color in Graphics, Imaging, and Vision, CGIV2002. Poitiers, France, pp. 126 – 131, April 2002.

[7] W. Lu and Y. P. Tan, "A color histogram based people tracking system," In Proc. ISCAS 2001, Volume 2, pp. 137 – 140, 2001.

[8] J. Czyz, B. Ristic and B. Macq, "A particle filter for joint detection and tracking of color objects," Image and Vision Computing, 25(2007), pp. 1271 – 1281, 2006.

[9] M. Israd and A. Blake, "CONDENSATION – conditional density propagation for visual tracking," Int. journal of computer vision, 29(1): 893 – 908, 1998.

[10] L. Mihaylova, P. Brasnett, N. Canagarajah and D. Bull, "Object Tracking by Particle Filtering Techniques in Video Sequences," Advances and Challenges in Multisensor Data and Information Processing, Vol. 8, NATO Security Through Science Series: Information and Communication Security, E. Lefebvre (Ed.), IOS Press, the Netherlands, pp. 260-268, 2007.

[11] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo and J. Miguez, "Particle Filtering," IEEE signal processing magazine, pp. 19 – 38, September, 2003.

[12] B.D.R. Stenger, "*Model-Based Hand Tracking Using A Hierarchical Baysian Filter*," Ph.D. Thesis, University of Cambridge, March 2004.

[13] F. Zhao, Q. Huang and W. Gao, "Image Matching by Normalized Cross-Correlation," In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006. Volume 2, pp. II-II, 14-19, May 2006.

[14] D.M. Gavrila, "Multi-feature Hierarchical Template Matching Using Distance Transforms," In Proc. IEEE ICPR, Brisbane, Australia, 1998.

[15] B. Gunilla, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 10, No. 6, December 1988.

[16] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model-switching," In proc. ICCV98, pp. 107 – 112, January, 1998.

[17] A. M. Sanjeev, M. Simon, G Neil and C. Tim, "A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking," IEEE transactions on signal processing, vol. 50, no. 2, pp. 174 – 188, February 2002.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Microprocessor based Temperature Monitoring and Control System using Fuzzy Logic Controller

Md. Rabiul Islam, M. A. Goffar Khan* and M. F. Rahman*

Department of Computer Science & Engineering
Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology, Rajshahi-6204,Bangladesh
E-mail: rabiul_cse@yahoo.com, *qmagk@yahoo.com and *frahman3@yahoo.com

Abstract - A closed loop control system incorporating fuzzy logic has been developed for a class of maintenances of temperature of a heating chamber. A unique fuzzy logic controller (FLC) structure with an efficient realization and a small rule base that can be easily implemented in existing industrial controllers was proposed. The performance of this system has been measured in various environmental conditions such as compensating for thermo mass changes in the system, dealing with unknown and variable delays, operating at different temperature setpoints without retuning, etc. It is achieved by implementing, in FLC, a classical control strategy and an adaptation mechanism to compensate for the dynamic changes in the system. This paper deals with the detail issue for the improvement of microprocessor based design and implementations of the heating chamber temperature maintenances on practical demand and finally focuses the performance of this system to other environmental conditions in various systems.

I. Introduction

Temperature control is widely used in various processes. [1, 2, 3] These processes, no matter it is a process of large industrial plant, or a process in home appliance, share several unfavorable features such as non-linearness, interference, dead time, and external disturbance, etc. Conventional control approaches usually cannot achieve satisfactory results for this kind of processes. [4, 5, 6] The full enhanced design of a fuzzy logic controller to control the temperature according to the different environments has been presented in this paper [1].

II. Paradigm for the Fuzzy Temperature Control

A fuzzy control system is a closed loop system that uses the process of fuzzification. Fig.1 shows the block diagram for the maintenance of the heating chamber temperature.

III. Design Goals

Control of the environment for large computing systems is often a greater challenge. Not only the systems themselves generate heat, but also they are often specified by their manufacturers to be maintained in as little as a plus-or-minus 1 degree (Fahrenheit) range. Humidity is also a challenge, causing, for example, corrosion and jamming of associated mechanical systems at high humidity levels and the enhanced possibility of static discharge with low levels. Humidity control is often specified as 50% relative humidity, with a maximum swing of plus-or-minus 3% per hour.

In this work, the above conditions have been considered correctly in the implementation phases.

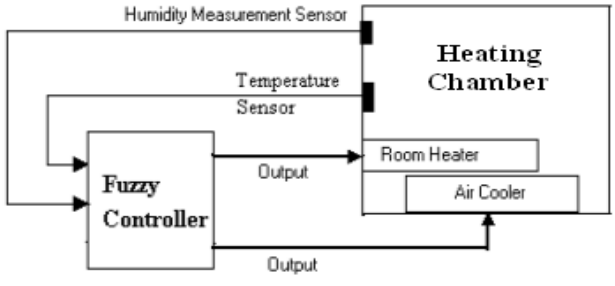


Fig. 1 Paradigm for the control of heating chamber temperature using fuzzy controller

IV. Design of the Fuzzy Logic Controller

To design the fuzzy logic [7,8, 9, 10, 11, 12], two inputs and two outputs has been taken. One input has been used to measure the temperature (Named as Temp_Sense) and another has been used to get the humidity (Named as Hum_Sense). As the same way of inputs, two outputs have been used to control the temperature on the heating chamber. One output has gone to the heating chamber heater (Named as Heat_Response) and other one has gone to the air cooler (Named as Cool_ Response). The steps that are required to design the control of the FLC for heating chamber temperature are given below.

Step 1: Define Inputs and Outputs for the Fuzzy Logic Controller and Set up the Fuzzy Membership Function for the Outputs

Labels and membership functions (whose shapes are simple, such as triangles) of input variables are defined in Fig. 2 and Fig. 3. The two outputs of the FLC has been designed as same as fuzzy range and membership function. Similarly, the output variables are in Fig. 4. All the values of variables here are normalized into the range of -1 to 1 or 0 to 1.

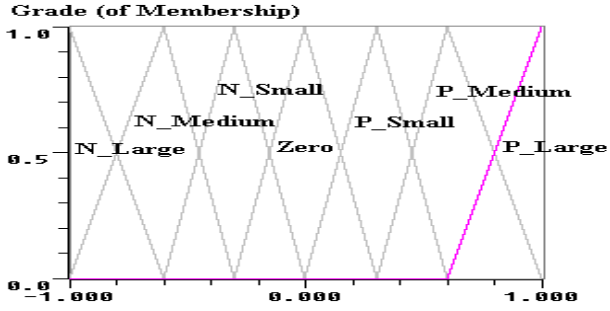


Fig. 2 Level and Triangular Membership function for the input variable Temp_Sense for Temperature

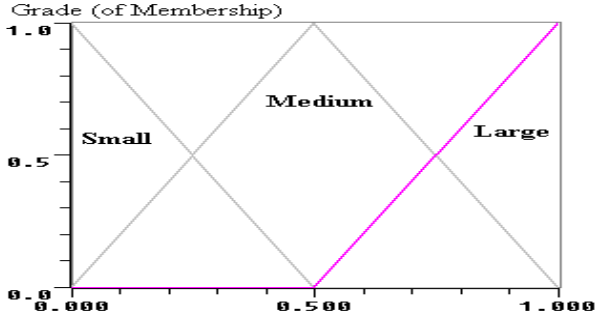


Fig. 3 Level and Triangular Membership function for the input variable Hum_Sense for Humidity

Step 2: Create a Fuzzy Rule Base

By using rule base, FLC makes the decisions according to the inputs. Since the changing rate of Heat_ Response and Cool_ Response is considered as same, Table-1 shows the fuzzy rule base for two outputs.

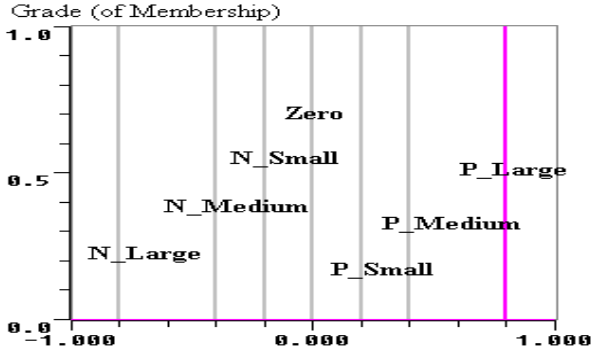


Fig. 4 Level and Membership function for the output variables Hot_Response for heating chamber heater and Cool_Response for heating chamber cooler.

Table 1 Fuzzy rule base

		Hum_Sense		
		Small	*Medium*	*Large*
Temp_Sense	*N_Large*	P_Large	P_Medium	-
	N_Medium	P_Large	P_Medium	-
	N_Small	P_Small	P_Small	-
	Zero	N_Small	P_Small	-
	P_Small	N_Small	N_Small	-
	P_Medium	N_Small	N_Medium	-
	P_Large	-	-	-

The IF-THEN rule can also be created such as,

IF Temp_Sense is **N_Large** AND Hum_Sense is **Small** THEN SET **Heat_Response** and **Cool_Response** are **P_Large**.

*Step.3:*Defuzzify the Outputs

There are several different strategies of defuzzification process. In this work, Centroied method has been used. As a result, by using inputs (i.e. Hum_Sense and Temp_Sense) outputs (i.e. Heat_Response and Cool_Response) have been easily found.

V. Implementation of the FLC

There are two inputs and two outputs to control the heating chamber temperature, FLC takes two inputs (i.e. temperature and humidity) from the heating chamber and sends two outputs to control the temperature.

The overall block diagram of the temperature control and monitoring system is shown in the figure 5.

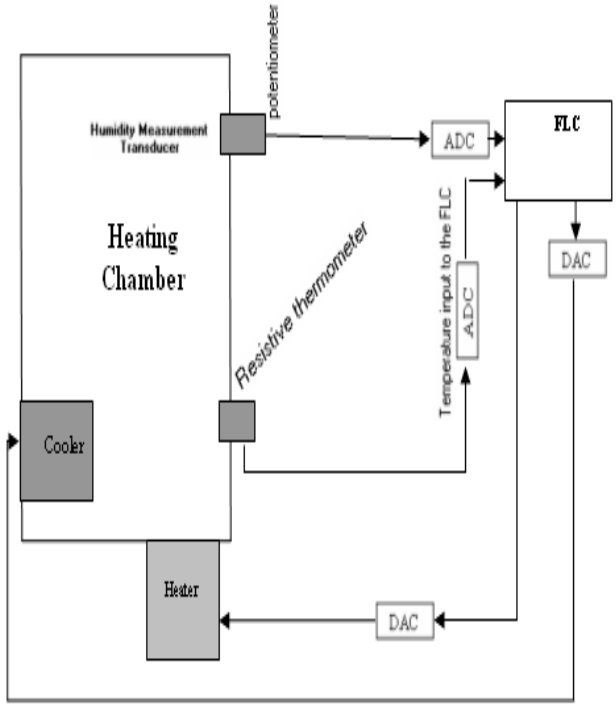


Fig. 5 Block diagram of the maintenance the temperature of a heating chamber.

A. Implementation of the control of temperature as an input to the FLC

A bead type thermistor is used to sense the temperature. Now the temperature is used to the operational amplifier and the output of the operational amplifier is gone to the FLC. The following mechanism has been used for this purpose.

The thermistor is connected to the signal conditioning circuit as shown in the Figure 6. The resistance of the sensor changes with the change of temperature. Therefore the output analog voltage depends on the change of temperature. The relation between output voltage and is non-linear. However over a small range of operation range, it can be considered as linear.

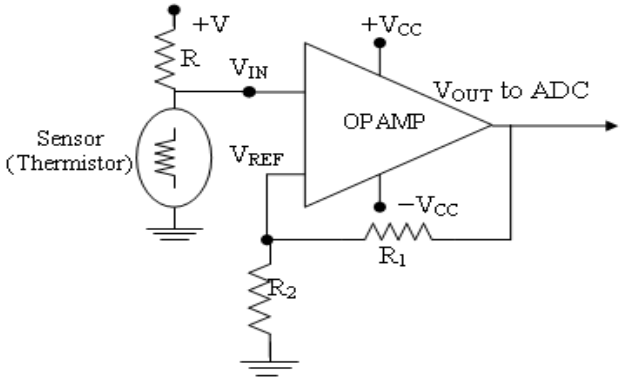


Fig. 6 Temperature sensor and signal conditioning circuit

The analog signal obtained at the output of the previous section is converted to an equivalent digital signal so as to feed to the computer. Conventional 0804 IC is used for this purpose. The circuit diagram is as shown in Fig.7. The analog signal is connected to pin no. 6 whereas the 8 bit digital output is available at pin no. 11 through 18. The conversion time is 100 microsecond, which is quite fast compared to the variation of the analog signal. The conversion process is inhibited by passing a HIGH signal at pin 3, which is generated by the computer programme, as described below. The digital signal is read by the computer and the temperature is monitored at an interval of 10ms.

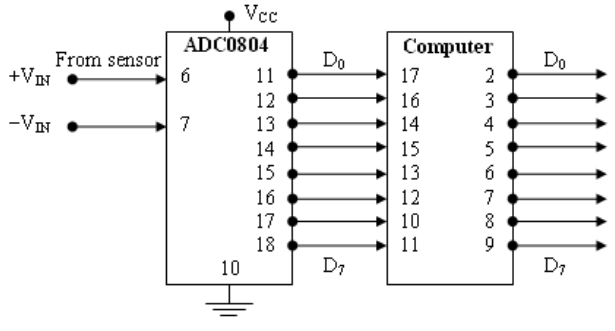


Fig. 7 Analogue to digital conversion and pc connection

To measure the humidity of the heating chamber, a humidity measurement transducer has been used. Based on the value of the transducer, the operational amplifier (i.e. Op amp 741 IC) gets some values. The reason for the use of operational amplifier is to amplify the signals from the transducer.

By getting two inputs from two sensors, the FLC can make decisions from the fuzzy rule base. These values have been used to control the heater and cooler. The output is sent to the heater and cooler using the following mechanism.

The computer program generates a control signal depending on the temperature. It is a 8 bit digital signal available at pin no. 2 through 9 of the parallel port of the computer. The digital signal is converted to an equivalent analogue signal by using an active R-2R ladder DAC, as shown in Fig. 8. Thus a 256 level analog signal is obtained which is connected to the inverting terminal of a comparator, the non-inverting terminal is fed with a triangular wave having time period equal to half cycle of the ac signal. Thereby a rectangular wave, having variable delay between the zero of ac signal and beginning of rectangular wave is obtained. The output is passed through a pulse generator, which generates a pulse of short duration, used as firing pulse for the SCR. An isolation circuit is used at the gate of the SCR for safety reason of the control circuit and the computer.

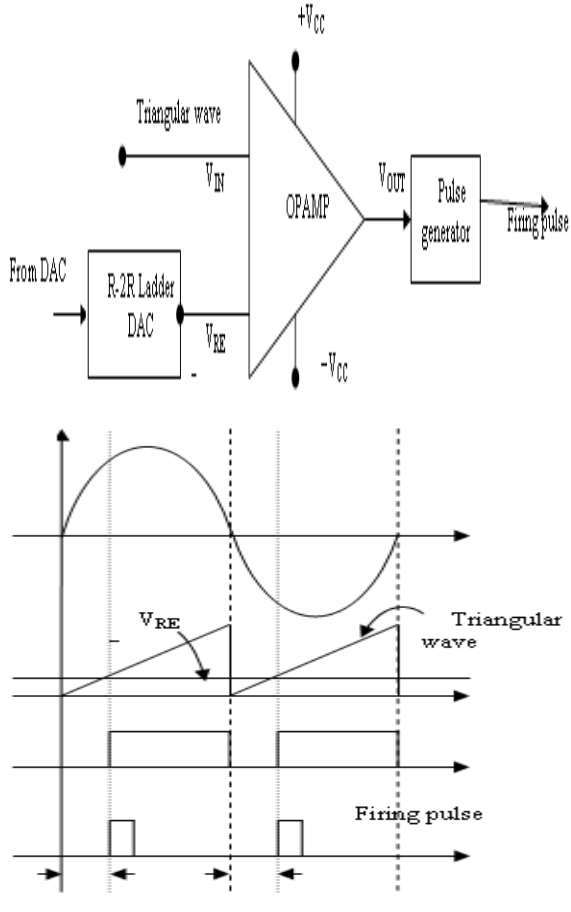


Fig. 8 Generation of firing pulse for SCR

B. Implementation of the Working Principle of the FLC

The FLC is a close loop controller that controls the total process. The control system may be digital or analog. In this work, digital system (i.e. a computer) has been used to control the overall process. The inputs and outputs are taken and sent to the computer (i.e. FLC) using printer port. The LPT1 port has been used to do this work.

C. Algorithm of the FLC Control Program

The algorithm of the temperature control program is shown below.

Algorithm for the control of the FLC:

Step1: Initialize the LPT1 Port.

Step2: Read inputs of the Hum_Sense and Temp_Sense through the LPT1port using pin no. 11 to 18.

Step3: If the range of Hum_Sense and Temp_Sense satisfies the inputs then go to the next process otherwise read inputs again.

Step4: By using Fuzzy Rule Base and two inputs the response is detected.

Step5: Select the values for the heating chamber heater knob (i.e. Heat_Response) and the heating chamber cooler Knob (i.e. Cool_Response).

Step6: Send these selected data to two knobs through LPT1 port.

Step7: End.

The flow diagram of the above algorithm for the control of FLC is shown in figure 9.

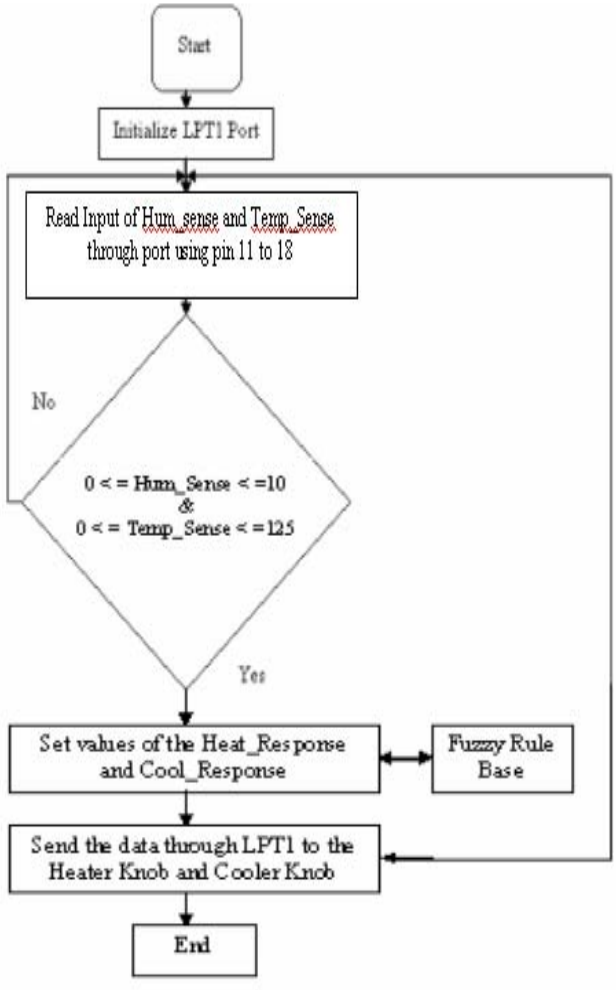
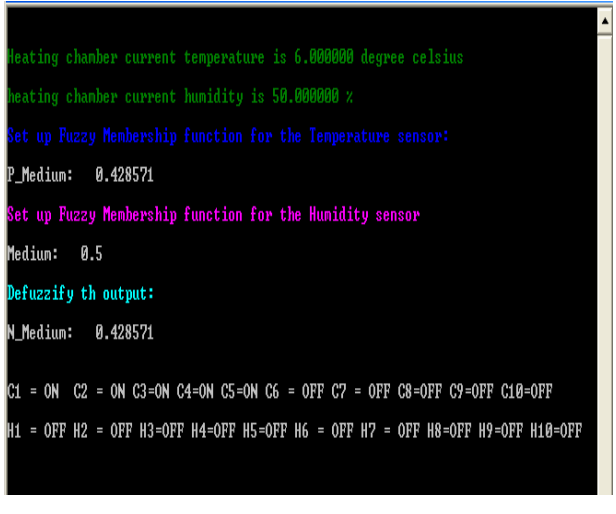


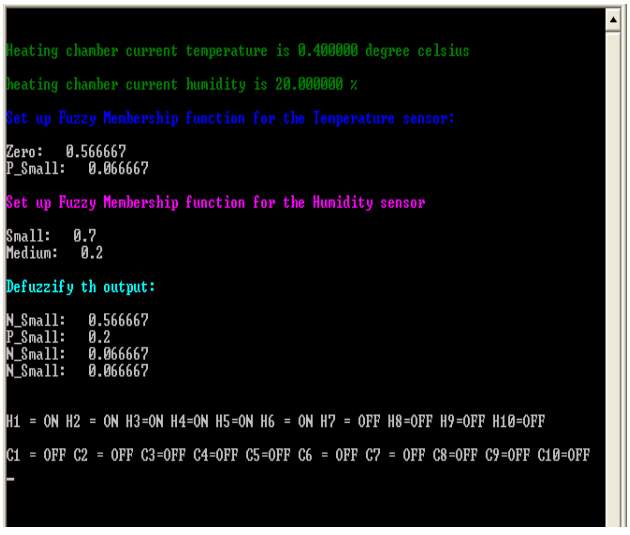
Fig. 9 Flow diagram for the control of the FLC

VI. Performance Analysis

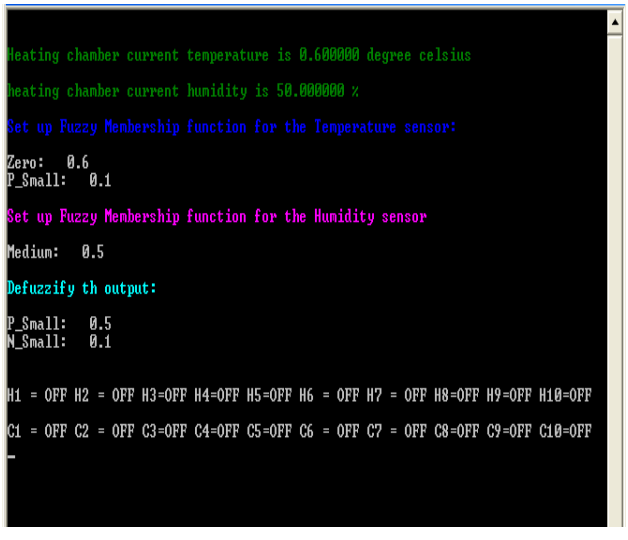
Figure 10 shows the excitation and responses of developed temperature monitoring and control using FLC. In this design, 10 cooling units and 10 heating unites are considered. These heating and cooling units are activated according to the inputs. The system is fully developed according to the previously discussed fuzzy logic design. In this fuzzy logic based temperature control system, we can take many factors into account when developing the controller. Some of these factors are considered to be noises in a conventional control system and make controller design a very hard work. Using fuzzy logic, it is much easier to design a controller with better performance in this type of temperature control problem.



(a) Some of the cooling units are on and all the heating units are off



(b) Some of the heating units are on and all the cooling units are off



(c) All the heating and cooling units are off

Fig. 10 The input/output responses of the FLC to control the temperature as outputs

VII. Conclusion and Observations

Unlike some fuzzy controllers with hundreds, or even thousands, of rules running on dedicated computer systems, a unique FLC using a small number of rules and highly improvement straightforward implementation has been done to solve a class of temperature control problems with unknown dynamics or variable time delays. Additionally, the FLC can be easily programmed into many currently available industrial process controllers. The FLC was simulated on the heating chamber temperature control problem with promising results. Then, it was applied to an entirely different industrial temperature apparatus. The results show significant improvement in maintaining performance and stability over the widely used PID design method. The FLC also exhibits robust performance for plants with significant variation in dynamics. The stability characteristics were investigated and a stability safeguard was derived also. The efficiency of this system can be increased and it will be the further improvement of this work. To increase the performance, several environmental variables such as space, pressure etc. can be used as inputs to the FLC for the heating chamber and industrial temperature control purpose.

References

[1] Ying, H., and B.-G. Hu: "Introduction to Fuzzy Control," International Journal of Fuzzy Systems, Vol 5, (2003) 87-88.

[2] Online Documentation: from http://www.austinlinks.com/Fuzzy/control. html.

[3] M.Khalid, S. Omatu & R. Yusof: “Temperature Regulation with Neural Networks and Alternative Control Schemes”, IEEE Trans. Neural Networks,vol. 6, no. 3., (1995), 572-582.

[4] K.S.Narendra & K. Parthasaraty: “Identification and Control of Dynamical Systems Using Neural Networks”, IEEE Transaction on Neural Networks, vol.1, No.1, (1990), 4-2.

[5] Y. Takahashi, C. S. Chan, and D. M. Auslander: "Parameter Tuning of Linear DDC Algorithms," ASME paper #70-WA/AUT-16, (1970).

[6] J.E. Marshall, *Control of Time Delay Systems,* Stevenage, UK; NY P. Peregrinus, c1979.

[7] Prokhorov, D., "Adaptive Critic Designs and their Application", Ph.D. Dissertation, Department of Electrical Engineering, Texas Tech University, (1997).

[8] Van Nostrand Reinhold, *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*, New York, (1992) 493 - 525.

[9] Yen, J. & R. Langari, *Fuzzy Logic*, Prentice Hall, Upper Saddle River, NJ, (1999).

[10] Ogata, K., *Modern Control Engineering,* Prentice Hall, London, (1997).

[11] Ying, H., "Structure and Stability Analysis of General Mamdani Fuzzy ynamic Models,” *International Journal of Intelligent Systems*, Vol 22, (2005) 103-125.

[12] Ying, H., "Conditions for Analytically Determining General Fuzzy Controllers of Mamdani Type to be Nonlinear, Piecewise Linear or Linear," Soft Computing,Vol 9, (2005) 606-616.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Cognitive-Based Teaching of Power Electronics

Muhammad H. Rashid, Ph.D., Fellow IET (UK), Fellow IEEE (USA)
Professor
Department of Electrical and Computer Engineering
University of West Florida
11000 University Parkway
Pensacola, Florida 32514-5754, USA
e-mail: mrashid@uwf.edu http://uwf.edu/mrashid

Abstract: Power electronics has evolved as a distinctive subject area and into many sub-specialties. There are so many important topics to cover within a limited time period of one semester. The questions raise what materials to cover, how to cover to develop and cultivate student's intellectual abilities for life-long learning knowledge and how to assess the outcomes achieved by the students. This presentation summarizes means of teaching and answering these questions.

1. INTRODUCTION

With the rapid advancements in technology and changes in the operations of business, the job functions of engineers are also changing. The engineering programs in USA are undergoing through rigorous changes in response to meeting needs of the new century. These changes are mandated by accreditation agency (Engineering Accreditation Commission of the Accreditation Board for Engineering and Technology, EAC/ABET) and the accreditation of an engineering program will be judged with respect to defined program outcomes. Each program must have an assessment process for continuous improvement with documented results. Any well thought course required for an engineering degree should be able to contribute towards fulfilling the educational program outcomes, which are mandated by the ABET criteria 2000 [1].

2 ENGINEERING ATTRIBUTES.

The Knowledge economy will have an impact on engineering and creates challenges and opportunities. It creates a global market place that will require averaging knowledge around the world through standardized quality engineering education and sharing of knowledge. It has created global opportunity and challenges for engineering education and for advancing quality of life worldwide participation by all nations/societies. Some of the desired attributes [2] of an engineer in the global marketplace in the new knowledge economy are as follow:

- Good understanding of engineering fundamentals and design/manufacturing processes.
- Multidisciplinary, systems perspective.
- Basic understanding of context engineering is practiced in.
- Good communication skills.
- High ethical standards.
- Ability to think critically/creatively, independently/cooperatively.
- Curiosity and desire to learn for life.
- Profound understanding of importance of teamwork.

The engineering education must adapt to the changing world and to the new forms of engineering. It must also accept new quality educational standards that are acceptable by industries around the world so that engineers can practice in the knowledge economy and are transferable to anywhere in the global market place with minimum amount of difficulties.

3. ROLE OF POWER ELECTRONICS

Power Electronics has already found an important place in the modern technology and it is now being used in great variety of high power products including heat controls, light controls, motor controls, power supplies, vehicle propulsion systems and HVDC.

Since power electronics is an interdisciplinary subject by nature, a good understanding of power electronics requires a strong background in math, fundamentals of power engineering, control and digital. This makes power electronics as a potential ideal course to develop student abilities at different intellectual levels for life-long learning and a successful career as an engineer and/or a business leader. A power electronic system is shown in Figure 1.

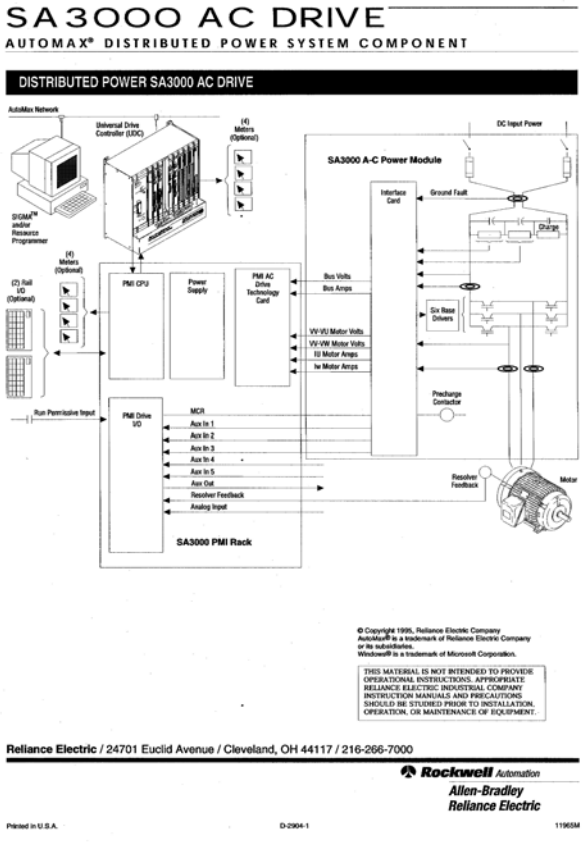


Figure 1 Power Electronic System

4. BLOOM'S TAXANOMY

At the 1948 Convention of the American Psychological Association, B S Bloom [3] had identified three "domains" of educational activities to achieve the "the goals of the educational process":

(1) Cognitive Domain,
(2) Affective Domain, and
(3) Psychomotor Domain.

The cognitive domain includes the recall of knowledge and cultivation of intellectual skills. The other two are not of major concerns for engineering. The affective domain describes interests, attitudes, and values. The psychomotor domain involves manipulative skills.

Certain cognitive processes such as problem-solving, critical thinking, reasoning, analysis and evaluations are particularly important in engineering tasks. Since most engineering methods involve some form of mathematics, this is a critical learning domain. In addition, engineering often involves innovation or invention; hence creativity is very important.

The taxonomy becomes a framework of analysis for educational outcomes. To date, most of the work in education has been in the cognitive domain. The taxonomy classifies the educational outcomes as a hierarchy of six major levels: knowledge, comprehension, application, analysis, synthesis, and evaluation. The objectives of one class should build on abilities. The comprehensive learning objectives of a course should include abilities at different intellectual levels as shown in Table 1. The level of cognitive domain is illustrated in Figure 2.

Table 1: Achievement of Bloom's Taxonomy of Educational Objectives in Cognitive Domain [2]

Cognitive Level	Educational Objectives	Learning Ability
# 1	Knowledge	List, recite
# 2	Comprehension	Explain, paraphrase
3	Application	Calculate, solve, determine
# 4	Analysis	Classify, predict, model, derive, interpret
# 5	Synthesis	Propose, create, invent, design, improve
# 6	Evaluation	Judge, select, critique, justify, optimize

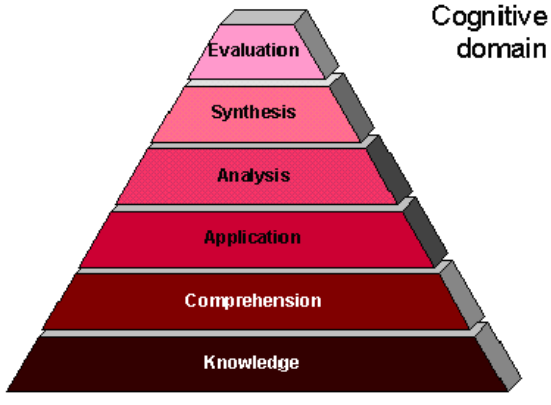


Figure 2: Level of Cognitive Domain [8]

5. APPLICATIONS OF BLOOM'S TAXANOMY

The teaching of power electronics can develop student's abilities at different cognitive levels. To illustrate the applications to power electronics, let us consider a diode rectifier [4,5] as shown in Figure 3 which is the most commonly used circuit in power electronics will be used an example to cultivate students learning abilities and assessments to demonstrate student mastery of the course material.

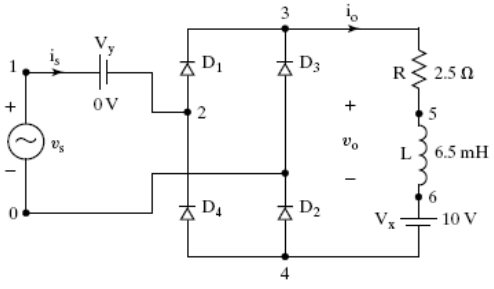


Figure 3: Level of Cognitive Domain

5.1 Knowledge

Can the students recognize the converter topology, identify the component layout and list the purposes of the circuit components?

The purpose of the inductor L is to limit
A. The average load current
B. The peak load current
C. None of the Above

5.2 Comprehension

Can the students explain the operation of the converter?

Increasing the load inductor L will
A. Increase the load current ripple
B. Decrease the load current ripple
C. None of the Above

5.3 Application

Can the students apply the circuit in applications, calculate the various currents and output voltages and predict the output?

If v_s = 170 sin (377t), the average load current will be
A. Between 25 to 35 A
B. Between 36 to 45 A
C. Between 46 to 55 A
D. None of the above

5.4 Analysis

Can the students identify and classify the performance parameters? Can the students model and predict the performance of the converter?

If v_s = 170 sin (377t) and Vx = 0 , the input power factor will be
A. PF = 0.50 to 0.75
B. PF = 0.76 to 0.85
C. PF = 0.86 to 0.95
D. None of the above

5.5 Synthesis

Can the students design a converter system? How can they improve the converter performance?

If v_s = 170 sin (377t), the minimum value of L to limit the load ripple to ±5% of the average load current
A. L = 24 - 35 mH
B. L = 36 - 45 mH
C. L = 46 – 55 mH
D. None of the above

5.6 Evaluation

Can the students identify the limitations, critique and optimize the converter components and system?

If R changes by ± 20%, _____to keep the same load ripple current
A. Change L by ±10%
B. Change L by
D. No effect of L

6. CONCLUSION

Power electronics should be able to develop critical thinking skills and prepare graduates who would have the ability for knowledge management such

- Ability to access the knowledge.
- Ability to locate the knowledge source.
- Ability to acquire and comprehend the knowledge.
- Ability to apply the knowledge in practical use.
- Ability to expand the knowledge into new knowledge and ideas leading to new innovation.
- Ability to share and communicate the knowledge.

It does not matter what we teach? Most course materials are likely to be obsolete in few years, even before the students graduate from a degree program. Then, why teach more? Rather, we should teach 'smart" and provide students with the necessary knowledge and skills to survive in the New Knowledge based-economy.

REFERENCES

1. "Criteria for Accrediting Engineering Programs", Engineering Accreditation Commission of the Accreditation Board for Engineering and Technology (EAC/ABET), 2005. http://www.abet.org/
2. Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain.* New York: David McKay Co Inc.

3. David O Swain, “Global Corporations Leveraging Knowledge”, ABET Annual Meeting, Incline Valley, Nevada, November 1, 2001.
4. M. H. Rashid (2003, 3/e), *Power Electronics - Circuits, Devices and Applications*. Prentice-Hall Inc.
5. M. H Rashid and H. M. Rashid (2005, 2/e), *SPICE for Power Electronics and Electric Power*. CRC Press.
6. M. H. Rashid (1999), Microelectronics, PWS Publishing.
7. M. H. Rashid (2003, 3/e), Introduction to PSpice Using OrCad for Circuits and Electronics, Prentice-Hall Inc.
8. R. Felder & R. Brent, *Designing and Redesigning Courses to Address EC2000, NCSU, 2001*

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Applications and Market Analysis of DC-DC Converters

S. D. Mitchell[2], S. M. Ncube[3], T. G. Owen[1], and M. H. Rashid, Fellow IEEE
Department of Electrical and Computer Engineering
University of West Florida
Pensacola, Florida 32514-5754, USA
e-mail: mrashid@uwf.edu
website: http://uwf.edu/mrashid

Abstract – This paper investigates the basic principles of operation of DC-DC converters and their applications. A market analysis was performed. This work was performed to determine the significance of DC-DC converters on the power electronic market. Research was further performed to focus on how buck-boost converters are used in everyday life along with their benefits and shortcomings when designed.

***Index Terms*--DC-DC power conversion, Road vehicle electric propulsion, Buck-boost Converter, Effective Efficiency.**

1. INTRODUCTION

The flexible capability to create an AC signal from a DC voltage, DC from AC, and DC from DC is all accomplished from the use of converters. Voltage conversion is an important element in power electronics. The area of interest for this paper involves the capability of DC to DC conversion. DC-DC converters provide the capability to either step up or step down a fixed DC voltage source. The converters can not only be used in devices such as automobiles and as power supplies, but they can be used to do things such as extend battery life for a considerable amount of time.

2. PRINCIPLES OF OPERATION

DC to DC conversion falls into two types, either step-down or step-up. Basic step-down operation is performed by a circuit represented in Figure 1. The converter steps down the DC voltage V_i to V_o. The component responsible for the step-down of the voltage is the switch, also referred to as a chopper. Any device that can be made to operate as a switch can perform the function of the chopper: bipolar junction transistor (BJT), metal oxide semiconductor field-effect transistor (MOSFET), insulated-gate bipolar transistor (IGBT), etc. The diode D1 is a freewheeling diode used for inductive loads [1].

The operation of the switch produces an average output voltage of

$$V_{o(avg)} = kV_i \qquad (1)$$

where k is the duty cycle of the opening and closing of the chopper and can vary from 0 to 1[1].

Basic step-up operation is performed by a circuit represented in Figure 2. The step-up is performed by transferring stored energy in the inductor to the load by opening and closing the chopper. The capacitor attached across the load helps to reduce the amount of ripple in the output voltage [1].

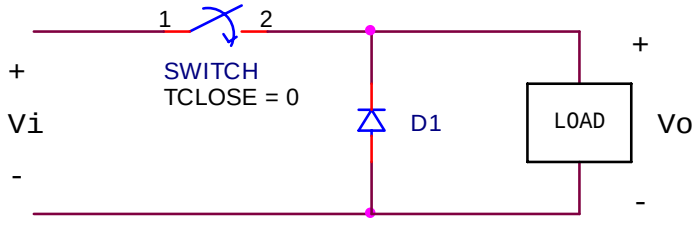


Fig. 1. Step-down converter.

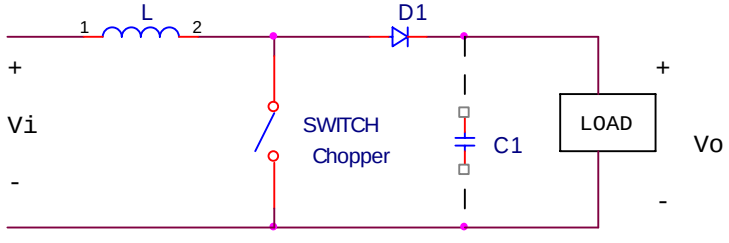


Fig. 2. Step-up converter.

The operation of the switch and inductor produces an average output voltage of

$$V_{o(avg)} = V_i \frac{1}{1-k} \qquad (2)$$

where k is the duty cycle of the opening and closing of the chopper and can vary from 0 to 1[1].

3. MARKET ANALYSIS AND PROSPECTUS

Many new applications for DC/DC converters have become available to developers in recent years. North American computer system designs and the rising market for

1,2,3 Undergraduate students

communications equipment in Asia are at the forefront of this market surge.

3.1 Prospectus

Rapid advancement and marketing in intermediate bus architecture (IBA) has increased the sales of non-isolated point-of-load (POL) and bus type DC/DC converters. Moreover, engineering trends toward reducing the operating voltage while increasing current and power handling capabilities of new electronics components have also had a major market impact. Currently, the North American computer market is accounting for the majority of DC/DC converter sales while Asian and Latin American demands for communications equipment are running a close second. Additionally, as the economies of nations around the world steadily recover,investments into new and/or replacement equipment has provided the required momentum to propel the DC/DC converter market to new levels of prosperity. Table 1 shows a representation of the Darnel Groups prospective for the growth in the DC/DC converter market at approximately 10.5% per year for the next 5 years [2].

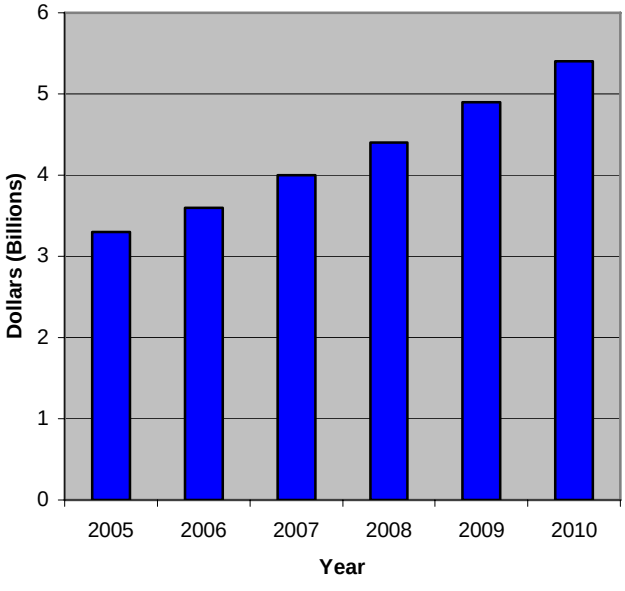


Table 1 Expected Market Growth of Dc-Dc Converters

3.2. Emerging Technologies

At this time, the server market is the fastest growing sector for DC/DC converters in both the communications and computer related markets. The current use of blade servers that typically employ either the distributed power architecture or the intermediate bus architecture are driving sales of bus converters, point of logic modules, and voltage regulators. Additionally, the telecommunications market is now gearing toward the use of these advanced IBA systems [3]. Other technologies such as the 3G, general packet radio service (GPRS), voice over Internet protocol (VoIP), and power over Ethernet (PoE) are also expected to increase the demand for DC/DC converters. [4]. Many of these new communications technologies are based on the Advanced Telecom Computing Architecture (ATCA) and will require extensive use of these converters [3].

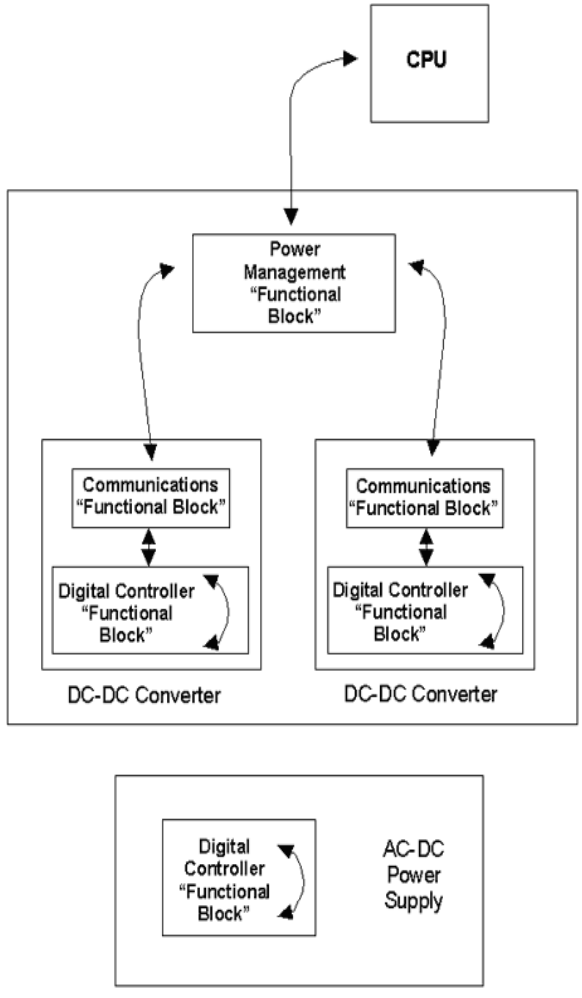


Figure 3. Large converter system

Another development that will see extensive use of DC/DC converters will be in the development of digitally controlled power supplies. "A digital dc-dc converter can be defined as one that is configurable and can be interrogated via a standard digital interface (e.g. a communications bus). At the system level, we would also include a "power management functional block" which conveys information from the system itself (e.g. the CPU) to the converter. In some cases, the converter "functional block" and the digital controller "functional block" could be a single, integrated chip. Figure 3 gives an example of a large system and where these functions would be implemented." [3]

"One issue that will become increasingly important is standards and communications protocols. The PMBus™ Initiative was announced at DPF, and it is gaining widespread support in the power supply community, including the Point-

of-Load Alliance (POLA) and the Distributed Power Open Standards Alliance (DOSA). In addition, digital power management solutions are still making use of different communications protocols, including I 2C and SMBus. Companies are introducing products that vary in their approach toward digital power management, and this could be one of the more controversial topics to emerge in the next year or two.

Darnell Group expects the competitive environment to change over the forecast period, with companies introducing either "revolutionary" products that address fundamental architecture levels, or "evolutionary" products that build on existing analog or hybrid solutions. Niche markets are likely to emerge, as well, with certain products addressing higher-end systems and others focusing on high-volume, lower-margin markets. Since these technologies are in their early phases, many companies are partnering to develop initiatives and products for future implementations of digital control. More of these partnerships are expected.

The successful adoption of digital power management, along with the introduction of commercial products, is dependent on (1) whether digital control will lead to system efficiency and improvement; and (2) the perceived value of digital control in OEM systems and power conversion. At that point, good business models will determine which companies reap the benefits of this potentially large opportunity." [3]

3.3. Consequences of Advancement

As with any new technology, many of these advancements come at the cost of replacing existing systems and thereby affecting the market. The current move toward improved digital power control systems and power conversion techniques may provide good opportunities for some, while posing a threat to others. One of the most significant examples is in the growing blade server market that is replacing the established independent enterprise servers [3].

However, the largest individual threat for the DC/DC converter market may actually come from the increased demand for these devices. The reason for this is due to the increased commoditization that will inherently reduce per unit cost of these devices. Standardization may also increase this effect. One estimate given by the Darnell Group indicates that prices for these converters are projected to decline at a rate of approximately 4.5% per year [3].

Additionally, "despite a strong dependence on the economy, there is a risk that demand may stagnate. To rise above this challenge, vendors must diversify geographically into new markets and develop technical innovations to create additional demand." [4] "Additionally, the business climate is also providing a number of challenges. Several recent alliances, partnerships, and acquisitions will continue to change the face of the dc-dc converter market. Several companies, such as C&D Technologies, have been able to gain substantial market share, while others have not been as fortunate. "[3]

"Emerging technologies may also have a very large say in how market share turns out in the somewhat near future. The ability to digitally configure output voltages, set voltage sequencing, the ability to monitor the dc-dc converter and so on not only offers technological advantages, but more importantly, economic advantages. Power-One claims that implementation of their digital solution can cut average development time from eight weeks to three days. Given the increasing importance of time-to-market and the costs associated with development, this benefit is significant."[3]

4. APPLICATIONS OF BUCK-BOOST REGULATORS

A buck-boost regulator can be used to help conserve power in hand-held devices run with lithium-ion-cell batteries. "A buck-boost regulator produces an output voltage that may be required to be greater than or less than the initial input voltage that was fed into the circuit. The reason this type of regulator is needed in many hand-held devices is that many of these device loads (such as digital ICs and memory) require voltage lower than the battery, while others (such as LED's used to display backlighting) need a higher voltage than the battery." [7]. Therefore, the need for a buck-boost regulator is needed as aforementioned so that the voltage can be boosted or lowered for whatever application is in use at that current time. A picture (Figure 4) below displays the circuit of the buck-boost regulator for inputs that need to be boosted above 3.3 V and also lowered below 3.3 V.

"When the voltage is above 3.3V, the IC stops switching and the in put voltage is stepped down to 3.3V by a linear regulator comprising Q1, R1, R2, R3 and an op-amp internal to the IC. When the input is below 3.3V, the IC operates as a step-up switching regulator and boosts the output to 3.3V. For this condition, the MOSFET is fully on, offering a virtual short from drain to source. [5]"

Since the regulator will be used to conserve power for future use, the efficiency of the process must be looked at also.

Figure 5 shows the efficiency of the circuit shown in Figure 4 varies as the load (output) current increases or decreases in value. "As expected, the efficiency is a minimum for battery voltage at its 4.2V peak as shown in Figure 5. For a 3.6V input and output, currents less than 500 mA have a numerical efficiency of 89%. This behavior is significant because the output of a lithium-ion cell is near 3.6V for most of its discharge cycle. For inputs ranging from 3.3 to 3.6V, the efficiency is even better. Efficiency is also excellent when the IC operates as a step-up switching converter, which it does for battery voltages which are below the critical 3.3V. [6]"

It is evident how much power a buck-boost converter can save a portable hand-held device just by increasing voltage above or below the critical voltage level of 3.3V. A good example of a buck-boost regulator chip that is very efficient is the LTC3440.

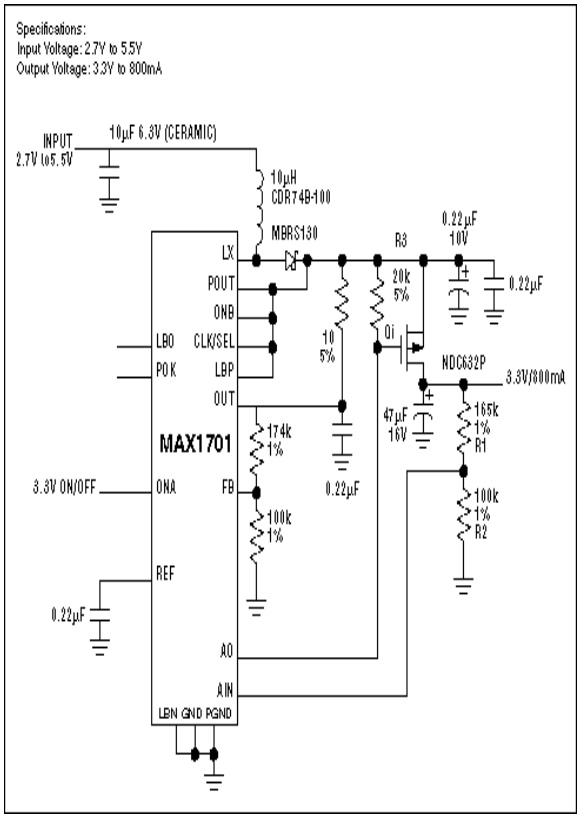


Figure 4: Buck-Boost Converter in Common Handheld Devices [5]

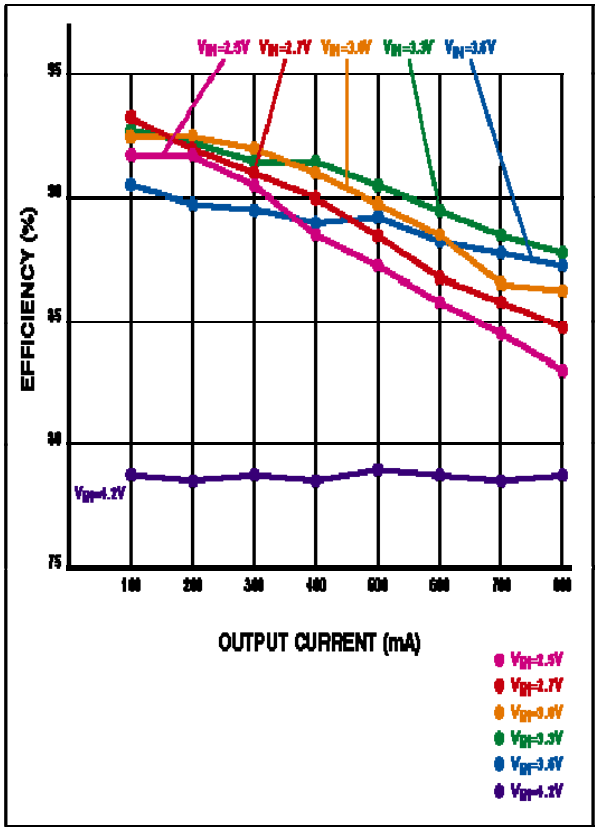


Figure 5: Circuit Efficiency vs Change in Output Current [5]

"The chip is based off of a full-bridge (4 switch) architecture using a unique control scheme. The dc /dc converter switches from step-down to step-up operation as the battery voltage changes, and delivers 94 percent efficiency across the entire battery voltage range. Because all four MOSFETs are included in the MSOP package, and switching frequency may be user-adjusted up to 2MHz. A complete 3.3V, 600mA power supply consumes only 150mm2 of pcb area. The LTC3440 can be pin-selected for continuous switching or Burst Mode operation where quiescent current drops to 25µA. [6]"

Burst mode operation is the process that the buck-boost regulator will undergo when the hand-held device is really not in use. Basically, this means that the current needed to keep the circuit barely running should be very low. This poses a problem at first because even though the circuit may not be in use, the efficiency is still very important. "In burst mode operation (a "keep alive" function for a switching regulator) reduces quiescent current to only 20µA under light load. Also desirable is an efficiency remaining up 87 percent even with only 1mA load current."

However, burst mode operation increases the amount of noise that the buck-boost circuit will produce; mainly, noise that interferes with cellular phones and apparatuses of that nature. "Selecting continuous switching or synchronization inhibits burst mode operation, and eliminates audio noise generation, but again, at the expense of light load efficiency. [6]" Hence, as always in engineering, there is always a trade-off of one thing benefit for another.

A paper published by Schuessler [8] says, "In many cases it is cost prohibitive or difficult to find a power supply to accept the full range of input voltage, but if the entire extent of input voltage from the battery could be used, device life would certainly be maximized. Effective Efficiency produces a single efficiency number taking into account any unused energy left in the battery. [8]" If the range of the input voltage can somehow be extended then more power can be preserved for future use and power use would be used much more efficiently.

"Only now becoming available, a buck-boost switched capacitor regulator used multiple fractional gains in a switching matrix controlled by a DSP-like digital control loop. As input voltage or output load changes, the controller dynamically changes the capacitor conversion ratios to maintain optimal frequency. This results in a saw-tooth type efficiency curve as gains are changed over the input voltage range. [8]"

Just recently (as in the last year) some chips have been designed that actually use lower input voltages and in general have a wider input voltage range from approximately 1.5V to 5.7V. An example of a chip that is in use is "the SiP1759 chip which has a stable 3.3V fixed output voltage with a 1.6V to 5.5V input voltage range. This device is used to power handheld devices such as PDA's, cellular phones and many other devices. [9]" With a lower input voltage range the maximum battery life can be sustained as low as 1.6V according to the preceding sentence. "The regulator offers 100mA guaranteed output current, and it has a 1.5MHz operating frequency to help reduce the capacitor size. [9]"

Another example of a chip that has a wide range for input voltages is the "Elantec EL7515 (Figure 6) which becomes a buck-boost converter when a PFET (Q1) and a diode (D2) are added to the circuit. This chip is mostly helpful with USB applications because when these devices are in use their input voltage varies although the standard voltage is considered to be 5V. However, the input voltage in reality varies from 4.5 to 5.25V which is why a buck-boost converter is needed in order to produce stable operation. [10]"

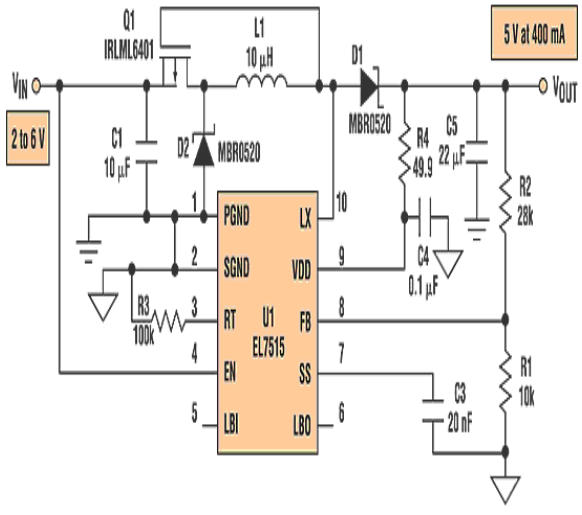


Figure 6: Buck-Boost Converter used in USB Applications

This buck-boost converter used in USB applications allows for a wide range of input voltages which allow it to be used in higher voltage applications. The maximum output voltage that the converter produces is 17 V.

These are some of the many applications that buck-boost converters have not only in power electronic circuit, but also in our daily lives.

ACKNOWLEDGMENT

The work was prepared by S. D. Mitchell, S. M. Ncube, T. G. Owen for the course EEL 4242 – Power Electronics under the guidance of Dr. Muhammad H. Rashid. There was no financial support attributed to the creation of this document. They have developed the content and information in this report using original material and the information cited in the references.

REFERENCES

[1] M. H. Rashid, *Power Electronics: Circuits, Devices, and Applications*, 3rd ed. Upper Saddle River, New Jersey: Pearson Prentice Hall, 2004.

[2] DC/DC Converters: Global Market Forecasts, Low-Voltage Powering Trends and Competitive Environment. (2005). Retrieved June 19, 2005, from the Darnell.com web site at: http://www.darnell.com/services/01-dcmkt.stm.

[3] Emerging Markets in Digital Power Electronics: Component, Converter and System Level Opportunities. (2005). Retrieved June 19, 2005, from the Darnell.com web site at: http://www.darnell.com/services/01-digital.stm.

[4] Frost & Sullivan. (2005). World DC-DC Converter Markets. Retrieved June 19, 2005, from the Global Information, Inc. web site at: http://www.the-infoshop.com/study/fs25122_world_dcdc.html.

[5] Author Unknown, 11/01/01 "Lithium-Ion-Cell Buck/Boost Supply Requires One Inductor,". 6/15/05 <http://www.maxim-ic.com/appnotes.cfm.appnote_number/438>

[6] Bell, Dave, 07/09/02 "Micropower dc-dc converters promote battery life in handhelds,". 6/15/05 <http://www.eet.com/in_focus/mixed_signals/OEG20020709S0021>

[7] Williams, Richard, 4/14/05 "Features Drive Supply Complexity In Handhelds,". 6/15/05 < http://www.elecdesign.com/Articles/Index.cfm?ArticleID=10011m>

[8] J.E.Schuessler. (1999, Feb.). Efficiency and Power Utilization Data Guide DC/DC Conversion Choices in Battery Operated Devices. Grass Valley Design Center., Grass Valley, CA. [Online]. Available: <http://www.national.com/appinfo/power/files/WPbD-02a.pdf>.

[9] Author Unknown, 08/03/04 "3.3 Buck-Boost Regulator-has compact 4.9 x 3.0 mm footprint,". 6/16/05 < http://www.powermanagementdesignline.com/products/51200011>

[10] Yang, Liu, 03/29/04 "Simple Buck-Boost Converter Shines in USB Applications,". 6/19/05 < http://www.elecdesign.com/Articles/ArticleID/7657/7657.html >

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Performance Analysis of a MIMO-OFDM Wireless Communication System With Convolutional Coding

*Kazi Mostaq Ahmed ** Satya Prasad Majumder
*Lecturer , Department of Electrical and Electronic Engineering,
Stamford University Bangladesh
E-mail : engrkma@gmail.com
**Professor , Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering & Technology (BUET)
E-mail : spmajumder2002@yahoo.com

Abstract- The bit error rate (BER) performance analysis of an OFDM wireless communication system is carried out considering the effect of fading and multipath dispersion. Analysis is extended to include multiple-input and multiple output (MIMO) transceiver configuration to combat the effect of fading. Further, forward error correction coding such as Convolutional Coding (CC) is also applied to improve the BER of a MIMO-OFDM system. Significant improvement in receiver performance is obtained by numerical computation. The analytical results are found to be in good agreement with the simulation results reported earlier.

Key words : OFDM, Fading, ICI, Convolution-Code and MIMO.

1. Introduction

The application of multiple antennas at both transmitter (TX) and receiver (RX) side of wireless communication systems is proposed in many contributions over the last few years. This systems is widely known as multiple-input multiple-out (MIMO), which provides the benefit of increased range, robustness and improved data rate. When applying this techniques to wideband communication, the combination of the MIMO architectures with the multicarrier techniques to every subcarrier, separately. Research concerning MIMO-OFDM based systems mainly focuses on systems impaired by additive white Gaussian receiver noise and spatial correlated channels. When impairments will arise, which can largely influence the performance of the wireless system.

Several research work has been carried out on the performance evaluation of an OFDM system both analytically and also by simulations [1]-[6]. The BER performance results for OFDM system over fading channels are reported in [1]-[4]. The performance of non-coherently detected BFSK/OFDM over multipath fading channels with noise is investigated in [1]. This analysis demonstrate that it is possible to transmit information in selective channels with no symbol interference. Expression of the Bit Error Probability(BEP) are derived in the context of frequency selective Rician fading channels with and without convolutional coding. Reference [2] proposes an approximate derivation method of the bit error rate (BER) in DQPSK/OFDM systems over frequency non-selective Nakagami-Rice and Rayleigh fading channels. Performance of OFDM has been evaluated in fading channels exhibiting both time-selectivity and frequency-selectivity in [3]. Reference [4] proposed a simple method, that is , an approximation closed-form equation of the bit error rate (BER) in DPSK/OFDM systems mentioned above over both time and frequency selective Rician fading channels. Reference [5] shows the impact of ICI Cancelling Space-Frequency Block Code for MISO-OFDM over the Fast Fading Channels. Performance evaluation of a multi-antenna OFDM systems in fading channels with additive transmitter and receiver impairments has been carried out in Reference [6]. BER of MIMO-OFDM Systems with Carrier Frequency Offset and Channel Estimation Errors discussed in reference [7]. Reference [9] shows the Inter carrier interference in MIMO-OFDM system. In this research work, we evaluate the performance from a approximate closed-form equation of the unconditional BER of MIMO-OFDM systems over time selective fading channels with convolutional coding.

The remainder of this paper is organized as follows. Section- 2 derives the system model for a multiple antenna system applying OFDM. The probability of bit error for a MIMO-OFDM system in AWGN and fading channels for both coded and un-coded system are derived in Section-3. Section-4 then provides computational results. Finally, conclusions are drawn in Section-5.

2. System Model

Fig-1 shows a block diagram of a MIMO-OFDM wireless communication system. In this paper we have considered two receiving antennas and one transmitting antenna. In a modulation block, the followings are carried out to a binary data sequence : Serial to Parallel conversion, DPSK modulation with differential coding in the time or frequency domain.

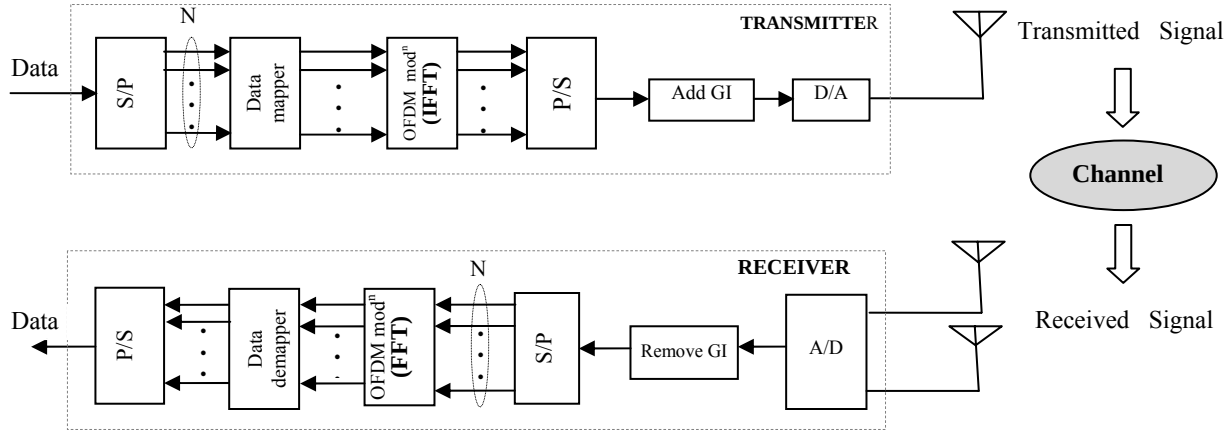


Fig-1 : Block diagram of a MIMO-OFDM System

OFDM modulation using IFFT, parallel-to-serial conversion and addition of the GI. It is assumed that the ISI can be completely avoided with the GI which is longer than the maximum multipath delay. In a transmission block, an OFDM signal is subjected to the fading and is added the white Gaussian noise (AWGN). In a demodulation block the desired binary data sequence is demodulated from the received OFDM signal through a reverse process of the modulation block.

3. Theory and Discussion

A. Effect of Inter-carrier Interference :

The time and frequency fluctuations give rise to not only the BER degradation but also the interference to the other symbols of all subscribers, which degrades the BER furthermore in OFDM systems. We will analyze the interference in the following subsections, which is based on the concept of complex Fourier series.

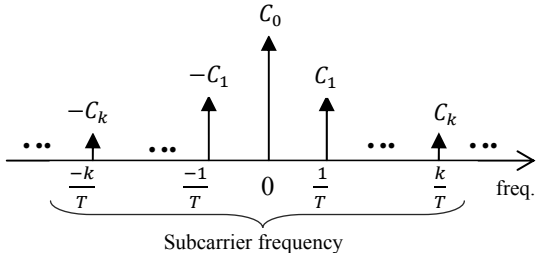


Fig-2 Complex Fourier coefficients.

Fig.2 illustrates the complex Fourier coefficients C_k which is given by :

$$C_k = \frac{1}{T}\int_0^T f(x)e^{-jk\frac{2\pi}{T}x}dx \qquad (1)$$

When a carrier $\cos(2\pi f_c t)$ passes through the fading channels, the received carrier frequency is affected by Doppler frequency $f_D\cos\theta$ where θ is a random phase with the uniform distribution of (0, 2π) . Then the received carrier r(t) can be expressed as [4]

$$r(t) = cos\{2\pi(f_c + f_D cos\theta)t\}$$
$$\cong \cos(2\pi f_c t) - (2\pi f_D t\, cos\theta)\sin(2\pi f_c t) \qquad (2)$$
$$for\ |2\pi f_D t\, cos\theta| \ll 1,$$

From (1) the interference power to a subcarrier $k(k \neq 0)$ can be obtained as :

$$|C_k|^2 = \left|\frac{1}{T_d}\int_0^{T_d} 2\pi f_D t\, cos\theta e^{-jk\frac{2\pi}{T_d}t}\, dt\right|^2$$
$$= \frac{(f_D t\, cos\theta)^2}{k^2}\ for\ k \neq 0 \qquad (3)$$

The total power of ICI from subcarriers $k(-k_l \leq k \leq k_l, k \neq 0)$ can be calculated as :

$$I_a^{f_D} = 2\sum_{k=1}^{k_l}|C_k|^2$$
$$= \frac{(\pi f_D T_d\, cos\theta)^2}{3}, k_l \gg 1 \qquad (4)$$

Finally, the average power I_a of ICI caused by the Doppler frequency shift can be derived by averaging (4) by uniformly distributed θ as follows :

$$I_a = \frac{1}{2\pi}\int_0^{2\pi}\frac{(\pi f_D T_d\, cos\theta)^2}{3}d\theta$$
$$= \frac{(\pi f_D T_d)^2}{6}$$
$$= \frac{(\pi f_D T_s)^2}{6(1+\alpha)^2} \qquad (5)$$

B. Effect of Intra-Symbol Interference

Let, ψ is a phase rotation angle of the demodulating symbol Z_d caused by the fading . The differential detection in DMT

is carried out by the phase difference Z_r and Z_d which have a time interval of T_s in the same subcarrier (figure-3). Then the ψ in terms of time can be expressed as :

$$\psi(t) = 2\pi f_D t\, cos\theta, \quad \text{for DMT} \tag{6}$$

Which is caused by the Doppler frequency shift.

Then the received carrier $r(t)$ with the phase rotation ψ become: $r(t) = cos(2\pi f_c t + \psi)$

$$= cos\psi \cos(2\pi f_c t) - sin\psi sin(2\pi f_c t) \tag{7}$$

Assuming $|2\pi f_D t\, cos\theta| \ll 1$, $cos\psi$ and $sin\psi$ in (7) can be approximated as:

$$cos\psi = \cos(2\pi f_D t\, cos\theta) \cong 1 \tag{8}$$

$$sin\psi = sin(2\pi f_D t\, cos\theta) = 2\pi f_D t\, cos\theta \tag{9}$$

The interference power $I_d^{f_D}$ and its average power $\overline{I_d^{f_D}}$ caused by the Doppler frequency shift become :

$$I_d^{f_D} = \left(\frac{1}{T_s}\int_0^{T_s} 2\pi f_D t\, cos\theta\, d\theta\right)^2 = (\pi f_D T_s cos\theta)^2$$

$$\overline{I_d^{f_D}} = \frac{1}{2\pi}\int_0^{2\pi}(\pi f_D T_s\, cos\theta)^2\, d\theta = \frac{(\pi f_D T_s)^2}{2}, for\ DMT \tag{10}$$

Finally, the average interference power I_d to the quadrature channel component can be expressed as

$$I_d = \frac{(\pi f_D T_s)^2}{2}, for\ DMT\ in\ DQPSK$$

C. Carrier-to-Noise plus Interference (CNIR) Power Ratio:

The CNR, γ_N can be calculated from the received signal (Rician wave) as follows:

$$CNR,\ \gamma_N = \frac{2\frac{\Gamma_{EN}}{1+\alpha}}{k+1} = \frac{2\Gamma_{EN}}{(1+k)(1+\alpha)}, for\ DQPSK \tag{11}$$

Where ,$\frac{1}{1+\alpha}$ means the energy loss caused by the removal of the GI . By regarding the interference power I_a and I_d as the increase of disturbance powers besides the AWGN against the signal power, the CNIR of the received Rayleigh wave γ_{NI} can be expressed as :

$$CNIR, \gamma_{NI} = \frac{P_S}{P_n + P_I} = \frac{1}{\frac{P_n}{P_S} + \frac{P_I}{P_S}} = \frac{1}{\frac{1}{\gamma_N} + I_a + I_d}$$

$$= \frac{1}{\frac{(1+k)(1+\alpha)}{2\Gamma_{EN}} + I_a + I_d}$$

$$= \frac{1}{\frac{(1+k)(1+\alpha)}{2\Gamma_{EN}} + \frac{(\pi f_D T_s)^2}{6(1+\alpha)^2} + \frac{(\pi f_D T_s)^2}{2}} \tag{12}$$

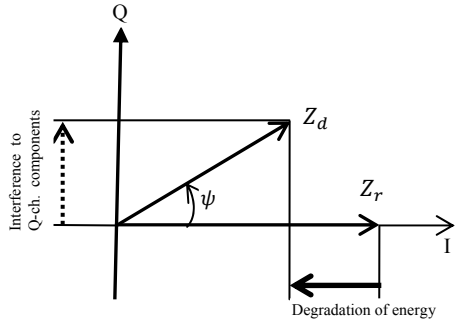


Fig-3 Intra-symbol Interference.

D. Expression for BER

Considering both Rician and Rayleigh distribution, the BER of the system can be expressed as follows [4] :

$$P_b = \frac{(1-\rho)\frac{CNIR}{2} + 1}{2\left(\frac{CNIR}{2} + 1\right)} exp\left(\frac{K\frac{CNIR}{2}}{\frac{CNIR}{2} + 1}\right) \tag{13}$$

Where, $\rho = J(2\pi f_D T_S)$ is the time correlation function and K= $\frac{A}{2{\sigma_s}^2}$ is the Rician K-factor. This method can simply calculate the BER from the values of the maximum Doppler frequency normalized by the symbol frequency $f_D T_s$.

E. Expression for BER with Convolutional Coding

Applying Convolutional Coding the transmission performance can be significantly improved. In presence of channel coding, an expression of the bit error probability P_e cannot be worked out exactly, showing the need for a good upper bound. It is well known that using a rate $R = \frac{K}{N}$ convolutional coding and a Viterbi algorithm decoding, the bit error probability for an information symbol is bounded by :

$$P_{coded} = \frac{1}{k}\sum_{d=d_f}^{\infty} W(d)P(d) \tag{14}$$

Where $P(d)$is the probability for the decoding algorithm to choose a path at a distance d from the correct path in the decoding trellis, d_f is the free distance of the encoder and W(d), a characteristic coefficient of the encoder, defined by:

$$W(d) = \sum_{i=1}^{\infty} ia(di)$$

Where a(di) is the number of paths at distance d from the correct path and corresponding to i information symbols equal '1'. In general, the $a(di)$ are deduced from the transfer function of the encoder. Now $P(d)$ can be expressed as follows :

$$P(d) = [4P_{uncoded}(1 - P_{uncoded})]^{1/2} \tag{15}$$

F. Analysis of MIMO-OFDM system

Maximal ratio combining method relies on the complex channel gains, so the weights re chosen as $w_n = h_n^*$ . . In this case, post-combining signal reads as :

$$z = \sqrt{E_s}\ |h|^2\, c + n' \qquad (16)$$

Where, E_s = transmitted symbol energy, $h = [h_1,\ h_2,\ h_3 \ldots\ldots\ldots\ldots h_{n_r}]^T$ and $n' = h^H n$. This scheme is known as maximal ration combining (MRC), for the reason that it maximizes the output SNR ρ_{out} . The latter is indeed equal to :

$$\rho_{out} = \frac{1}{\sigma_n{}^2}\sum\left\{\frac{E_s|h|^4}{|h|^2}\right\}$$

$$\rho_{out} = \rho\sum|h|^2 \qquad (17)$$

Denoting $u = |h|^2$, it is well-known that u follows a χ^2 (chi-square) distribution. Now the probability of conditional BER can be expressed as :

$$BER(u) = \frac{(1-\rho)\frac{(CNIR*u)}{2}+1}{2\left\{\frac{(CNIR*u)}{2}+1\right\}} exp\left\{\frac{k\frac{(CNIR*u)}{2}}{\frac{(CNIR*u)}{2}+1}\right\} \qquad (18)$$

When the different channels are i.i.d Rayleigh distributed, the pdf of u can be expressed as :

$$P_u(u) = \frac{1}{(n_r-1)!}u^{n_r-1}e^{-u} \qquad (19)$$

Where, n_r = 1, 2, 3 , ……. = no. of receiving antenna. The expression for the Probability of BER of a MIMO-OFDM System over fading channel can be expressed as :

$$P_{mimo} = \int_0^\infty BER(u).P(u)\,du \qquad (20)$$

4. Result and Discussion

Following the analytical approach given in section- 3, we evaluate the bit error rate (BER) performance of a MIMO-OFDM system for several values of f_DT_s . It is revealed from the Figures 4 and 5 that the system performs better with comparatively lower values of f_DT_s . BER is about 10^{-6} at GI-factor, α=0.25, SNR=40 dB and f_DT_s=0.001 for Rician factor- K=4dB. Whereas, BER is measured as 10^{-7} at K=6dB. The performance of this transmission system has been evaluated for various values of K. Figure-5 shows the BER performance of a Convolutionally-Coded MIMO-OFDM versus SNR plots for different values of f_DT_s at K=0 dB and Constraint length, L=9 (Rate=1/2) . From the computational result, it is observed that BER is about 10^{-30} and 10^{-34} at SNR=40dB for K=0 dB and 6dB respectively. As compared with the single antenna OFDM system the values are respectively 10^{-24} and 10^{-26} .

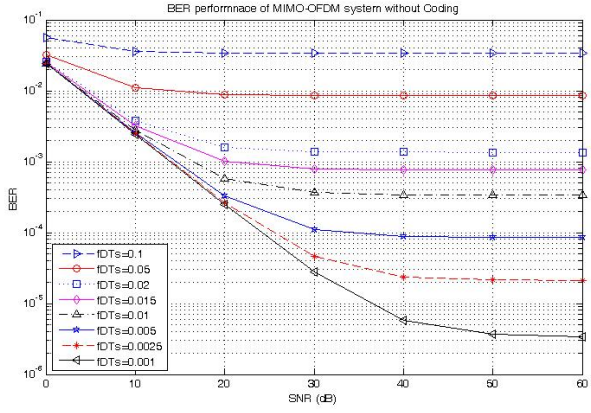


Figure-4 : BER performance of Un-coded MIMO-OFDM versus SNR for different values of f_DT_s at K=0dB with one transmitting antenna and two receiving antenna.

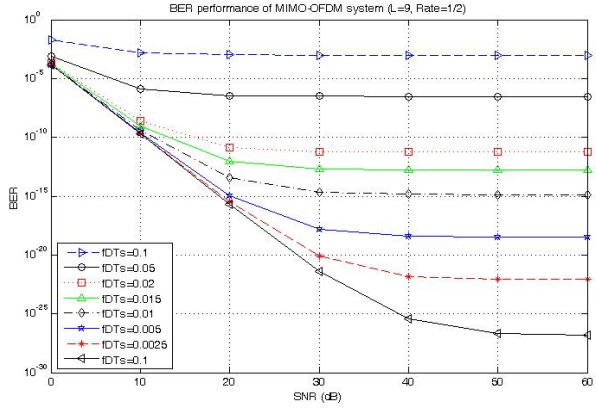


Figure-5 : BER performance of Coded MIMO-OFDM versus SNR for different values of f_DT_s at K=0 dB and Constraint Length, L=9 (Coding Rate=1/2) with one transmitting antenna and two receiving antenna.

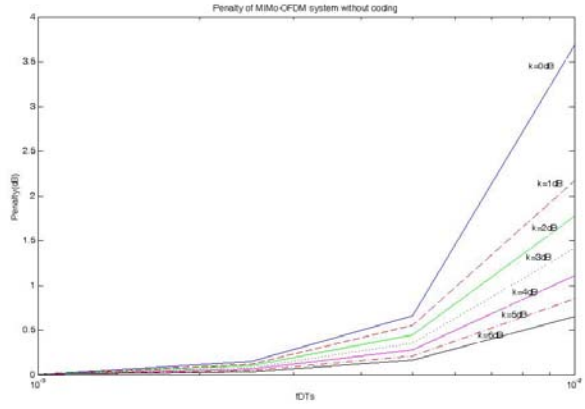


Figure-6 : Power-Penalty versus f_DT_s of an Uncoded MIMO-OFDM system for different values of Rician K-factor.

Figure-6 illustrates the power penalty versus the $f_D T_s$ with K as a parameter for an uncoded MIMO-OFDM system. It is remarkable from the plot that the penalty is about 3.7dB, 2.20dB, 1.76 dB, 1.40 dB, 1.20 dB, 0.8 dB and 0.065 dB for K=0dB, 1dB, 2dB, 3dB, 4dB, 5 dB and 6dB respectively at a BER of 10^{-3} . As compared with the penalty for both coded and uncoded OFDM system, it is observed that the OFDM with Multi-antenna system suffers significantly less amount of power penalty. Figure-7 shows the BER performance of a coded MIMO-OFDM system considering several no. of receiving antenna (i.e. Rx=2, 3, 4, 5, and 6). It is noticed that there is remarkable improvement in the achievable probability of bit error as number of receiving antenna increase. The BER is measured as 10^{-12} for a un-coded MIMO-OFDM at SNR=40 dB, fDTs=0.001 and number of receiving antenna, Rx=6 . In case of a coded system the BER occurs at about 10^{-56}, which indicates substantially much more higher performance. We also evaluate the performance of BER-Floor as a function of Rician K-factor for different $f_D T_s$ values for both coded and un-coded system. This is noticeable from figure-8 that the system is robust against lower $f_D T_s$ and whereas at higher $f_D T_s$ the BER performance degrades. Unlike the OFDM system, the MIMO-OFDM system is substantially dominated by the Rician K-factor and BER can be reduced by increasing the K-factor. We also verify the BER-floor versus $f_D T_s$ plots for both coded (figure-9) and un-coded system. Again, it is noticed that a higher K and a lower $f_D T_s$ provide better ICI cancellation and BER performance. It is revealed that the BER-Floor occurs at about 10^{-6} for un-coded MIMO-OFDM system at SNR=40 dB with $f_D T_s$=0.001 and K=2 dB. It is noticeable that for the same parameters the coded-MIMO-OFDM provides improved performance and results in BER-floor occurs at about 10^{-24} for L=7, rate=1/2. Figure-10 illustrates the comparison of the performance between a coded and un-coded MIMO-OFDM system as a function of number of receiving antenna (Rx.) . The BER performance improves as the number of receiving antenna increase.

The power-penalty versus number of receiving antenna plots for k=0, 2, 4, and 6 dB respectively is shown in figure-11. It is noticed that using two receiving antenna (Rx=2) penalty is 33.18 dB , 31.67 dB , 29.41 dB and 25.07dB for K=0, 2, 4 and 6 dB respectively at fDTs=0.001 and BER=10^{-5}. On the other hand, increasing the no. of receiving antenna (Rx=3) the penalty is measured as 16.06 dB, 15.69 dB, 14.68 dB and 10.94 dB respectively. It should be mentioned that the penalty decreases as the number of receiver antenna (Rx.) increases and obviously penalty is less for higher values of K. Performance obtained for 1-Tx and 2-Rx antenna as shown in figure-12 along with the performance reported in Ref. [11] for 2-Tx And 2-Rx antennas. It is observed that there is no significant deviation in the results.

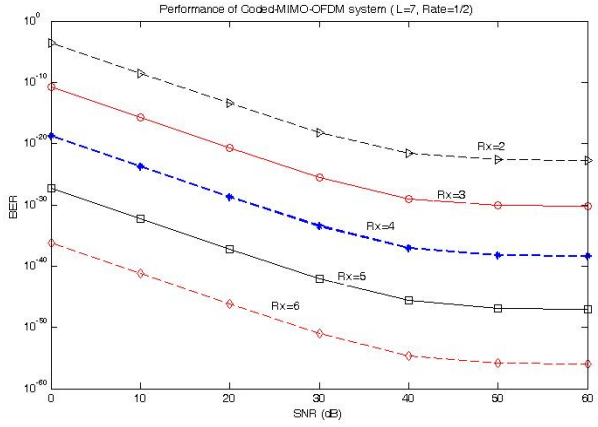


Figure-7 : BER versus SNR of a Coded MIMO-OFDM system for different no. of receiving antenna at L=7 , $f_D T_s$ =0.001 and K=0 dB.

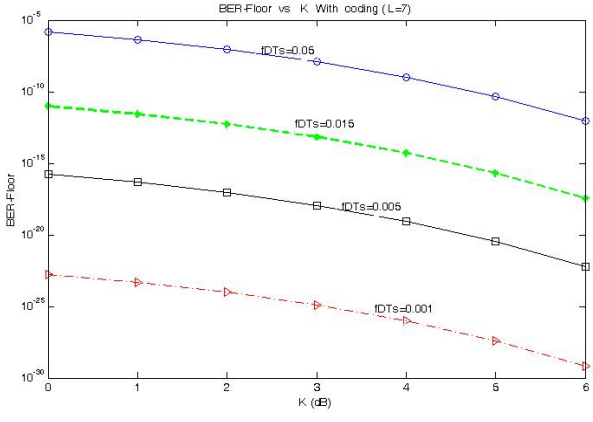


Figure-8 : BER versus K of an Coded MIMO-OFDM system for $f_D T_s = 0.001,\ 0.005,\ 0.015,\ and\ 0.05$ respectively and at SNR=40dB.

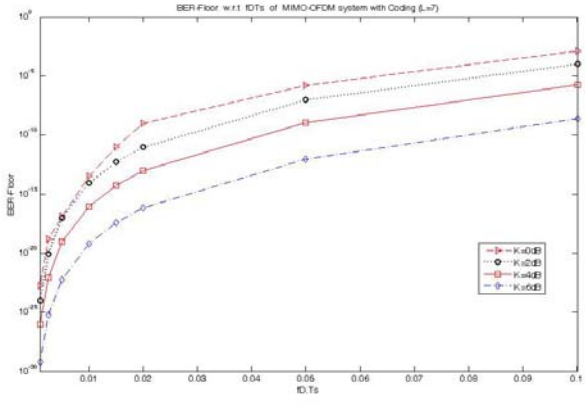


Figure-9 : BER-Floor versus $f_D T_s$ plots of a Coded-MIMO-OFDM system for L=7, rate =1/2, and K=0, 2, 4 and 6 dB respectively.

5. Conclusion

Expressions for the probability of bit error of a MIMO-OFDM wireless communication system with Inter carrier interference (ICI) and over fading channels are derived. Unlike the OFDM with single antenna system, it is revealed that the MIMO-OFDM system performs better with comparatively lower values of normalized Doppler frequency $f_D T_s$ with higher Rician K-factor, i.e., the system is substantially dominated by the K-factor and $f_D T_s$. It should be mentioned that the power penalty decreases as the number of receiving antenna increases. Also, from the numerical results it is observed that the performance of the MIMO-OFDM wireless communication system over fading channels can be significantly improved with Convolutional Coding. The analytical results are found to be in good agreement with the simulation results reported earlier.

6. References

[1] Alain Glavieux, Pierre Y. Cochet and Annie Picart, "Orthogonal Frequency Division Multiplexing with BFSK Modulation in Frequency Selective Rayleigh and Rician Fading Channels" IEEE Transaction on Communication, Vol-42, No. 2/3/4, February/March/April 1994.

[2] Steendam, H., Moeneclaey, M., "Analysis and Optimization of the performance of OFDM on Frequency-Selective Time-Selective Fading Channels", IEEE Transaction on Communication, Vol-47, No.12 December 1999.

[3] Bulumulla, S. B. Kassam S.A. and Venkatesh, S. S., "A Systematic Approach to Detecting OFDM Signals in a Fading Channels" , IEEE Transaction on Communication, Vol-48, No. 48 May 2000.

[4] Fumihito Sasamori, Shiro Handa and Shinjiro Oshita, "A Simple Method of BER Calculation in DPSK/OFDM Systems over Fading Channels", IEICE Transaction Fundamentals, Vol.E88-A, No-1, page 366-373, January 2005.

[5] Jae Yeun, Eui-Rim Jeong, Jong Guk Ahn , Sae-Young Chung and Yong H. Lee. , " ICI Cancelling Space-Frequency Block Code for MISO-OFDM in Fast FadingChannels, IEEE GLOBECOM 2007.

[6] Tim C. W. Schenk, Peter F.M. Smulders and Erik R. Fledderus "Performance of MIMO-OFDM systems in fading channels with additive TX and RX Impairments", IEEE BENELUX/DSP vally signal Processing Symposium, proceedings of SPS-DARTS, page 41-44, 2005.

[7] Zhongshan Zhang, Wei Zhang and Chinta Tellambura, "BER of MIMO-OFDM Systems with Carrier Frequency Offset and Channel Estimation Errors" ICC, 2007 proceedings , page 5473.

[8] G.J. Stuber, J. R. Barry, S. W. Mclanghlin, Y. Li, M. A. Ingram, and T. G. Pratt, Broadband MIMO-OFDM wireless communication,"Proc.IEEE, vol. 92, no. 2 pp. 271-294, Feb-04.

[9] A. Stamoulis , S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO-OFDM." IEEE Trans. Signal Process., vol. 50, no.10 pp. 2451-2464, oct. 2002.

[10] Alamouti S.,S. M., "A simple Transmit Diversity Technique for wireless Communications .",IEEE Journals on selective areas in communications, Vol-16., No. 8, October 1998.

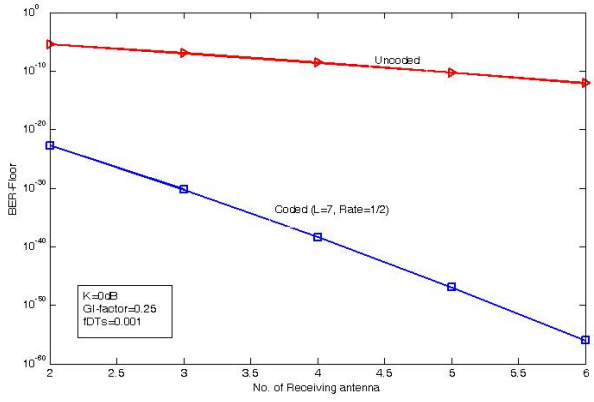


Figure-10 : Comparison of BER-Floor versus no. of receiving antenna between Coded and Un-coded MIMO-OFDM system at $f_D T_s = 0.001$, GI-factor=0.25 and K=0 dB.

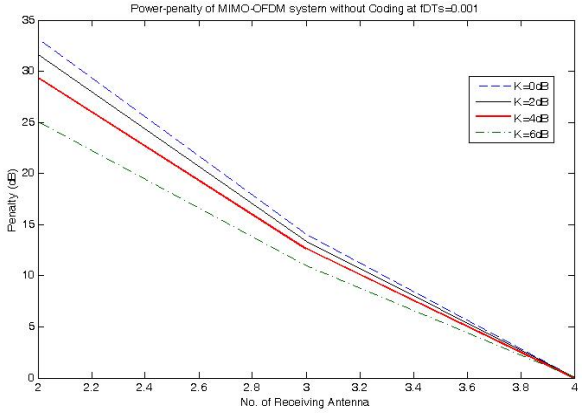


Figure-11 : Power-penalty with respect to number of Receiving antenna for an Un-coded MIMO-OFDM system at $f_D T_s = 0.001$, GI-factor=0.25, BER=10^{-5} and K=0, 2, 4, and 6 dB respectively.

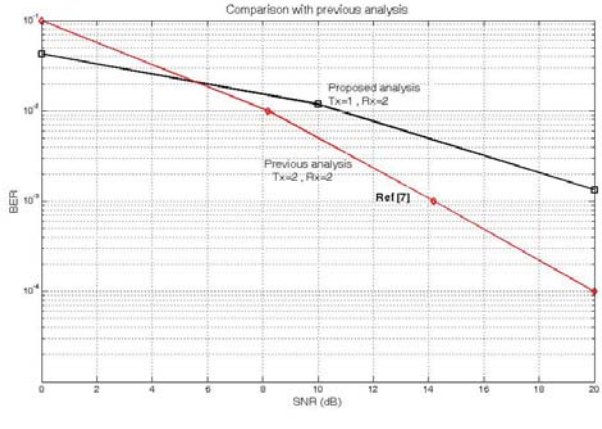


Figure-12 : Comparison with the simulation results reported earlier for 2-Tx and 2-Rx as in Ref [7]

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Noise Optimized Minimum Delay Spread Equalizer Design for DMT Transceivers

Toufiqul Islam [1], Satya Prasad Majumder [1] and Md. Kamrul Hasan [1,2]
[1] Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.
[2] Department of Electrical and Electronic Engineering
East West University, Dhaka, Bangladesh.
E-mail: {toufiq56, khasan, spmajumder}@eee.buet.ac.bd

Abstract— Time-domain equalizer (TEQ) design for multicarrier transceivers has recently received much attention. In this paper, we consider generalization of one such design method which takes into account the noise observed in discrete multitone (DMT) systems. We propose an iterative TEQ design method which jointly minimize delay spread of the channel and filtered noise at the output of the equalizer. Experimental results show that our method can successfully minimize delay spread and noise when compared to other reported techniques.

I. Introduction

Channel shortening, a generalization of equalization, has recently become necessary in receivers employing multicarrier modulation (MCM). MCM techniques like orthogonal frequency division multiplexing (OFDM) and DMT have been deployed in applications such as the wireless LAN standards IEEE 802.11a and HIPERLAN/2, digital audio broadcast (DAB) and digital video broadcast (DVB) in Europe, and asymmetric and very-high-speed digital subscriber loops (ADSL, VDSL). MCM is attractive because intercarrier interference (ICI) and intersymbol interference (ISI) can be avoided by inserting a cyclic prefix (CP) between consecutive symbols. If the channel impulse response (CIR) spans no more than $\nu + 1$ samples, where ν is the CP length, then linear convolution with channel turns into circular convolution, enabling one-tap frequency domain equalization (FDE). However, the CP reduces the bandwidth efficiency by the factor $N/(N+\nu)$, where N is the length of the DMT symbol. Hence, ν cannot be high as it in turn reduces bit rate.

A well-known technique to combat the ICI/ISI caused by the inadequate CP length is the use of a time-domain equalizer (TEQ) in the receiver front end. The TEQ is a finite impulse response filter that shortens the channel so that the delay spread of the combined channel-equalizer effective impulse response (EIR) is not longer than the CP length. Several TEQ designs have been proposed [1],[2],[3],[4]. One of the drawback of these methods is the need to search over the range of admissible transmission delays, Δ in order to design optimum TEQ. In [5], a unique TEQ design method is proposed which claims to minimize the delay spread of the overall channel impulse response (called MDS method). MDS method is computationally less intensive as it does not require to search for optimum Δ. This approach is independent of the CP length and attempts to squeeze the EIR as much as possible. This is advantageous since EIR squeezing allows further to reduce CP length to increase data rate and provides additional robustness to synchronization offset. Recently, in [6], MDS method is modified to account for the true time reference about which delay spread needs to be minimized. Unfortunately, both these methods, [5] and [6], do not account for any knowledge of noise statistics. Due to noise source models for DMT systems, such as near-end crosstalk (NEXT) and far-end crosstalk (FEXT) [7], it is only natural to exploit such knowledge to obtain a more robust equalizer.

In this paper, we generalize the method of [6] to some aspects. First, we modify the cost function to account for minimization of noise. Secondly, a trade-off parameter is also introduced to set appropriate weight for minimizing delay spread and noise. Thirdly, we propose an iterative algorithm for TEQ update where trade-off parameter is adaptively adjusted at each iteration under a specific criteria. Simulation results provided show that our method performs comparatively well in terms of delay spread minimization and filtered noise suppression.

II. System model

The channel/equalizer model is shown in Fig. 1 and the notation is summarized in Table 1. We make the following assumptions here.

- The channel, $\mathbf{h}$ and the TEQ, $\mathbf{w}$ are finite impulse response (FIR) filters.
- The input x_k is zero mean and white with variance σ_x^2.
- The noise u_k is zero mean wide-sense stationary (WSS) random process with covariance matrix $\mathbf{R}_{\mathrm{u}}$.
- The processes x_k and u_k are uncorrelated.
- The channel, $\mathbf{h}$ is known at the receiver. In practice, channel state information is obtained via training sequence.

The effective channel is given by $c_k = h_k * w_k$ where '$*$' denotes linear convolution. Then the output y_k can be written as

$$\begin{aligned} y_k &= r_k * w_k \\ &= x_k * h_k * w_k + u_k * w_k \\ &= x_k * c_k + q_k \end{aligned} \quad (1)$$

978-1-4244-2014-8/08/$25.00 (c)2008 IEEE

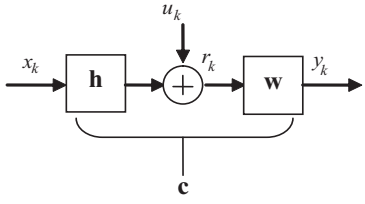


Fig. 1. System Model with channel/equalizer.

Table 1
Channel shortening notation

Notation	Meaning
ν	length of CP
N	FFT size
$\triangle$	delay of effective channel
$\mathbf{h}$	$L_h \times 1$ CIR vector
$\mathbf{w}$	$L_w \times 1$ TEQ impulse response vector
$\mathbf{c}$	$L_c \times 1$ EIR vector
$\mathbf{H}$	$L_c \times L_w$ channel convolution matrix
$\mathbf{R_u}$	$L_w \times L_w$ noise covariance matrix
$\mathbf{I_n}$	$n \times n$ identity matrix
$\mathbf{A^T}$	transpose
$diag[\ ]$	diagonal matrix with entries in the main diagonal

where q_k is the filtered noise. In vector form, the effective channel impulse response, $\mathbf{c}$ can be written as

$$\mathbf{c} = \mathbf{H}\mathbf{w} \tag{2}$$

where $\mathbf{H}$ is the Toeplitz convolution matrix of the unequalized channel, $\mathbf{h}$. $\mathbf{H}$ is configured as

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{L_h-1} & h_{L_h-2} & \cdots & h_{L_h-L_w} \\ 0 & h_{L_h-1} & \cdots & h_{L_h-L_w+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{L_h-1} \end{bmatrix}$$

III. Problem Formulation

We restate the definition of the following fundamental quantities from [6] for a given impulse response $\{g_k, -\infty \le k \le \infty\}$:

$$E_g = \sum_k g_k^2 \quad \text{(energy)} \tag{3}$$

$$k_g = \frac{1}{E_g}\sum_k k g_k^2 \quad \text{(centroid)} \tag{4}$$

$$D_g = \sqrt{\frac{1}{E_g}\sum_k (k-k_g)^2 g_k^2} \quad \text{(delay spread)} \tag{5}$$

The MDS design of [5] chooses $\mathbf{w}$ in order to minimize the following cost function:

$$J_{ds} = \frac{1}{E_c}\sum_k (k-k_{ref})^2 c_k^2 \tag{6}$$

where E_c is the energy of the effective channel. J_{ds} is, in fact, the square of the delay spread. Hence, [5] actually attempts to minimize the delay spread of the effective channel. This method assumes centroid of the original channel, k_h, as the time reference, k_{ref}. Now, we show that choice of [5] for k_{ref} is not optimum. From (6), after some modification, we get that

$$J_{ds} = \frac{1}{E_c}\{\sum_k k^2 c_k^2 - 2k_{ref}\sum_k k c_k^2 + k_{ref}^2 \sum_k g_k^2\} \tag{7}$$

Now, using (3),(4) and setting $\partial J_{ds}/\partial k_{ref}$ equal to zero, we obtain

$$k_{ref} = k_c \tag{8}$$

where k_c is the centroid of EIR. It reveals that we need to minimize (6) using k_c, not k_h. The cost function of (6) can be written as

$$J_{ds} = \frac{\mathbf{c^T}\mathbf{\Lambda^2_{k_{ref}}}\mathbf{c}}{\mathbf{c^T c}} \tag{9}$$

where $\mathbf{\Lambda}_{k_{ref}}$ is a diagonal weighting matrix defined as

$$\mathbf{\Lambda}_{k_{ref}} = diag[0, 1, \ldots, L_c - 1] - k_{ref}\mathbf{I_{L_c}} \tag{10}$$

Now using (2), we get

$$J_{ds} = \frac{\mathbf{w^T H^T}\mathbf{\Lambda^2_{k_{ref}}}\mathbf{H w}}{\mathbf{w^T H^T H w}} \tag{11}$$

Noise and filtered signal power at the output of the TEQ can be written respectively as

$$\sigma_q^2 = \mathbf{w^T R_u w} \tag{12}$$

$$\sigma_{x_f}^2 = \sigma_x^2 \mathbf{w^T H^T H w} \tag{13}$$

The cost function to minimize noise power is defined as ratio of filtered noise power and filtered signal power:

$$J_n = \frac{\sigma_q^2}{\sigma_{x_f}^2} = \frac{\mathbf{w^T R_u w}}{\sigma_x^2 \mathbf{w^T H^T H w}} \tag{14}$$

Finally, the objective function can be formulated by coupling the two cost functions of (11) and (14) via a trade off parameter, α :

$$J \triangleq \alpha J_{ds} + (1-\alpha) J_n, \quad 0 \le \alpha \le 1 \tag{15}$$

Hence, the problem is to find optimum TEQ $\mathbf{w}_{\text{opt}}$ by minimizing J in (15):

$$\mathbf{w}_{\text{opt}} = \arg\min_{\mathbf{w}} J = \arg\min_{\mathbf{w}} \frac{\mathbf{w^T X w}}{\mathbf{w^T Y w}} \tag{16}$$

where

$$\mathbf{X} = \alpha \mathbf{H^T}\mathbf{\Lambda^2_{k_{ref}}}\mathbf{H} + \frac{(1-\alpha)}{\sigma_x^2}\mathbf{R_u} \tag{17}$$

$$\mathbf{Y} = \mathbf{H^T H} \tag{18}$$

IV. Optimum TEQ Design

From (4), it is obvious that we need to know EIR, $\mathbf{c}$, to find k_c which in turn require to solve (16) first. So we proceed to iteratively minimize the (15) using centroid obtained at the previous iteration as time reference for the next [6].

At ith iteration, optimum $\mathbf{w}$ is obtained as

$$\mathbf{w}_i = \arg\min_{\mathbf{w}} \; J \mid_{k_{ref}(i-1)} \tag{19}$$

Note that at ith iteration, $\mathbf{\Lambda}_{k_{ref}}$ in (10) present inside matrix $\mathbf{X}$ in (17) is formed using $k_{ref}(i-1)$. Finally, k_{ref} for ith iteration is set as the centroid of the effective channel:

$$\begin{aligned} k_{ref}(i) &= \frac{1}{E_c}\sum_k k c_k^2 \\ &= \frac{\mathbf{c^T \Upsilon c}}{\mathbf{c^T c}} \\ &= \frac{\mathbf{w}_i{}^T \mathbf{H^T \Upsilon H w}_i}{\mathbf{w}_i{}^T \mathbf{H^T H w}_i} \\ \text{where} \quad \mathbf{\Upsilon} &= diag[0, 1, \ldots, L_c - 1] \end{aligned} \tag{20}$$

α can be adapted based on proportionate ratio of delay spread and filtered noise power :

$$\begin{aligned} \alpha(i) &= \frac{\mathbf{w}_i{}^T \mathbf{H^T \Lambda^2_{k_{ref}(i)} H w}_i}{\mathbf{w}_i{}^T \mathbf{Z w}_i} \\ \text{where} \quad \mathbf{Z} &= \mathbf{H^T \Lambda^2_{k_{ref}(i)} H} + \frac{1}{\sigma_x^2}\mathbf{R}_{\mathrm{u}} \end{aligned} \tag{21}$$

The resulting algorithm is as follows

- Precompute $\mathbf{Y}$ of (18) and initialize $k_{ref}(0) = k_h$ i.e as centroid of the CIR and $\alpha(0)$ = 0.5 (arbitrary choice).
- for i=1, 2, . . . do the following

1. Compute weighting matrix $\mathbf{\Lambda}_{k_{ref}}$ of (10) using $k_{ref}(i-1)$.
2. Compute matrix $\mathbf{X}$ of (17) and obtain optimum TEQ for ith iteration, $\mathbf{w}_i$ solving (19).
3. Compute $k_{ref}(i)$ using (20).
4. Compute $\alpha(i)$ using (21). Here, $\mathbf{\Lambda}_{k_{ref}}$ is calculated using $k_{ref}(i)$ as it is available.

Within few iterations, k_{ref} becomes fixed and J ceases to be minimized more. Then iteration is stopped and optimum TEQ, $\mathbf{w}_{\mathrm{opt}}$ is obtained. Fig. (2) shows that J cannot increase from iteration to iteration. It is obvious because when k_{ref} is set to k_c (see (8)), J_{ds} will minimize which effectively minimizes J as true time reference is gradually reached with iterations.

V. Simulation Results

We now proceed to analyze how our design works. The channels used are eight standard downstream carrier service area (CSA) loops combined with a plain old telephone service (POTS) splitter and a twelfth-order Chebyshev bandpass filter for the 30-1000 kHz frequency band, and truncated to 512 samples. Data for the channel and noise was obtained from the Matlab DMTTEQ Toolbox [9]. We used the following asymmetric digital subscriber line (ADSL) input parameters:

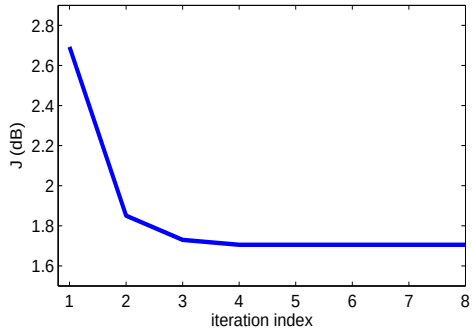


Fig. 2. Cost function J gradually stabilizes with iterations (for CSA loop 8).

- DFT size, N = 512 and sampling frequency, $fs = 2.208$ MHz.
- ν = 32, L_w = 16 and L_h = 512.
- Input power, σ_x^2 = 21 dBm. Input signal x_k consists of QAM symbols.
- Input noise consists of near-end crosstalk (NEXT) noise plus additive white noise with power density -110 dBm/Hz.

Fig. (3) shows the original and equalized channel impulse responses upto 470 samples for CSA loop 8 designed using the proposed method . As we can see, our method shortened the channel quite well. Fig. (4) shows the corresponding magnitude responses of original and shortened channel. As the cost function jointly minimize delay spread and filtered noise, in Fig. (5) and Fig. (6), we have plotted delay spread and output SNR variation as function of TEQ length, L_w. Signal-to-noise ratio (SNR) at the output of the equalizer can be obtained using (12) and (13) as:

$$\begin{aligned} \text{SNR} &= \frac{\sigma^2_{x_f}}{\sigma_q^2} \\ &= \frac{\sigma_x^2 \mathbf{w^T H^T H w}}{\mathbf{w^T R}_{\mathrm{u}} \mathbf{w}} \end{aligned} \tag{22}$$

As expected, increasing L_w results in improved performance (delay spread reduces, output SNR rises) which comes at the expense of the increased complexity in implementing the additional equalizer taps. It is also noticed from Fig. (5) and Fig. (6), delay spread and output SNR reaches a floor with increasing L_w. In Table 2, we compared our method with other existing methods on the basis of delay spread. Except [6], our method achieves delay spread which is lower than other methods. Minimum delay spread attained by modified MDS method in [6] is expected as it corrects the original time reference problem present in [5] and only targets to minimize the delay spread. Whereas in this paper, we perform a trade-off between minimizing delay spread and minimizing noise power. In Table 3, we compared output SNR for the methods which are derived from [5]. It is clear from Table 3 that our method achieves higher SNR than other methods. Incorporation of α allows appropriate weight for J_{ds} and J_n so that both output noise and delay spread are minimized in optimum manner.

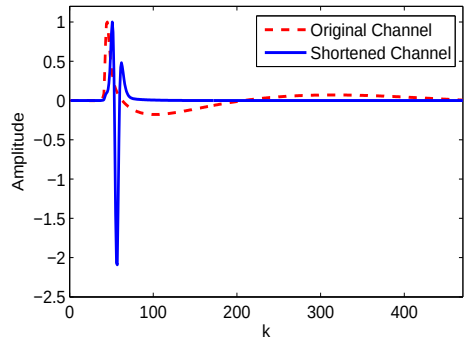


Fig. 3. Original and shortened channel impulse responses (for CSA loop 8).

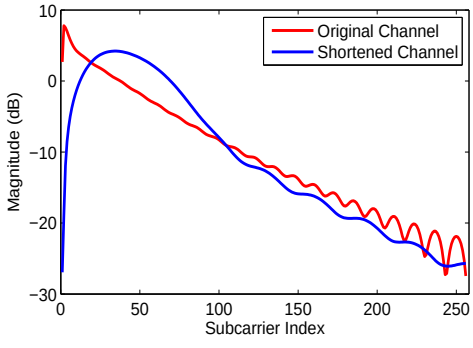


Fig. 4. Original and shortened channel frequency responses (for CSA loop 8).

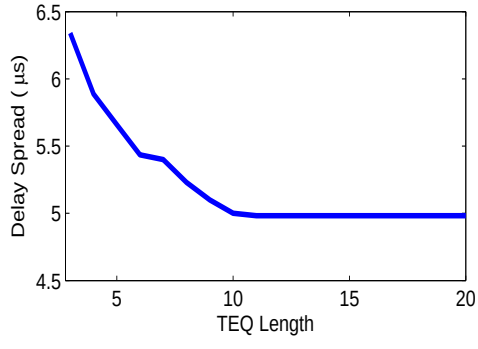


Fig. 5. Variation of delay spread as function of TEQ length (for CSA loop 1).

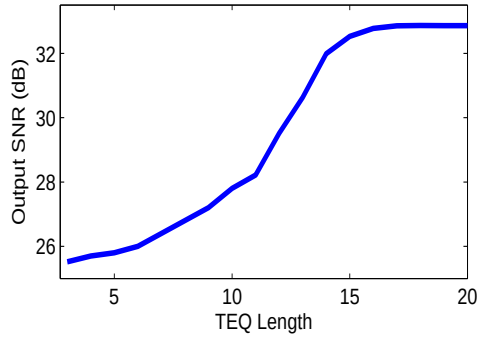


Fig. 6. Equalizer output SNR as function of TEQ length (for CSA loop 1).

Table 2
Observed delay spread for various methods (averaged over eight CSA loops)

Method	Delay Spread (μs)
MDS Design [5]	6.27
Eigenfilter Method [4]	7.61
MSSNR Method [3]	4.50
Modified MDS Design [6]	4.40
MMSE Method [1]	4.64
Proposed	4.42

Table 3
Output SNR (in dB) comparison of various delay spread minimizing methods

CSA loop no.	MDS [5]	Eigenfilter [4]	Modified MDS [6]	Proposed
1	26.94	22.01	29.31	32.91
2	28.05	20.75	27.10	28.63
3	26.27	20.82	28.91	31.81
4	27.27	18.35	28.00	31.56
5	26.30	21.57	29.10	31.54
6	24.55	19.15	27.87	30.29
7	27.05	17.49	29.81	32.57
8	24.54	15.08	30.01	31.49

VI. Conclusion

In this paper, we have generalized the modified MDS method to account for noise encountered in the system. Trade off parameter involved in the cost function successfully adjusts weight for delay spread and noise minimization. Simulation results show that our method jointly achieves superior delay spread minimization and filtered noise suppression at the output of the equalizer. This method can be easily extended to incorporate multi-antenna system following the guidelines of [8].

References

[1] N. A. Dhahir and J. M. Cioffi, "Effficiently computed reduced-parameter input-aided MMSE equalizers for ML detection ," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 903–915, May 1996.

[2] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone transceivers to maximize bit rate," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.

[3] P. J. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, Dec. 1996.

[4] A. Tkacenko and P. P. Vaidyanathan, "A low-complexity eigenfilter design method for channel shortening equalizers for DMT systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1069–1072, July 2003.

[5] R. Schur and J. Speidel, "An efficient equalization method to minimize delay spread in OFDM/DMT systems," in *Proc.IEEE Int. Conf. Commun.*, Helsinki, Finland, June 2001, vol. 1, pp. 1–5.

[6] R. López-Valcarce, "Minimum delay spread TEQ design in multicarrier systems," *IEEE Signal Processing Letters*, vol. 4, pp. 112–114, 1997.

[7] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology.* Upper Saddle River, NJ: Prentice-Hall, 1999.

[8] A. Tkacenko and P. P. Vaidyanathan, "Eigenfilter design of MIMO equalizers for channel shortening," in *Proc. IEEE Int. Conf. Acoustic Speech Signal Processing.*, Orlando, FL, May 2002, vol. 3, pp 2361–2364.

[9] G. Arslan, M. Ding, B. Lu, M. Milosevic, Z. Shen and B. L. Evans , "MATLAB DMTTEQ Toolbox 3.1," The University of Texas at Austin, May 10, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Finding a Unique Association Rule Mining Algorithm Based on Data Characteristics

Mohammed M. Mazid, A.B.M. Shawkat Ali, Kevin S. Tickle

School of Computing Sciences, Faculty of Business and Informatics

Central Queensland University, Rockhampton, QLD-4702, Australia
Email : m.mazid@cqu.edu.au; s.ali@cqu.edu.au; k.tickle@cqu.edu.au

Abstract - This research compares the performance of three popular Association Rule Mining algorithms, namely Apriori, Predictive Apriori and Tertius based on data characteristics. The accuracy measure is used as the performance measure for ranking the algorithms. A wide variety of Association Rule Mining algorithms can create a time consuming problem for choosing the most suitable one for performing the rule mining task. A meta-learning technique is implemented for a unique selection from a set of association rule mining algorithms. On the basis of experimental results of 15 UCI data sets, this research discovers statistical information based rules to choose a more effective algorithm.

I. Introduction

Association Rule Mining (ARM) is one of the most important and substantial techniques in machine learning areas. It is particularly important for extracting knowledge from large databases by discovering frequent itemsets and associating item relationships between or among items of a data file. ARM is a powerful exploratory technique with a wide range of applications such as marketing policies, medical diagnosis, financial forecast, credit fraud detection and many other research areas. Machine learning researchers have already proposed a number of ARM algorithms, including Apriori [3], Predictive Apriori [19], Tertius [9], CLOSET [17], MAFIA [8], ELACT [25], CHARM[23] and many others. From among these rapidly increasing arrays of ARM algorithms, users easily become confused in choosing the right algorithm for specific kinds of data. The aim of this research is to determine characteristics of datasets using meta-learning processes and provide users a method to choose the right algorithm based on data characteristics. Of course, it is important to keep in mind Wolpert and Macready's well-known No Free Lunch (NFL) theorem:

"If algorithm A outperforms algorithm B on some cost function, then loosely speaking there must exist exactly as many other functions where B outperforms A" [4].

Credit for development of Association Rule Mining is mostly attributed to Agrawal [3]. Association rule is an expression of X→Y (read as 'if X then Y'), where X and Y are itemsets in a database D. The expression can be illustrated as 'if a customer buys item X then likely to buy item Y' or 'if a patient is infected by disease X then likely to be infected by disease Y', and so on. The itemset of the left hand side of the arrow is called antecedent and itemset of the right hand side of the arrow is called consequent. Each expression is called a rule. A rule can contain from two to an unrestricted number of items with or without conjunctions using of AND or OR operands. An item of a rule is selected from the frequent itemset of the data file. Frequent itemsets are the items that occur more frequently. Basically, ARM follows two major steps to produce rules from a data file: first, find all the frequent itemsets; second, generate strong association rules from the frequent itemsets. The best rules are picked the basis of different types of interestingness measurement rules. Some of these will be discussed later on in this paper.

The Apriori [3] algorithm has become the standard approach and most widely used tool for Association Rule Mining [25]. Many ARM algorithms are being used at present but Apriori has become one of the most prevailing algorithms. However a major problem of Apriori is size of datasets and subsequent frequent itemsets while generating rules. Hegland [11] reviewed the Apriori algorithm and discussed variants for distributed data, inclusion of constraints and taxonomies. The author also talked about some potential tools which deal with long itemsets and considerably reduce the number of (uninteresting) itemsets returned. "A frequent pattern of length l implies the presence of 2^l - 2 additional frequent patterns as well, each of which is explicitly examined by such algorithms" [24]. For a simple example, a datafile with 4 items could produce 14 itemsets. Details of possible itemsets are shown in Figure 1.

Researchers have proposed two ways to solve this problem. The first one is to mine only maximal frequent itemsets [2, 6, 10, 13] which are typically fewer in order of magnitude than all frequent patterns. The second is to mine only the frequent closed sets [5, 16, 21, 26]. "Closed sets are lossless in the sense that they can be used to uniquely determine the set of all frequent itemsets and their exact frequency" [24]. However, this study found that performance of Apriori is still the best among the three algorithms considered.

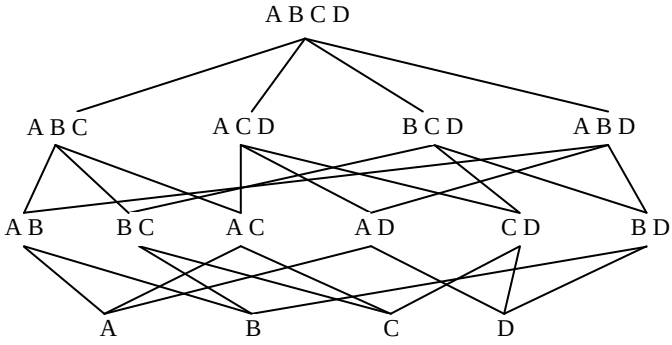


Fig. 1 Possible itemsets of four items

In this study, we have examined 15 problems from the UCI Repository [7]. The details of the data sets description is provided in Table 1. We evaluated rules produced by three Association Rule Mining algorithms using accuracy measures. A Java based machine learning tool Weka3.4 [22] was used to perform the experiment. For all the three algorithms, the first 10 best rules were chosen. The machine configuration was Intel Core2 Duo CPU 2.33GHz and 4GB RAM.

The rest of the paper is organized as follows: Section 2 describes briefly the three association algorithms. Section 3 describes the different measurements of interestingness we have used to perform the experiment. The experimental process, data preparation and result are presented in Section 4. Finally we draw conclusions from our research in Section 5.

II. Algorithms

This section provides a brief description of all the algorithms we considered in our experimental design. All of the algorithms belong to the category of unsupervised learning methods, but we can further categorise them into Association Rule Mining algorithms as described in the following sections.

A. Apriori

The Apriori algorithm was first proposed by Agrawal [1]. It uses prior knowledge of frequent tools for association rule mining. The basic idea of the Apriori algorithm is to generate frequent itemsets of a given dataset and then scan the dataset to check if their counts are really large. The process is iterative and candidates of any pass are generated by joining frequent itemsets of the proceeding pass. Apriori is a confidence-based Association Rule Mining algorithm. The confidence is simply accuracy to evaluate rules, produced by this algorithm. The rules are ranked according to the confidence value. If two or more rules share the same confidence then they are initially ordered using their support and secondly the time of discovery.

B. Predictive Apriori

In the case of Apriori, every so often we can find rules with higher confidence but low support on respective items of generated rules. Sometimes, rules are produced with large support but low confidence. Scheffer [19] introduced this algorithm with the concept of "larger support has to trade against a higher confidence". Predictive Apriori is also a confidence-based ARM algorithm. But rules ranked by this algorithm are sorted according to "expected predicted accuracy". The interestingness measure of Predictive Apriori suits the requirements of a classification task [15]. It tries to maximise expected accuracy of an association rule rather than confidence in Apriori.

C. Tertius

The Tertius system implements a top-down rule discovery system employing the confirmation measure [9]. Tertius uses a first-order logic representation that deals with structured, multi-relational knowledge. It allows users to choose the most convenient option among several possible options. Generated rules from Tertius are descriptive rather than predictive. Descriptive approaches include clustering, association rule learning, learning of attribute dependencies, subgroup discovery, and multiple predicate learning [9]. Tertius is able to deal with extensional knowledge, either with explicit negation or under the Closed-World Assumption. This algorithm also considers induction of integrity constraints in databases, learning mixed theories of predicate definitions and integrity constraints.

III. Description of Algorithm Performance Measures

In the past few years researchers have contributed to finding the interestingness measurement of a rule. There are numbers of measures for this purpose such as Support, Confidence, Predictive Accuracy, Conviction, Lift, Coverage, etc. In our research, we have used the following measurement for the performance of rules.

A. Support

Support for ARM is introduced by Agrawal [1]. It measures the frequency of association, i.e. how many times the specific item has been occurred in a dataset. An itemset with greater support is called frequent or large itemset. In terms of probability theory we can express support as:

$$\text{Support} = P(A \cap B) = \frac{\text{number of transactions containing both A an B}}{\text{Total number of transactions}}$$

where A and B are itemsets in a database D.

B. Confidence

Confidence measures the strength of the association [5]. It determines how frequently item B occurs in the transaction that contains A. Confidence expresses the conditional probability of an item. The definition of confidence is

$$\text{Confidence} = P(A \mid B) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{\text{number of transactions containing both A and B}}{\text{number of transactions containing A}}$$

C. Predictive Accuracy

Predictive accuracy is another way to measure interestingness of a generated rule. Basically this accuracy is used for the Predictive Apriori rule measurement. According to Scheffer [19], the definition of predictive accuracy is:

Let ***D*** be a data file with ***r*** number of records. If ***[x → y]*** is an Association Rule which is generated by a static process ***P*** then the predictive accuracy of [***x → y***] is

c([x → y])=***Pr***[***r*** satisfies ***y***|***r*** satisfies ***x***]

where distribution of ***r*** is govern by the static process ***P*** and the Predictive Accuracy is the conditional probability of ***x→r*** and ***y→r***.

IV. Experiments

Our research compares three algorithms of ARM which are Apriori, Predictive Apriori and Tertius. We organised our research into three main steps: first we compare the algorithms across a number of different measures of accuracy providing a comprehensive empirical evaluation of the performance of three association algorithms on 15 datasets. We then characterise the datasets using the central tendency measures and some other basic statistical measures. Finally, the empirical results are combined with the dataset characteristic measures to generate rules describing which algorithm is best suited to which type of problems.

A. Data Preparation

After selecting data files without any missing values, we have applied three algorithms of Association Rule Mining that are Apriori, Predictive Apriori and Tertius using WEKA 3.4. For Association Rule Mining, WEKA uses 'unsupervised discretized' value of attributes. We have used 3-bin discretization and the other parameter settings of Discretize option of the software have been kept as default. As well, parameter settings for each algorithm were the default and we always chose the first 10 best rules.

B. Meta Learning Process

Meta learning is one of the sections of Machine learning technique. In this technique automatic learning algorithms are applied on pre processed data to conduct machine learning experiments. In our experiment we converted nominal data to numeric by using Java programming and analysis those data using Matlab [14]. Then a powerful data mining tool See5 [18] had been used to get final rules.

Table1 : Basic Properties of 15 data files

Data file Name	No of Instance	No of Attribute			
		Total	Nominal	Numeric	No. of Class
page-blocks	5473	11	1	10	1
vehicle	846	19	1	18	1
Market Basket	1000	12	3	10	2
liver disorders	345	7	1	6	1
ionosphere	351	35	1	34	1
Zoo	101	18	17	1	1
ecoli	336	8	1	7	1
cmc	1473	10	2	8	X*
breast cancer	286	10	10	0	1
prostoperative patient data	90	9	9	0	1
bridges version1	107	13	10	3	1
Iris	150	5	1	4	1
tae	151	6	3	3	1
haberman	306	4	2	2	1
car	1728	7	7	0	1

* Denotes none

C. Experimental Design

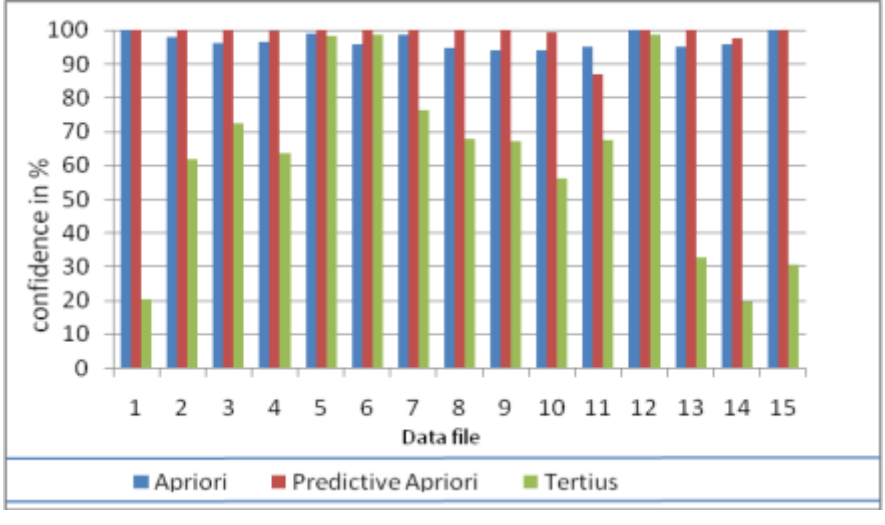
In the first step of our experiment, we have compared average confidence of the 10 best rules of 15 data files. On the basis of confidence, Apriori always showed supremacy compared to the other two algorithms (Predictive Apriori and Tertius). The details of the comparative study are provided in Figure 2- 4. Rules generated from Tertius are worst in terms of confidence comparison. Then we compared between Apriori and Predictive Apriori in terms of average confidence and predictive accuracy of the initially selected 15 data files (The details of the data sets description are provided in Table 1). We have converted Nominal data to numeric values to perform the statistical analysis. We have considered simple, statistical and information theoretic measures to identify the dataset characteristics. The details of statistical characteristics comparisons are provided in Table 2. Some of the statistical formulation is available in MATLAB Statistics Toolbox [14]. By using See5 [18] data mining tools with the most co-related attributes, we generate rules to identify which algorithm is suitable for which type of problem. Finally, we examined the rules by the statistical measures. The details of the rule descriptions are provided in Table 5.

D. Performance Analysis

In the experiment, we have compared Apriori, Predictive Apriori and Tertius with 'confidence' in the first step. Apriori always shows the best performance in this experiment. The details of the 'confidence' based performance of 3 algorithms are shown in Figure 2, 3, and 4.

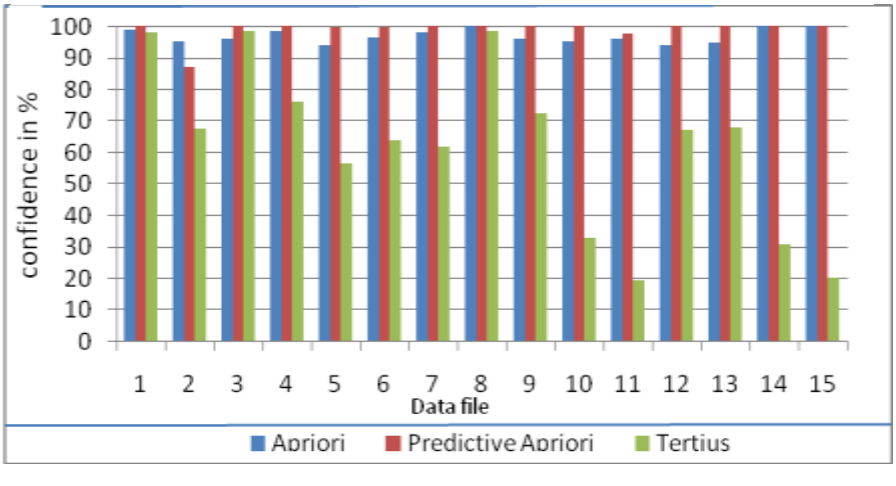
As in the first step of this experiment, Apriori and Predictive Apriori showed better performance. Tertius had the worst among those in terms of 'confidence' based comparisons. So we compared Apriori and Predictive Apriori with 'confidence' and 'predictive accuracy' respectively. Then we found out which algorithm is better between them. The details of the data sets comparisons are provided in Table2.

Figure 2 : Performance of Apriori based on average confidence



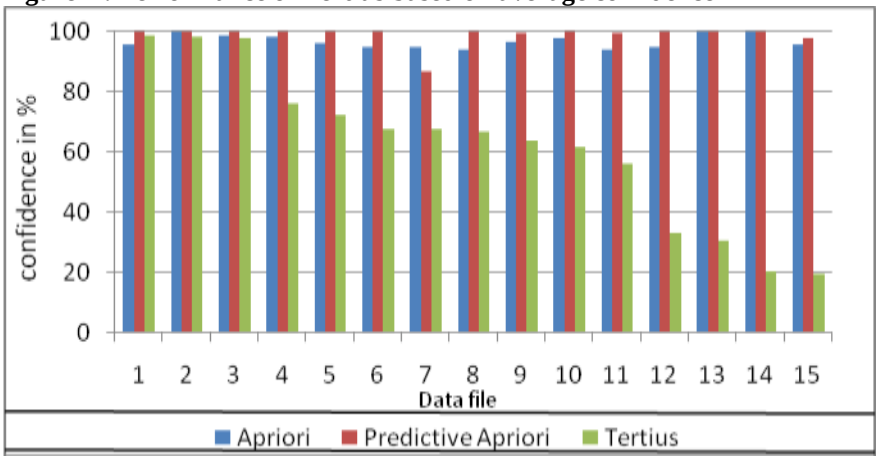
1	ionosphere
2	bridges version1
3	Zoo
4	Ecoli
5	postoperative patient data
6	liver disorders
7	vehicle
8	Iris
9	Market Basket
10	tae
11	haberman
12	breast cancer
13	cmc
14	car
15	page-blocks

Figure 3 : Performance of Predictive Apriori based on average confidence



1	page-blocks
2	vehicle
3	Market Basket
4	liver disorders
5	ionosphere
6	Zoo
7	ecoli
8	cmc
9	breast cancer
10	postoperative patient data
11	bridges version1
12	Iris
13	tae
14	haberman
15	car

Figure 4 : Performance of Tertius based on average confidence



1	Zoo
2	Iris
3	ionosphere
4	ecoli
5	Market Basket
6	cmc
7	bridges version1
8	breast cancer
9	liver disorders
10	vehicle
11	postoperative patient data
12	tae
13	car
14	page-blocks
15	haberman

Table 2 : Comparisons of algorithms

Data file Name	Apriori	Predictive Apriori	Better Performance
	Ave. Confidence of 10	Ave. accuracy of 10 rules	Among two
bridges version	0.946	0.861495	*Apriori*
Car	1	0.994993	*Apriori*
Cmc	0.947	0.994989	*predApriori*
Ecoli	0.984	0.994923	*pred Apriori*
Haberman	0.96	0.969948	*predApriori*
ionosphere	0.988	0.994017	*predApriori*
Iris	1	0.9929	*Apriori*
liver disorders	0.970833333	0.968394167	*Apriori*
Market Basket	0.959	0.971457396	*predApriori*
page-blocks	1	0.9949	*Apriori*
prostoperative patient data	0.94	0.983791366	*predApriori*
Tae	0.950	0.991509	*predApriori*
Vehicle	0.982	0.994866	*predApriori*
Zoo	0.959	0.993823	*predApriori*
breast cancer	0.944	0.994545	*predApriori*

V. Algorithm selection

A. Datset Characterization Measurement

Each dataset can be described by a number of simple, statistical and information theoretical measures [20]. We average some statistical measures of all the attributes and take these as a global measure of the dataset characteristics (The details of the data sets statistical characteristics are provided in Table 4). The Table 3 lists some of the statistical measures used in our experiment to find out data characteristics.

Table 3: Statistical measures for characterisation of each dataset

Measure	Notation
Arithmetic Mean	Mean
Median	Median
Mode	Mode
Geometric mean	GM
InterQuartile Range	IQR
Range	Range
Variance	Var
Standard deviation	Std
Z-score	Z_score

A brief description of the above measures are as follows:

Arithmetic mean

The arithmetic mean or the simple mean of a list of sequence $\overline{X}$ is:

$$\frac{1}{n}\sum_{i=1}^{1} x_i = \frac{1}{n}(x_1 + \ldots\ldots\ldots + x_n)$$

Mode

The Mode is the most frequent incident number in a sequence. For example, the mode of 16, 15, 18, 9, 18, 11, 18 and 33 is 18.

Median

The Median is the middle number in a list of sequence; that is, half the numbers have values that are greater than the median, and half the numbers have values that are less than the median. For example, the median of 16, 15, 18, 9, 18, 11, 18 and 33 is 17.

Geometric mean (GM)

The geometric mean of a sequence $\{X\}_{i=1}^{n}$ is:

$$\left[\prod_{i=1}^{n} X\right]^{1/n} = \sqrt[n]{x_i, x_2, ...x_n}$$

InterQuartile Range (IQR)

The InterQuartile Range (IQR) is the difference between the 75th percentile and the 25th percentile.

Variance (Var)

The variance is used to characterize the dispersion among the measures in a given population. It calculates the mean of the scores, and then measures the amount that each score deviates from the mean and then squares that deviation for a given population. Numerically, the variance equals the average of the squared deviations from the mean.

Standard Deviation (Std)

The std measures the spread of a set of data as a proportion of its mean

$$\text{std} = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}$$

where $\bar{x}$ is Arithmetic mean.

Z Score

Z-score is derived from subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. It helps to measure, how far the observation is from the mean in units of standard deviation. The value of Z-score is greater than 3 indicates that the data distribution has outliers [4].

$$\text{Zscore} = \frac{x - \bar{x}}{\sigma}$$

Table 4: Using statistical formulation and comparison between confidence and 'predictive accuracy' we have generated the following table

	Mean	Median	Mode	GM	IQR	Range	Var	Std	Zscore	Class
Breast-Cancer	584.19	607.2	607.2	576.85	96.3	252.7	8416.6	64.934	23.462	2
bridge_version	212.48	233.75	225.92	198.64	55.938	193.58	6034.9	54.408	6.7602	1
car	312.88	250.92	222.67	282.95	207.25	268	16230	116.15	3.91	1
cmc	5.4593	5.4444	4.7778	4.7094	2.5833	7.1111	8.559	1.7011	17.247	2
ecoli	187.87	192	192.43	186.47	3.6429	149	326.72	17.149	3.263	2
haberman	39.779	38.667	36.667	38.032	9	38.667	59.655	7.0808	11.44	2
ionosphere	269.18	337.68	51.176	209.21	200.76	344.26	18575	132.11	4.4927	2
iris	149.19	149	148	149.16	4.25	11	7.8595	2.7971	40.272	1
liver-disorders	112.88	112	111.17	112.36	4.2083	71	114.82	9.8548	18.25	2
market-basket	175.81	177.67	178.5	173.69	18.125	138.75	661.93	24.643	15.954	2
page-blocksW	160.81	158.9	151.8	157.86	32.6	142.8	750	24.799	16.018	1
pos-patient-data	497.88	435.31	421.13	486.96	142.88	230.25	15001	92	5.8092	2
tae	71.115	71	71.4	69.56	20	34.8	233.42	11.303	3.2965	2
vehicle	123.06	122	121.72	122.37	11.5	32.889	123.46	8.2291	10.275	2
zoo	472.73	480.71	463.35	469.99	57.368	96.059	3139.7	40.134	10.75	2

B. Rule Generation

As a part of the meta-learning process, we have used data from Table 4 and generated rules using See5 data mining tools. See 5 basically classified the data characteristics on the basis of Mode and Geometric Mean. The details of the characteristics rules are provided in Table5.

Table 5 : Characteristics rules generated by See5.

```
See5 [Release 2.03]
Options:
        Rule-based classifiers
        Pruning confidence level 99%
Read 15 cases (9 attributes)
Rules:
Rule 1: (6/2, lift 2.3)
        Mode > 121.72 & Geomean <= 282.95 ->  class 1  [0.625]
Rule 2: (6, lift 1.2)
        Mode <= 121.72   ->  class 2  [0.875]
Rule 3: (3, lift 1.1)
        Geomean > 282.95 ->  class 2  [0.800]
Default class: 2
Evaluation on training data (15 cases):
            Rules
          ----------------
            No      Errors
             3    2(13.3%)   <<
           (a)   (b)    <-classified as
          ----  ----
             4          (a): class 1
             2     9    (b): class 2
Time: 0.0 secs
```

Rule for Apriori

```
If
Mode > 121.72 and Geomean <= 282.95 of a data file then choose Apriori.
```

Rules for Predictive Apriori

```
if
Mode <= 121.72 or Geomean > 282.95 then choose Predictive Apriori
```

VI. Conclusion

Comparative study of three algorithms showed that Apriori had performed best on confidence based ranking and Predictive Apriori had performed better on accuracy based ranking. However the main aim of this research is to assist in the selection of an appropriate Association Rule Mining algorithm without the need for trial-and-error testing of the vast array of available algorithms. Based on statistical formulation measurement and meta-learning process, we can recommend an algorithm by analysing Mode and GM of a data file. Although we have used nine statistical measurements, the rules emphasise only two of them. We aim to continue this research by considering more association problems, extracting better rules using other participating statistical measurements.

Bibliography

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, September 1994.

[2] R. Agrawal, C. Aggarwal, and V.V.V. Prasad, "Depth First Generation of Long Patterns," Proc. Seventh Int'l Conf. Knowledge Discovery and Data Mining, August 2000.

[3] R. Agrawal, T. Imielinski, and Swami A. "Mining associations between sets of items in large databases," In Proc. of the ACM SIGMOD Int'l Conference on Management of Data. - Washington D.C. , pp. 207-216, May 1993

[4] S. Ali and K. A. Smith, "On learning algorithm for classification", Applied Soft Computing, pp. 119-138, December 2004.

[5] Y. Bastide, R. N. Taouil, Y. Pasquier, G. Stumme, and L. Lakhal, "Mining Frequent Patterns with Counting Inference ," SIGKDD Explorations, vol. 2. December 2000.

[6] R.J. Bayardo, "Efficiently Mining Long Patterns from Databases", Proc. ACM SIGMOD Conf. Management of Data. - June 1998.

[7] C. Blake and C.J. Merz, UCI Repository of machine learning databases, University of California, 2007. available at http://archive.ics.uci.edu/ml/. viewed on Feb 2008.

[8] D. Burdick, M. Calimlim and J. Gehrke, "MAFIA: a macimal frequent itemset algorithm for transactional databases," In Intlernational conference on Data Engineering. April 2001.

[9] P.A. Flach and N. Lachiche, "Confirmation-guided discovery of first-order rules with Tertius," Kluwer Academic Publishers. - The Netherlands, Vol. 42. - pp. 61-95, 2001

[10] K. Gouda and M.J. Zaki, "Efficiently Mining Maximal Frequent Itemsets," in Proc. First IEEE Int'l Conf. Data Mining. November 2001.

[11] M. Hegland, "Algorithms for Association Rules," Lecture Notes in Computer Science, Vol. 2600. pp. 226 – 234. January 2003.

[12] T.-S. Lim, Knowledge Discovery Central datasets. - 2002. available at http://www.kdcentral.com/.

[13] D-I. Lin and Z.M. Kedem, "Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set," in Proc. Sixth Int'l Conf. Extending Database Technology, March 1998.

[14] Matlab Statistics Toolbox User's Guide [Book]. - USA : The MathWorksInc, version 6.2., 2008

[15] S. Mutter, M. Hall and E. Frank, "Using Classification to Evaluate the Output of Confidence based Association Rule Mining," Lecture Notes in Artificial Intelligence,Advances in Artificial Intelligence - AI 2004. Berlin, Springer, vol. 3339. - pp. 538-549. 2004.

[16] N. Pasquier, Y. Bastide, R. Taouil, & L. Lakhal, "Discovering Frequent Closed Itemsets for Association Rules," Proc. Seventh Int'l Conf. Database Theory, January 1999.

[17] J. Pei, J. Han and R. Mao, "Closet: An Efficient Algorithm for Mining Frequent Closed Itemsets", Proc. SIGMOD Int'l Workshop Data Mining and Knowledge Discovery. May 2000.

[18] RuleQuest, RuleQuest Research Pty Ltd, available at http://www.rulequest.com/download.html. November 2007, viewed on April 2008.

[19] T. Scheffer, "Finding Association Rules that Trade Support Optimally Against Confidence" in proceedings of the 5th European Conference on Principles and Practice of Knowlege Discovery in Databases(PKDD'01). - Freiburg, Germany : Springer-Verlag, pp. 424-435. September 2001

[20] K.A. Smith, F. Woo, R. Ibrahim, and V. Ciesielski, "Matching data mining algorithm suitability to data characteristics using a selforganising map", Hybrid Information Systems, Heidelberg : Physica-Verlag, pp. 169–180, 2002.

[21] J. Wang, J. Han and J. Pei "Closet+: Searching for the Best Strategies for Mining Frequent Closed Itemsets," in Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, August 2003.

[22] I.H. Witten and Frank E., "Data Mining: Practical Machine Learning Tool and Technique with Java Implementation," San Francisco : Morgan Kaufmann, 2000.

[23] M. J. Zaki and C.J. Hsiao, "CHARM: An efficient algorithm for closed association rule mining," Computer Science Department ; Rensselaer Polytechnic Institute. - New York, October 1999.

[24] M. J. Zaki and C-J. Hsiao, "Efficient Algorithms for Mining Closed Itemsets and Their Lattice Structure," IEEE Computer Society, vol. 17, April 2005.

[25] M. J. Zaki, "Scalable algorithms for association mining," IEEE Transaction on Knowledge and Data Engineering, vol. 12. May 2000.

[26] M.J. Zaki, "Generating Non-Redundant Association Rules," in Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining. August 2000.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

ALEACH: Advanced LEACH Routing Protocol for Wireless Microsensor Networks

Md. Solaiman Ali, Tanay Dey, and Rahul Biswas

Department of Computer Science & Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
solaiman_kuet@yahoo.com, dtanay2004@yahoo.com, rahul_kuet@yahoo.com

***Abstract*— A wireless network consisting of hundreds or thousands of cheap microsensor nodes allow users accurately monitor the characteristics of the remote environment or detect an event. As the sensor nodes have limited energy resources, so the routing protocol designed for the wireless sensor networks should be energy efficient and provide low latency. For this reason, we propose advanced low-energy adaptive clustering hierarchy (ALEACH), a clustering-based protocol architecture where nodes make autonomous decision without any central intervention. ALEACH proposes a new cluster head selection algorithms that enables selecting best suited node for cluster head, algorithms for adaptive clusters and rotating cluster head positions to evenly distribute the energy load among all the nodes. Simulation results show that ALEACH can improve system life time and energy efficiency in terms of different simulation performance metrics.**

***Index Terms*— Wireless microsensor networks, set-up phase, steady-state phase, threshold, current state probability.**

I. Introduction

In wireless system consisting of numerous cheap microsensor nodes, the main target is to design energy efficient routing protocol for the sensor networks where nodes sense data from the environment, process the collected data and routes those data to the base station. Sensor networks have been receiving significant attention due to their potential applications in environmental surveillance, military operations, and other domains. In such networks, each sensor not only serves as a host to generate sensed data and to process the collected data, but also as a router to transmit messages and receive messages from other sensors within its transmission range [1].

In the design of wireless system, there are two key resources such that communication bandwidth and energy. That are limited than the tethered network environment [2]. In wireless microsensor networks can contain different amount of sensing nodes. Eventually, the data being sensed by the nodes in the network must be transmitted to a control center or the base station where the end user can access the data. Since all nodes can transmit/receive data, nodes' energy decreases. The main constraint of the sensor node is their low finite energy. Therefore, energy efficiency in the design of routing protocols for sensor networks is of paramount importance. Our protocol is based on the clustering technique and follows the following two considerations: firstly- maximizes the lifetime of the network - the lifetime is taken to be the time at which the first node runs out of energy. Since sensor networks need to self configure in many situations. Secondly- wireless sensor networks are energy limited. Instead of enforcing transmission energy constraints on every individual node, the total energy consumption of the network should be considered [3]. The cluster head node collects data from all its neighbor nodes, then aggregate and send those data to the base station (BS). In our protocol we devise a new technique to select the most energy effective nodes as cluster heads in every round. So the nodes' die rate will decrease and eventually it will increase network life time.

The remaining part of the paper is organized as follows: related work is briefly discussed in Section II. In Section III we discuss the network and radio models. Section IV describes ALEACH routing protocol in detail. In Section V, we simulate the proposed protocol by using NS-2 simulator and compare its performance with other prevalent routing protocols. Finally, some conclusions are drawn in Section VI.

II. Related Work

In [4], each node computes the quotient of its own energy level. With this value each node decides if it becomes cluster head for this round or not. High energy nodes will more likely become cluster head than low energy nodes. The disadvantage of this approach is, that each node has to estimate the aggregate remaining energy in the network since this requires additional communication with the base station and other nodes.

Recently, there has been much work on "power-aware" routing protocols for wireless networks [5], [6]. In these protocols, optimal routes are chosen based on the energy at each node along the route. Routes that are longer, but which use nodes with more energy than the nodes along

the shorter routes are favored, helping avoid "hot spots" in the network. In ALEACH, we use randomized rotation of the cluster head positions to achieve the same goal.

A method of choosing routes is to use "minimum transmission energy" (MTE) routing [7], [8], where intermediate nodes are chosen such that the sum of squared distances is minimized. This approach ignores the energy dissipated in the radio to send and receive the data and, therefore, may not actually produce the lowest energy routes.

LEACH [2] is a self-organizing, adaptive clustering protocol that uses randomization to distribute the energy load evenly among the sensors in the network. In LEACH, the nodes organize themselves into local base station or cluster-head according to probability equation which does not depend on the present condition (energy state) of the nodes'. For this reason in a cluster, a node having much energy may not become cluster-head for that rounds but on contrary, a node having minimum energy may become cluster-head for that round. As usually, this node will drain its rest energy in that round and will die in the next round due to insufficient energy. This is the main pitfall of this protocol.

III. Network and Radio Models

A. The Network Model and Architecture

The foundation of ALEACH lies in the realization that the base station is a high-energy node with a large amount of energy supply. In this paper we assume that a sensor network model [2], [9] with the following properties:

- A fixed base station is located far away from the sensor nodes.
- The sensor nodes are energy constrained with a uniform initial energy allocation.
- The nodes are equipped with power control capabilities to vary their transmitted power.
- Each node senses the environment at a fixed rate and always has data to send to the base station.
- All sensor nodes are immobile.

B. The Radio Model

In our analysis, we use the same radio model discussed in [2]. We assume the transmitter dissipates energy to run the radio and electronics, as shown in Fig. 1. The transmit and receive energy costs for the transfer of a p-bit data message between two nodes separated by a distance of r meters is given by Eqs. 1 and 2 respectively.

$$E_T(p,r) = E_{Tx}p + E_{amp}(r)p \quad (1)$$

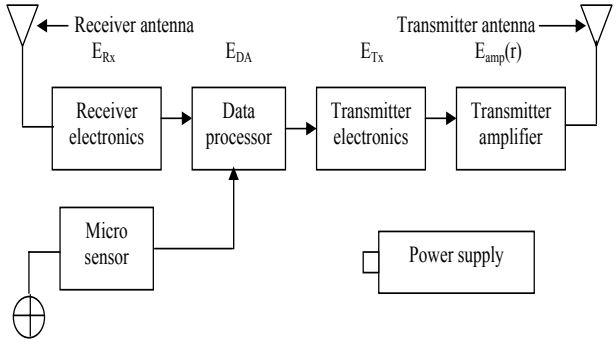


Fig. 1. Radio energy dissipation model

$$E_R(p) = E_{Rx}p \quad (2)$$

Where $E_T(p,r)$ in Eq. 1 denotes the total energy dissipated in transmitter of the source node, and $E_R(p)$ in Eq. 2 represents the energy cost incurred in the receiver of the destination node. The parameters E_{Tx} and E_{Rx} in Eqs. 1 and 2 are the per bit energy dissipations for transmission and reception respectively. $E_{amp}(r)$ is the energy required by the transmit amplifier to maintain an acceptable signal-to-noise ratio in order to transfer data messages reliably. As in [2], both the free space and the multi path fading channels models are used, relying on the distance between the transmitter and receiver.

IV. Proposed ALEACH Protocol Architecture

ALEACH forms clusters by using a distributed algorithm, where nodes make autonomous decisions without any centralized control. The advantages of this approach are that no long-distance communication with the base station is required and distributed cluster formation can be done without knowing the exact location of any of the nodes in the network. In addition, no global communication is needed to set up the clusters and nothing is assumed about the current state of any other node during cluster formation. The goal is to achieve the global result of forming good clusters out of the nodes, purely via local decisions made autonomously by each node. The operation ALEACH is divided into rounds. Each round begins with a set-up phase when the clusters are organized, followed by a steady-state phase when data are transferred from the nodes to the BS via through their respective cluster-head.

In ALEACH, the nodes organize themselves into local cluster, with one node acting as the cluster-head. All non cluster-head nodes transmit their data to the cluster-head, while the cluster-head node receives data from all the cluster members, performs signal processing functions on the data (e.g., data aggregation), and transmit data to the remote BS. Since being a cluster-head node is much more

energy intensive than being a non cluster-head node, so appropriate selection of cluster-head in every round may cause a sound affect on the network. Again, in selecting cluster-head, we have to consider the nodes current state energy. In this way, the energy load being a cluster-head is evently distributed among the nodes.

The following sections describe the cluster head determination and steady-state operation of proposed ALEACH protocol.

A. Cluster Head Determination Algorithm

Our propose ALEACH algorithm selects a certain number of clusters during each round using distribute algorithm without central intervention. We want to try evenly distributing the energy dissipation among all the nodes in the network so that there are no overly-utilized nodes that will run out of energy before the others. This will maximize the time until the first node death. As being a cluster-head node is much more energy-intensive than a non-cluster-head node (since the cluster-head node must receive data from all the nodes in the cluster, perform signal processing functions on the data and transmit the data to an end-user who may be far away), evenly distributing the energy load among all the nodes in the network requires that each node takes its turn as cluster-head. Therefore, apt cluster head determination algorithm should be designed such that nodes are cluster-heads approximately the same amount of time, and in a cluster, a node having much energy compared with the other nodes should be cluster head for that round, assuming all nodes start with the same amount of energy. Finally, we would like the cluster-head nodes to be spread throughout the network, as this will minimize the distance the non-cluster-head nodes need to send their data.

From LEACH, the uniform threshold equation is

$$T(n) = \frac{k}{N - k(r \ mod \ \frac{N}{k})} \tag{3}$$

Where, k= Expected number of cluster heads in a round
N= Total number of nodes in the networks
r= Any round

According to LEACH, at the start of any round a node calculates its threshold value from Eq. 3 must be greater than a randomly generate value range from 0 and 1, for becoming a cluster-head in a cluster. The threshold value calculated from Eq. 3 depends only on the round but does not depend on its current energy which represents its present condition.

In our propose protocol ALEACH, we improve the threshold equation by introducing two terms: General probability (G_p) and Current State probability (CS_p). The threshold equation of a node for the current round depends on both terms. So,

$$T(n) = G_p + CS_p$$

Here, $G_p = \frac{k}{N - k(r \ mod \ \frac{N}{k})}$

Previously we have explained, how performance may be decreased if we calculate the threshold considering only the Gp. For this reason, in our propose protocol we introduce a new term current state probability.

If the nodes in a cluster having different amount of energy at the same time, then the node with the highest energy should be cluster-head to ensure that all nodes die at approximately the same time. This can be achieved by setting the probability as a function of node's current energy $(E_{current})$ relative to the initial energy (E_{n-max}) in the networks, multiplying by the percentage of clusters $(\frac{k}{N})$ in the network.

so,

$$CS_p = \frac{E_{current}}{E_{n-max}} \times \frac{k}{N} \tag{4}$$

As the threshold equation does not rely only on the general probability means that the threshold value expresses more realistic view.

Final Threshold equation is,

$$T(n) = G_p + CS_p$$

$$= \frac{k}{N - k(r \ mod \ \frac{N}{k})} + \frac{E_{current}}{E_{n-max}} \times \frac{k}{N}$$

The cluster-heads in ALEACH act as local control centers to coordinate the data transmissions in their cluster. The cluster-head node sets up a TDMA schedule and transmits this schedule to the nodes in the cluster. This ensures that there are no collisions among data messages and also allows the radio components of each non-cluster-head node to be turned off at all times except during their transmit time, thus minimizing the energy dissipated by the individual sensors. After the TDMA schedule is known by all nodes in the cluster, the set-up phase is complete and the steady-state operation (data transmission) can begin.

The Set-up Phase Algorithm

The set-up phase algorithm uses the following symbols:
• AL= set of alive nodes in the network
•$Random(0, 1)$= generate random number between 0 and 1
•$T(i)$= generate threshold value of node i
•T_R= total rounds

•$Broadcast\_Cluster(i)$= broadcast cluster announcement message for cluster head i
•$Wait\_join(t)$= wait for t time unit for join request messages
•$TDMA\_send(i)$= create $TDMA$ schedule for cluster head i and send to cluster members
•$Send\_JRequest$= send Join Request to chosen cluster head

for round =0 to T_R
  for every node $i \in AL$
    if node i was cluster head in round then
      $T(i) = 0$
    end if
    if $T(i) > Random(0, 1)$ then
      $Broadcast\_Cluster(i)$
      $Wait\_join(t)$
      $TDMA\_send(i)$
    end if
    else
      $wait(t_i)$ for getting cluster head announcement
      $Send\_JRequest$ after getting announcement
    end else
  end for
end for

B. Steady-State Phase

The steady-state operation is broken into frames where nodes send their data to the cluster-head at most once per frame during their allocated transmission slot. Each slot in which a node transmits data is constant, so the time for a frame of data transfer depends on the number of nodes in the cluster. While the distributed algorithm for determining cluster-head nodes ensures that the expected number of clusters per round is k, it does not guarantee that there are k clusters at each round. In addition, the set-up protocol does not guarantee that nodes are evenly distributed among the cluster-head nodes. Therefore, the number of nodes per cluster is highly variable in ALEACH and the amount of data each node can send to the cluster-head varies depending on the number of nodes in the cluster.

To reduce energy dissipation, each non-cluster-head node uses power control to set the amount of transmits power based on the received strength of the cluster-head advertisement. Furthermore, the radio of each non-cluster-head node is turned off until its allocated transmission time. Since all the nodes have data to send to the cluster-head and the total bandwidth is fixed, using a TDMA schedule is efficient use of bandwidth and represents a low latency approach in addition of being energy efficient.

The Steady-State Phase Algorithm

The steady-state phase algorithm uses the following symbols:
•CH= set of Cluster Head
•T_{round}=total Time in a Round
•$RD(t)$= received data from cluster members for $t_{schedule}$ seconds

if $i \in CH$ node
    while $t < T_{round}$
      $RD(t)$
      Compute data aggregation and send to BS.
    end while
    if $t > T_{round}$
      call cluster set-up algorithm.
    end if
end if
else
    sleep for $t_{slot\_for\_node\_i}$ seconds.
    while $t < T_{round}$
      transmit data to the cluster member.
      Sleep for $t_{schedule}$ seconds.
    end while
    if $t > T_{round}$
      call cluster set-up algorithm.
    end if
end else

The cluster-head must be awake to receive all the data from the nodes in the cluster. Once the cluster-head receives all the data, it performs data aggregation to enhance the common signal and reduce the uncorrelated noise among the signals. Assuming perfect correlation, such that all individual signals can be combined into a single representative signal. The resultant data are sent from the cluster-head to the BS. Since the BS may be far away and the data messages are large, this is a high energy transmission.

V. Performance Simulations

A. Simulation Description

1) Simulation Environment: In order to analyze the performance of the proposed algorithm, we run the simulation under the ns-2 simulator [10] named as ns allinone 2.27 testbed with a mit [11] wireless sensor package. The simulation parameters are listed in Table 1: We run the simulation 10 times to achieve a 95 percent confidence interval for the results.

2) Simulation Metrics: To compare the performance of the our proposed routing algorithm with the prevalent ones we measure the following metrics:
a. Number of Data Messages: The amount of data messages metric determines how many data messages

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Simulation Area (x,y)	$1000\times$ 1000 m^2
Each Node Starts with	2 joules of energy
Simulation Time	3600 seconds
Base Station Location	$(50, 175)$
Number of Nodes	100
Desired Number of Clusters	5
Round Time	20 seconds
Confidence Interval	95%

are received to the base station from the network. The more amount of data messages received at the BS reveals less die rate of nodes and expenses of energy. So, it intensively related to the performance of the network.

b. Number of Alive node: The performance of a network depends on the lifetime of its nodes. If the lifetime of the nodes is high then the network performs well and also transmits more data to the base station.

c. Energy: This matric greatly affect the network as the lifetime of a node and how much data being transmitted by the node depend on the energy level of the node. So the low expenses of energy will incur great network performance.

B. Results and Analysis

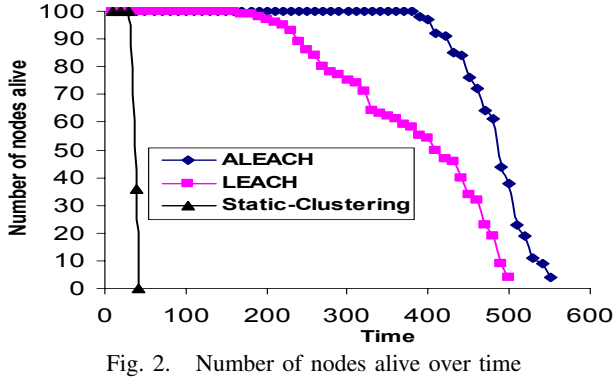


Fig. 2. Number of nodes alive over time

1) Sensitivity to Time: Fig. 2 and Fig. 3 are the corresponding graph to the number of nodes alive and the number of data messages received at the BS over time respectively. In Fig. 2, as ALEACH selects energy effective nodes in every round as its cluster heads, so its first node alive more time comparing with other protocol shows the better performance. Fig. 2 also shows that Static-Clustering and LEACH perform poorly as the cluster heads die quickly, So the node dying rate in the case of ALEACH is less compared to the ALEACH and MTE. Fig. 3 shows that total number of data messages received at the BS over time in the case our proposed ALEACH is less than other compared prevalent protocols (LEACH and

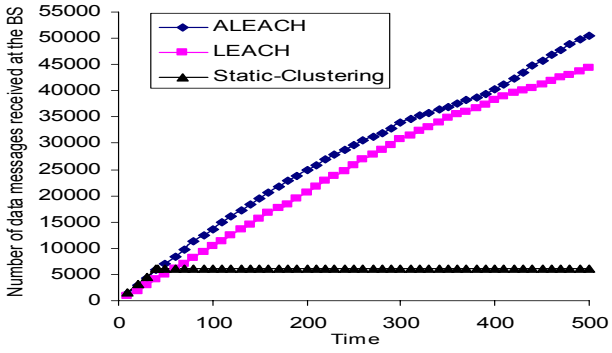


Fig. 3. Number of data messages received at the BS over time

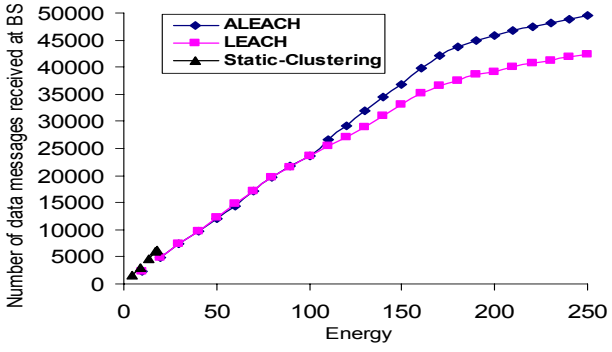


Fig. 4. Total amount of data received at the base station(BS) per given amount of energy

Static-Clustering). In all the protocols (ALEACH, LEACH and Static-Clustering), each message is transmitted over a single hop, to the cluster head, where data aggregation occurs. The aggregate signal is sent to the BS, greatly reducing the amount of data transmitted. As the nodes die rate in ALEACH (Fig. 2) is less, so more data messages will be sent to the BS comparing with other compared protocols.

2) Sensitivity to Energy: Fig. 4 shows the total data messages received at the BS at the expense of specific amount of energy. This graph shows that ALEACH delivers the most data messages per unit energy, achieving both energy and latency efficiency. Routing protocol MTE does not provide local aggregation to reduce the amount of data that needs to be transmitted to the BS. Both LEACH and ALEACH reduce data using data aggregation and expected number of clusters per round, k=5. But in ALEACH, we use CS_p, so the probability of having five clusters in each round is more in ALEACH than LEACH.

3) Sensitivity to Number of Data Messages: The simulation curve of Fig. 5 shows the number of nodes alive with respect to number of data messages received to the BS. The life time of a network is high when nodes die rate is less and vice versa(Fig. 2) and nodes send data messages at the BS as time goes by(Fig. 3). So we can infer that more alive node rate will send more data messages to the BS. In Existing LEACH and Static-Clustering the nodes

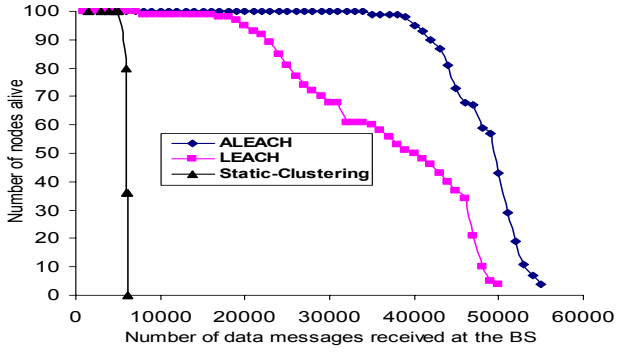


Fig. 5. Number of nodes alive per amount of data sent to the BS

die very quickly than ALEACH. As a result number of nodes alive with respect to data received in ALEACH is higher than the Existing LEACH and Static-Clustering.

VI. Conclusions

In this paper, we devise a new technique to select the cluster heads in every round which depends both on current state probability and general probability. In performance simulations, we compare our protocol with prevalent protocols LEACH and Static-Clustering. As we select the most eligible nodes as the cluster-heads in terms of its current state and general probability, so the nodes' die rate is less than the other compared protocols. This incurs great impact incase of data messages reception at the base station and energy loss of the nodes.

Since ALEACH appears to be a promising protocol, there are some areas for improvement to make the protocol best suited everywhere. As the sensor nodes have limited energy, so the nodes die after a certain time limit. Future work directions are to take the nodes as solar aware nodes which regain energy it selves, so that our protocol will be more energy sufficient. And another provision is to make our protocol in to hierarchal protocol by forming "super clusters" out of the cluster head nodes and super clusters will process all the data from the cluster heads. So the energy loss of every cluster heads will be reduced to send the data messages to the BS and aggregation. This improvement will make ALEACH effective for a wider range of wireless microsensor networks.

References

[1] Weifa Liang and Yuzhen Liu, "Online Data Gathering for Maximizing Network Lifetime in Sensor Networks", *IEEE Transaction on Mobile Computing*, vol. 6, no. 1, pp. 2-11, January 2007.

[2] Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", *IEEE Transaction on Wireless Communications*, vol. 1, no. 4, pp. 660-670, October 2002.

[3] Liang Song and Dimitrios Hatzinakos "Cooperative Transmission in Poisson Distributed Wireless Sensor Networks: Protocol and Outage Probability", *IEEE transactions on wireless communications*, vol. 5, no. 10, pp. 2834-2843, October 2006.

[4] M. J. Handy, M. Haase, D. Timmermann " Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-Head Selection", *Mobile and Wireless Communications Network*, pp. 368-372, 2002.

[5] S. Park and M. Srivastava, "Power aware routing in sensor networks using dynamic source routing,", *ACM MONET Special Issue on Energy Conserving Protocols in Wireless Networks*, 1999.

[6] S. Singh, M. Woo, and C. Raghavendra, "Power-aware routing in mobile ad hoc Networks", *Proc. 4th Annual ACM/IEEE Int. Conf. Mobile Computing and Networking (MobiCom)*, pp. 181-190, Oct. 1998.

[7] M. Ettus, "System capacity, latency, and power consumption in multihop-routed SS-CDMA wireless networks", *Proc. Radio and Wireless Conf. (RAWCON), Colorado Springs, CO*, pp. 55-58, Aug. 1998.

[8] T. Shepard, "A channel access scheme for large dense packet radio networks", *Proc. ACM SIGCOMM, Stanford, CA*, pp. 219-230, Aug. 1996.

[9] S.D. Muruganathan, D.C.F. Ma, R.I. Bhasin and A.O.Fapojuwo, A centralized energy-efficient routing protocol for wireless sensor networks, *IEEE,Communications Magazine*, vol. 43, issue 3,pp.8 - 13, Mar. 2005.

[10] The network simulator-2 [Online]. Available: http://www.isi.edu/nsnam/ns/

[11] MIT package and pdf files:-leach_doc.pdf and ns2leach.pdf for wireless sensor networks. "The MIT uAMPS Package Version 1.0" August 7, 2000 [Online]. Available: http://www.ece.rochester.edu/research/weng/code/leach/

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Publish/Subscribe based Reprogramming of Sensor Networks

Sazia Parvin
Dhaka University
Dhaka, Bangladesh
saziap@yahoo.com

Abstract

Remote reprogramming of sensor nodes through wireless channel is a very important function for efficient sensor networks management. Now-a-days secure reprogramming protocols use asymmetric cryptography schemes which are in need of many computational cost and large memory. In this paper an efficient authentication strategy based on the publish/subscribe scheme for reprogramming of sensor networks is presented. By means of integrating this scheme with previous reprogramming protocol MNP, this strategy can ensure security of reprogramming at lower cost.

1. Introduction

Wireless Sensor Networks (WSN) consists of a large number of wireless communicating ultra small autonomous devices, called sensor nodes, which are powered with low powered battery and equipped with integrated sensors. In typical application scenarios, sensor nodes are spread randomly to collect sensor data depending on the query. Sensor nodes are deployed in hostile environment. Because of the aloofness of some nodes due to the remote access, remote reprogramming of sensor nodes through the wireless channel is a very important function for efficient sensor network management. The software image (updated code) is typically propagated through epidemic style (hop-by-hop) in most of existing traditional reprogramming protocols [2-3]. The remote attestation is one of the core functionalities provided by trusted computing platforms. The recent researchers [10] make use of remote software-based attestation to detect and recover compromised nodes with the checksum of executed code on the nodes; however this technology itself does not prevent attacks, it can detect the compromised nodes and recover (reprogramming) tampered software after the fact. As wireless sensors are deployed once and are intended to operate unattended for a long period of time, one of the key challenges is to manage WSNs in such a way that they can be dynamically customized to various circumstances. Since resource limitations prevent sensors from having an extensive set of services pre-installed, sensor software should be dynamically reconfigurable.

The secure reprogramming researchers [4-6] assume that there exists a public/private key pair. That means a base station has private key and each sensor node has the base station's corresponding public key. This PKI based scheme needs many computational overhead, large memory size and may be not suitable for wireless senor network (WSN) as WSN has many resource constraints. In this paper Event-Based Middleware has been proposed for reprogramming in WSNs. The Event-Driven mechanism is based on the publish/subscribe paradigm. We propose an efficient authentication strategy based on publish/subscribe scheme for reprogramming of sensor networks. Depending on the publish/subscribe scheme, the base station (publisher) will publish a message about the updated code. The interested sensor nodes (subscriber) will send subscriptions to base station depending on their interest. A parameter is *r* is transmitted from the base station to sensor node. The private key K_{Code} is computed with procedure $Verify(r)$ at base station and node separately for code propagation. In this paper four phases are: publish-subscribe-authentication-data handshake used.

The reminder of this paper is as follows: we introduce several related works in section 2. In section 3, we show our system model. We present our approach in section 4. In section 5, we conclude this paper.

2. Related Work

In this section, we present related works in the area of code propagation, reprogramming and brief description on publish/subscribe based middleware approach.

The code propagation is the major component of network reprogramming, which is responsible for efficiently and reliably forwarding software image to

one or more sensor nodes. MNP [3] applies an advertise-request-data handshake interface. This approach is designed for state machine. MNP are such protocols which transferring the code image in chunks of pages on hop-by-hop basis to all nodes in a network. Recent works on secure networks reprogramming protocols in sensor networks are described in [4-6]. Sluice [4] uses a combination of a one-way page-level hash chain and a digital signature to authenticate the source of a code-update image, employs a digital signature to authenticate the contents of the first page and the pages are verified recursively at the page-level hashes embedded in the previous pages. Deluge [9] uses a one-way hash chains to verify authenticity of packets (rather than pages). Deng et al. [7] proposes a protocol similar to Deluge for secure code dissemination in sensor networks. SWATT [10] is a software-based memory attestation technique that allows an external verifier to perform an equality check on the memory contents of an embedded device; the verifier can detect the attacker's modifications by measuring the actual time taken by the routine to generate the checksum and comparing it with the expected execution time. In [1] several reprogramming approaches are discussed and their limitations established. There are lots of unsolved problems that need further investigation to make reprogramming highly usable and efficient. But because of signal collisions and interference, the reprogramming protocol design is more challenging. In this paper we define an Event-Driven Approach (EDA) for WSNs applications that provides its support services in the middleware, node application and network levels. Our main objective is reprogramming in WSNs through applying Event-Driven Approach. We use publish/subscribe scheme [6] as a core architectural model for providing the essential services as middleware and adapt this core to satisfy the requirements and constraints for reprogramming of WSNs. The main purpose of middleware is to support the development, maintenance, deployment and execution of applications, filling in the gap between the application layer and the hardware, operating system and network stack layers. In the case of a WSN, this includes mechanisms for formulating complex high-level sensing tasks, communicating them to the WSN, coordination of sensor nodes to split tasks and distribute them to the individual nodes, data fusion for merging the sensor readings into high-level result, and reporting it. The key elements in the publish/subscribe paradigm are the *Notification Service*, *Subscribe Matching Service* and the *Subscription Data Store* [6]. A more recent work has been carried our specifically in the WSNs are by Mires [12]. In Mires, the publish/subscribe scheme in the middle ware layer of WSNs is provided to facilitate the deployment of event-based applications. Our intension in this paper is to use publish/subscribe based middleware to publish the events about the new software image to sensor nodes and the interested sensor nodes will send interest to base station which is called subscription. Depending on these subscriptions, the updated code will be propagated to sensor nodes.

3. System and Treat Models

3.1 System Model

The conventional reprogramming protocol system model for sensor networks is depicted in figure 1, in which the code images are propagated from base station to every sensor node in the network [7]:

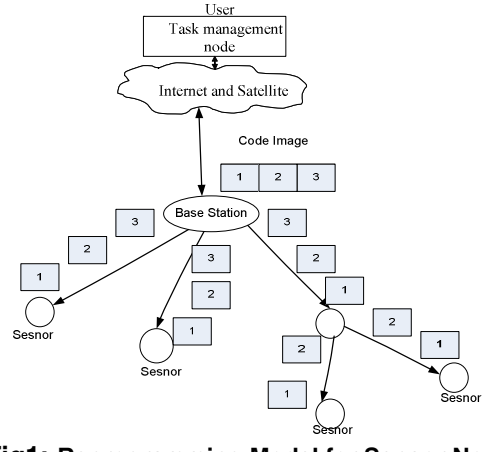


Fig1: Reprogramming Model for Sensor Nodes

▪Entire code image is segmented into pages; each page contains a certain number of packets.

▪Page level: a node has to receive page by page sequentially.

▪Packet level: a node broadcasts all packets in a page in "Round-Robin" way. After receiving ACK/NACK messages from neighbor's nodes, it broadcasts lost packets and so on until all packets are received by neighbors.

Our focus in this paper is to ensure the security of reprogramming for a sensor network after its initial deployment. In this model we assume that:

▪As researchers [4-6], all nodes in a WSN have a same running code image.

▪The base station has the entire code image running on nodes, and can compute the code attestation of the

sensor with a memory content verification procedure *Verify(r)* same as [10].

▪A sensor node contains a verification procedure *Verify(r)* same as base station, and can send data to its neighbor nodes directly (one-hop) or through multi-hop

▪The communication is bidirectional.

▪A publish/subscribe based middleware layer mechanism will be established in WSNs. Constructing our chosen structure to EDA, we can infer that base station play the role of event sources (ES) in event-driven mechanism, and the task of subscribing and notification is assigned to sensor nodes, thus we name these nodes *Event Broker* (EB).

3.2 Threat Model

In this section, we mainly focus the adversary that tries to attack the system through capturing and compromising some sensor nodes. In this section we assume that:

▪The base station can not be compromised by attackers as it is physically protected or has tamper-robust hardware [8] and enough resources. So an adversary can't compromise the base station.

▪Authentication broadcasting and flooding is good to prevent Hello flood attack.

▪An adversary can eavesdrop on any communication in the network, inject an arbitrary number of corrupt or updated code images into the system, compromise a sensor and acquire information inside it, but cannot modify the hardware of the system.

4. Proposed Approach

4.1 Establishment of publish/subscribe scheme in WSNs:

The event-driven mechanism is based on the publish/subscribe paradigm. According to publish/subscribe paradigm, a user expresses his/her interest in receiving certain types of events by submitting a predicate, called the user's subscription [6]. When a new event is generated and published to the system, the publish/subscribe infrastructure is responsible for checking the event against all current subscriptions and delivering it efficiently and reliably to all users whose subscriptions match the event.

The key elements in publish/subscribe paradigm in are the *Notification Service*, *Subscription Matching Service* and the *Subscriptions Data Store*.

At first the base station will publish an event (message about new version of software image) against to all sensor nodes. After getting the response from subscribers (sensor nodes), the Subscription Matching service is responsible to check the published events against the subscriptions issued by subscriber (sensor nodes) to forward the new version of software message towards interested nodes.

In this scheme subscriptions are defined based on events' (message) contents. For each message, the content of message is checked against the content of subscriptions reported from interested sensor nodes and registered in Data Store (where the subscriptions are preserved and messages are queued before they are passed to subscribers). If the result of matching becomes true, then the event (new software image) is forwarded to interested sensor nodes via *Notification Service*.

4.2 Services in Event Broker

Event Broker has three important parts: Notification Service, Subscription Checker and Subscriptions Storage.

When a new program code is generated to replace the old one, one message is published from the base station against all sensor nodes. According to sensor's interest, the sensor will submit a predicate for receiving the new program and these predicates are stored in Subscription Storage. The Subscription Checker checks the content of the message against the interests stored in Subscriptions Storage.

The output of subscription checker is a list of sensor nodes that are interested in receiving the new program code. After finding the interested sensor nodes, the new software code is forwarded through code propagation to interested sensor nodes by Notification service in turn.

4.3 Code Signing

As references by [5] before system deployment, each sensor is preinstalled a same key K_{Sensor} ; the sensor's K_{Sensor} is shared with base station, and is generated by pseudo-random function.

$$K_{Sensor} = f_{K^s}(Random)$$

Code signing is a common technique for authenticating the source and verifying the integrity, we sign a software image with the one-way hash chain, and the signing method is similar to [11] for signing digital streams. In this strategy, as referenced by [5] symmetric key K_{Code} is computed from verification procedure $Verify(r)$ of current software image

running on sensors acts as an encrypted key at base station; at sensor side, K_{Code} (Computed from $Verify(r)$ in sensor) acts as decrypted key.

In order to maintain dynamics characteristics of code signing, a parameter *r* is published against the sensor nodes from the base station before K_{Code} is computed and *r* is encrypted with K_{Sensor}.

The $Verify(r)$ using RC4 is same as literature [11].

4.4 Code Propagation in Detail

When a new program code is generated to replace the old one, one message is published from the base station against all sensor nodes. According to sensor's interest, the sensor will submit a predicate for receiving the new program. This is called the nodes subscription [6].

The main dissemination of code through publish/subscribe based scheme is described as follows:

1. According to the current software image running on sensors, the base station signs the new software image with methods described in 4.2.
2. The base station publishes a *message1* (events) that a new version of software image will be released to all sensor nodes after signing code.
3. After receiving the message, the sensor examines the version of the software to determine whether it receives this software image or not. If there is version inconsistent then the sensor show its interest in receiving the new program code by submitting a predicate, this is called node's subscription. This response means that this node is ready to receive the new software image through parameter *r;* else it does nothing.
4. The interested subscriptions from sensor nodes are stored in Data Store (where the subscriptions are preserved and messages are queued before they are passed to subscribers)
5. Then after checking the result from the Data Store, the base station broadcasts a *message2* containing the random parameter *r* encrypted with K_{Sensor} via notification service.
6. After receiving the *message2* the sensor decrypts the *message2* with K_{Sensor}, and achieves the parameter *r*, and then computes K_{Code} with $Verify(r)$.
7. The base station begins to broadcast packets $(p_o, p_1, p_2, p_3....)$ of software image in sequence; the sensor receives the packet p_o, it decrypts the signature using K_{Code} and verifies p_o. If the signature verification of p_o is successful, it confirms that the packet come from the base station truly and then verifies other packets.
8. After all received packets is verified, the sensor starts to execute new program through reboot.

Some packets may be retransmitted because the fallibility of wireless channel. At this time, the sensor sends request2 containing its ID and the number of retransmitted packet encrypted with K_{Sensor} to base station; after base station verifies the validity of the requesting, it retransmits the lost packet.

5. Conclusion and Future Works

In this paper an efficient authentication strategy based on the publish/subscribe scheme for reprogramming of sensor networks is presented. Our approach adopts previous MNP protocol with publish/subscribe based scheme. This approach is efficient for reprogramming in wireless sensor network as publish/subscribe scheme can keep track of message publication and subscription in an efficient way. We want to show the security of publish/subscribe based reprogramming at lower cost through simulation as a future works.

References

[1] Q.Wang, Y.Y. Zhu., L. Cheng, "Reprogramming wireless sensor networks: challenges and approaches", IEEE Networks, 2006, 20(3): 48.

[2] J.WHUi, D.E.Culler, "The Dynamic Behavior of a Data Dissemination protocol for Network programming at scale". In ACM International Conference on Embedded Networked Sensor Systems, Baltimore,, MD, USA, Nov.2004, pp. 81-94.

[3] S.S. Kulkarni,L.Wang, "MNP: Multihop Network Reprogramming Service for Sensor Networks", In Proc. of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS 2005), Columbus, Ohio, USA, Jun.2005,pp. 7-16.

[4] P.E. Lanigan, R.Gandhi, P.Narasimhan, " Sluice: Secure Dissemination of Code Updates in Sensor Networks", In Proc. of the 26th IEEE International Conference on Distributed Computing Systems, Lisboa, Portugal,Jul.2006,pp.53-53.

[5] J.Tan, J.Chen, Y.Liu, "An Efficient Authentication Strategy for Reprogramming of Sensor Networks", In Proc. of International Conference on Computational Intelligence and Security, 2007, Harbin, China, pp.833-837.

[6] A.Taherkordi, M.A.Taleghan, M.Sharifi, "Achieving Availability and Reliability in Wireless Sensor Networks Applications", In proc. of the First International Conference on Availability, Reliability and Security (ARES'06).

[7] J.Deng, R.Han, S.Mishra, " Secure Code Distribution in Dynamically Programmable Wireless Sensor Networks", In proc. of the International Conference on Information processing in Sensor Networks, Nashville, TN, USA, April 2006, pp. 292-300..

[8] E.Shi, A.Perrig, "Designing secure sensor networks", IEEE Wireless Communications, 2004 , 11 (6) : 38-43.

[9] P.K.Dutta, J.W.Hui, D.C. Chu, D.E.Culler, "Securing the Deluge network programming system", In Proc. of the Fifth Inernational Conference on Information Processing in Sensor Networks, Nashville, TN , USA ,Apr. 2006, 326-333.

[10] A.Seshadri , A.Perrig, L.V. Doorn, P.Khosla, "SWATT: SoftWare-based ATTestation for Embedded Devices", In Proc. of the IEEE Security and Privacy Conference, Oakland, CA, USA , May 2004, pp.278-282.

[11] R.Gennaro, P. Rohatgi, “How to Sign Digital Streams", Information and Computation, 2001, 165(1): 100-116.

[12] E.Souto, G.Guimaraes, G.Vasconcelos, M.Vieira, N.Rosa, C.Ferraz, " A Message-Oriented Middleware for Sensor Networks", IN Proceedings of the 2nd Workshop on Middleware for Pervasive and Ad-Hoc Computing, pp.127-134, New York, USA, 2004

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Load Balancing in DHT based P2P Networks

Md. Ahsanur Rahman

Department of CSE, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
E-mail: ah39san@yahoo.com

Abstract – Most basic DHT based peer-to-peer networks distribute objects among nodes in a way that tends to balance loads among the peers if the distribution of objects in the identifier space is uniform. But this doesn't hold in practice. The case is farther complicated by the fact that, in a typical scenario, object loads and node capacities vary enormously. Additionally, a node's load may vary greatly over time since the system can be expected to experience continuous insertions and deletions of objects, and continuous arrival and departure of nodes. This paper presents a scheme that can balance load in any DHT based P2P network even in a dynamic environment.

I. Introduction

Most of the structured P2P networks use distributed hash tables (DHT) in which the file location information is placed deterministically at certain peers. Specifically a DHT stores <file ID, value> mapping for each file in the network, where file ID denotes the globally unique identifier of the file and value represents the location of the file. Typically this <file ID, value> mapping information is distributed and stored over multiple DHT nodes. When assigning a file ID to a node, consistent hashing [1] is used so that the load tends to be evenly distributed across the nodes. When a user wants to locate a file, the exact file ID must be provided which is then hashed to obtain its location. Then the query message is routed to that location. Examples of DHT networks include Chord [2], CAN [3], Pastry [4], Tapestry [5] etc.

One of the most elegant features of DHTs is that Search is deterministic: given the key of an object, the search scheme guarantees to find the object within bounded cost. However, this scheme is useful only when we have exact file ID of the file we want, but often we have only partial description of the target file(s). So much attention has been paid to build a more flexible search service, such as search by keywords.

Keyword search in a DHT based P2P network is typically implemented through a special data structure called inverted index. An inverted index is a set of (k, F(k)) pairs where k is a keyword and F(k) is the set of files containing keyword k. Typically this (k, F(k)) mapping information is distributed and stored over multiple DHT nodes by assigning each keyword to a node through hashing. When a user searches for files associated with a set of keywords, s/he generates a query string. The first keyword in the query string is hashed to obtain the node hosting first keyword and the resulting list of files obtained from that node is then gradually intersected with the results obtained from nodes hosting other keywords in the query string to get final result. This is called chained query processing [7]. An example of this process is illustrated in Figure 1.

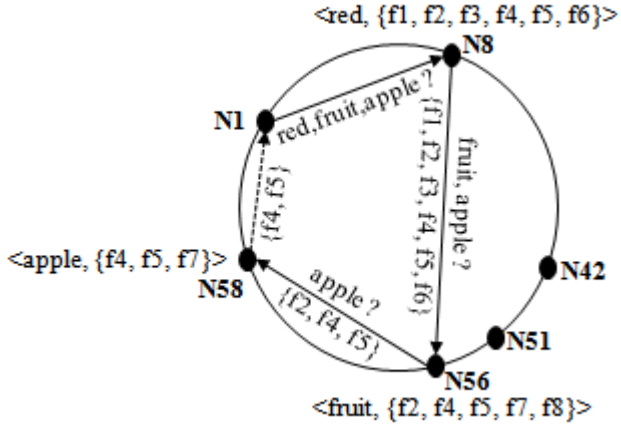


Fig. 1 Chained query processing in a Chord network

In this figure, node N1 searches files containing all of the keywords 'red' and 'fruit' and 'apple'. The node hosting 'red' is N8 which gives the output {f1, f2, f3, f4, f5, f6}. This resulting list of files is intersected with the results from N56 to get the result {f2, f4, f5}. This result is again intersected with the results from N58 to get the ultimate result {f4, f5}.

In real world corpus distribution of keyword frequency follows Zipf's law meaning that a few keywords occur very often (i.e. they are more common) while many others occur rarely [6]. As such, nodes hosting common keywords have to spend more storage than others. This makes the indexing load highly skewed among peers. This is the so called imbalanced load problem. The goal of this paper is to solve this problem by extracting loads from nodes hosting common keywords, without changing the mechanisms used in underlying DHT.

II. Proposed System

The proposed architecture uses a modified version of chained query processing to enhance performance: query processing stops immediately and returns to the query source if at any stage in the chain Ø (empty set) is found. Also, first keyword in a chain is retained in the query body until the query reaches the node hosting second keyword. This is done to determine common keywords.

Our scheme has 3 subparts: finding common keywords in the network, reducing loads from overloaded nodes and the new search scheme. The first two operations are executed periodically to cope with dynamicity of the environment. The 3 subparts are described below.

A. Finding Common Keywords

Every query source in the network caches the mapping from keyword to the size of the query result for a single keyword query in a local data structure, which we call Attenuated Inverted Index (AII). Also, each node in the network caches these mappings in its AII if it routes query results for single keyword query or query results for first keyword in a query string. For example, in figure 1 query result for keyword 'red' is routed from node N8 to node N56 via nodes N42 and N51 (according to Chord routing mechanism). So nodes N42, N51 and N56 will store the mapping <red, 6> in their AIIs. If after some time the entry <red, w> (w may be any number) is about to be inserted into the AII of any of these two nodes, then previous entry <red, 6> is deleted before this insertion. In this way, each node maintains uniqueness of keywords in its AII and remains up to date.

Let, h(k) denote the node hosting keyword k, F(k) denote the set of files containing keyword k, S(n) denote the set of keywords hosted by node n, and AII(n) denote the set of entries in AII of node n.

Whenever the size of $S(n) \cup AII(n)$ for a node n, goes beyond M (M is a tuneable parameter), n computes the median of |F(k)| for all $k \in S(n) \cup AII(n)$. Here median is used in stead of arithmetic mean to measure the central tendency of |F(k)|, because median is not attenuated by extreme values. We call the calculated median the current median and denote it by m. The keywords k (where $k \in S(n) \cup AII(n)$) for which $(|F(k)| - m) \geq \Delta$ are considered as common keywords, where Δ is another tuneable parameter.

If k is identified as a common keyword, h(k) is notified about that. Note that, If h(k) is the local node itself then there is no overhead. Otherwise the scheme will generate some network traffic. But common keywords are very rare in real world corpus and to reduce network traffic we can piggyback the notification message with outgoing messages. Also these notification messages should be cached along the path to avoid flooding of notification messages from some other nodes. Notification is required since some nodes may not reach M even after long period of time and as such don't know that they are hosting common keywords. If a node neither gets any notification nor reaches M even after sufficiently long period, then it consults with nodes that are physically closest to it.

B. Reducing Loads

Our scheme exploits the fact that, in real world corpus, very few files are annotated with only common keywords. The situation is represented in figure 2.

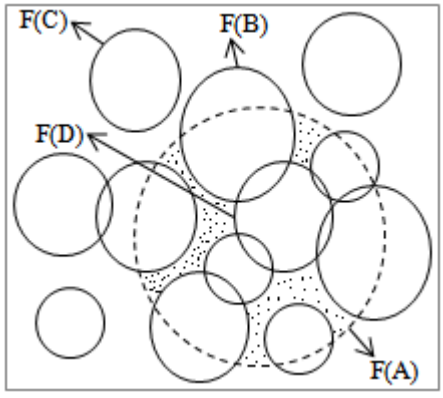


Fig. 2 Set of files annotated with keywords A, B, C, D etc. in a real world corpus

In this figure, A is a common keyword. So F(A) is very large. But there are lots of overlapping in F(A) and very few files (residing in the small dotted area) are members of only F(A).

Whenever a node comes to know that it hosts a common keyword c, it starts informing other nodes that: "c is a common keyword" using random walk [12].

At the same time, h(c) starts matching each incoming query results with F(c). It flags each matching file in F(c). Let the incoming query result for a keyword k is Q. If Q is a subset of F(c) then k is inserted into a list in h(c). We call the list I(c). Otherwise a new keyword is generated by combining k and c. For example if c is 'red' and k is 'apple' then the new keyword will be 'red&apple'. As in [7] we call such keywords 'synthetic keywords'. We assume that the function combine() is responsible for generating synthetic keywords. Note that, only the files in $Q \cap F(c)$ contain both k and c. In other words, these files contain the synthetic keyword c&k. This mapping information is inserted into a DHT node whose ID is found by hashing the synthetic keyword. The synthetic keyword is inserted into a list in h(c) which we call L(c). I(c) and L(c) are sufficiently large Bloom Filters [13] which can exactly answer whether a keyword is a member of them. The detailed algorithm is presented in figure 3.

The node h(c) iteratively calls this algorithm until the number of non-flagged files in F(c) drops below a threshold t. If it doesn't drop below t even after sufficiently long period, it consults with the nodes that are physically within ℓ hops from it (ℓ is a tuneable parameter) and calls this algorithm upon mappings obtained from them.

```
Algorithm Match (k, Q)
1    If Q ∩ F(c) ≠ ∅ then
2      Flag each file f ∈ Q ∩ F(c)
3      If Q is not a subset of F(c) then
4        x ← combine(c, k)
5        assign <x, Q ∩ F(c)> to a DHT node
6        insert x in L(c)
7      Else
8        insert k in I(c)
9      End If
10   End If
```

Fig. 3 A pseudo code for the algorithm to match F(k) with F(c)

When the number of non-flagged files in F(c) drops below t, the node h(c) deletes all flagged files from F(c). We say that h(c) is stable now.

This scheme relieves only overloaded peers. If any node feels overloaded due to its own storage limitation then it searches for node(s) capable of taking some of its loads (without making those node(s) overloaded) by randomly probing other nodes. It then handover loads to them. In such case it retains redirection pointers to them as in [8].

C. New search scheme

Before h(c) becomes stable, queries are handled as before. If h(c) becomes stable but the query source generating a query string (q) containing c isn't informed that c is a common keyword, then upon arrival of q on a node who is informed about that, informs the query source about it and requests to restart the query. Note that such a node of course exists, since h(c) will also be probed while query processing.

In the new scheme, a query source divides the query into two parts: one containing common keywords and the other containing uncommon keywords. There are 3 cases to consider:

a) If the query string contains just uncommon keywords then normal chained query processing is applied.

b) If a node generates a query for just a common keyword c, h(c) returns only F(c), assuming that the user wants to access the files annotated with just c. The user may also set certain number, z with the query string to indicate s/he wants to access minimum z files annotated with c. In that case, nodes hosting keywords in I(c) ∪ L(c) are iteratively probed from h(c) and the results found from them are intersected with the current result to maintain distinctness. This process iterates until h(c) accumulates at least z distinct files.

c) If a node generates a query string (q) containing one or more common keyword(s) and zero or more uncommon keyword(s), then the query source chooses a common keyword first. It should choose that one whose hosting node seems to be nearest from it. If the chosen node is c then the query is forwarded to h(c). h(c) searches for each keyword k ∈ q in I(c) ∪ L(c). We search for common keywords other than c first. Since there is less possibility that one file contains more than one common keyword, we can expect a mismatch. If any mismatch found then ∅ is immediately returned to the query source. Otherwise h(c) changes the query according to the algorithm in figure 4 and forwards it in the network for chained query processing.

```
Algorithm ProcessQuery (q)
1    Let, q is decomposed into 2 sets: set of common
     keywords (C) and set of uncommon keywords (U)
2    q ← ChangeQuery (q, C)
3    If q ≠ ∅ then
4      q ← ChangeQuery (q, U)
5      If q = ∅ then
6        return ∅ as the result to the query source
7    Else
8      return ∅ as the result to the query source
9    End If
10   forward q in the network for normal chained query
     processing

Algorithm ChangeQuery (q, K)
1    For each k ∈ K
2      If k ∈ I(c) then
3        continue
4      Else If k ∈ L(c) then
5        replace k by combine(k, c) in q
6      Else
7        return ∅
8      End If
9    return q
```

Fig. 4 A pseudo code for the algorithm to search for files containing keywords in q

Since c is common, typically it won't appear in q. Yet, if h(c) feels itself like a hot spot, it may replicate <c,F(c)>, I(c) and L(c) to nodes knocking it mostly, so that they can search keys in I(c) ∪ L(c) without probing h(c).

III. Conclusions and Future Work

The scheme described here has some elegant features:

a) It identifies overloaded peers without consuming any significant bandwidth.

b) It is applicable to any DHT based network.

c) It can be loaded in the peers in existing networks without altering the network.

d) It is applicable in a dynamic environment.

In future, we plan to extend this work to solve all the problems associated with keyword search in DHT based P2P networks, namely imbalanced load, hot spots, indexing overhead, failure to ensuring fault tolerance and providing object ranking etc.

IV. Related Works

Chord [2] proposed to allocate log n virtual nodes to each real node, to ensure that the number of files per node is within a constant factor from optimal. The problem with this approach is: it assumes that nodes are homogeneous (i.e. they have same storage/processing capability), and file IDs are uniformly distributed which is not typical.

Byers et. al. [8] proposed to hash newly inserted file's ID by multiple hash functions to compute a set of node IDs. The node having the smallest load in the set is chosen to store the meta-data of the file. Redirection pointers to this node are stored in all other nodes in the set to reduce searching overhead. While theoretically elegant, this scheme can't dynamically alleviate loads from peers that are already overloaded. Also, it changes the basic DHT mechanism to assign objects to nodes.

CFS [9] accounts for node heterogeneity by allocating to each node some number of virtual servers proportional to the node capacity. But CFS proposes to alleviate the overloaded nodes by removing some of its virtual servers, which may cause another node becoming overloaded and thus may result in thrashing.

Liu et. al. [7] used keyword fusion for load balancing. To alleviate the overloaded peers, they proposed to extract common keywords from hosting nodes and insert them in a distributed data structure called fusion dictionary. This scheme consumes huge storage and network bandwidth due to the maintenance of partial keyword list (another data structure). Also it fails to answer queries containing just common keyword(s).

Godfrey et. al. [10] proposed a scheme (which is basically an extension of [11]) that uses virtual server notion to solve the problem. Their scheme stores load information of the peer nodes in some directories which periodically schedule reassignments of virtual servers to achieve better balance. With this approach transfer of virtual servers incurs high overhead. Also, failure of directory node(s) may hamper the load balancing operation.

References

[1] D. Karger, E. Lehman, F. Leighton, M. Levine, D. Lewin, and R. Panigraphy, "Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the World Wide Web," In Proc. 29th Annual ACM Symposium on Theory of Computing, El Paso, pp. 654–663, May 1997.

[2] I. Stoica, R. Morris, D.R. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for Internet applications," in Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2001), pp. 149–160, 2001.

[3] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM 2001), pp. 161–172, 2001.

[4] A. Rowstron, and P. Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," in Proc. IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001), ser. Lecture Notes in Computer Science, vol. 2218. Springer-Verlag, pp. 329–350, 2001.

[5] B. Y. Zhao, J. Kubiatowicz, and A. D. Joseph, "Tapestry: An infrastructure for fault-tolerant wide-area location and routing," Tech. Rep. University of California, Berkeley UCB/CSD-01-1141, April 2001.

[6] Yuh-Jzer Joung, Li-Wei Yang, and Chien-Tse Fang, IEEE Journal on Selected Areas in Communications, vol. 25, no. 4, pp. 46 – 61, January 2007

[7] Lintao Liu, Kyung Dong Ryu, Kang-Won Lee, "Keyword Fusion to Support Efficient Keyword-based Search in Peer-to-Peer File Sharing," in Proc. 4th International Workshop on Global and Peer-to-Peer Computing, Chicago, April, 2004.

[8] John Byers, Jerey Considine, Michael Mitzenmacher, "Simple Load Balancing for Distributed Hash Tables," in the 2nd International Workshop on Peer-to-Peer Systems (IPTPS '03), Berkeley, CA, USA, February 2003.

[9] Frank Dabek, Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica, "Wide-area Cooperative Storage with CFS," in Proc. ACM SOSP, Banff, Canada, 2001.

[10] Brighten Godfrey, Karthik Lakshminarayanan, Sonesh Surana, Richard Karp, and Ion Stoica, "Load Balancing in Dynamic Structured P2P Systems," in the 23rd Conference of the IEEE Communications Society, Hong Kong, March 2004.

[11] Ananth Rao, Karthik Lakshminarayanan, Sonesh Surana, Richard Karp, and Ion Stoica, "Load Balancing in Structured P2P Systems," in Proc. IPTPS, Berkeley, CA, USA, February 2003.

[12] C. Gkantsidis, M. Mihail, A. Saberi, "Random walks in peer-to-peer networks," 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004), March 2004.

[13] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," Communications of the ACM, vol. 13, no. 7, pp. 422–426, 1970.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Generic Framework For Defining Security Environments Of Wireless Sensor Networks

Ali Nur Mohammad Noman

United International University
UIU Bhaban, Dhanmondi, Dhaka 1209, Bangladesh.
E-mail: noman@uiu.ac.bd

Abstract - At the beginning of its journey, Wireless Sensor Networks (WSNs) were not built keeping security in mind. Along with the proliferation of sensor applications, the need for security in sensor networks has been increased. However, due to the resource constrained sensor nodes as well as its unique nature of communication, it is very difficult to use traditional security solutions for sensor networks. Some initiatives have been taken to introduce security protocols for sensor networks but they are not enough and are based on different assumptions. To be more precise, it is very difficult to identify which security protocol is suitable for which kind of sensor applications. At the moment, there is an urgency to formally define all the security environments (or security classes) for sensor a network so that a sensor application can identify all the security protocols belongs to its security class. In addition, defined security classes will certainly help in building good security protocols for each security class. In this paper, we provide a generic framework to find out all the security classes for sensor networks. According to our best of knowledge, this work is the first attempt to fulfill this urgency of sensor networks.

I. Introduction

Wireless Sensor Networks (WSNs) is a rather new field in this high-techno era. Formally, wireless sensor network is a wireless network that consists of spatially distributed autonomous devices using sensor to cooperatively monitor physical or environmental situation like temperature, sound, vibration, pressure, motion, at different area [1,2]. A sensor node has three parts: a radio transceiver, a battery and a micro controller. Radio transceiver is responsible for the communication that is data transmission where as the task of a micro controller is to sense and process the data and all the powers needed for transceiver and micro controller are coming from battery.

The application of sensor networks is spreading day by day. Security was not the main concern when it was first deployed in real application. Security starts to play a big role since at present sensor networks are deployed in sensitive applications like battle field surveillance, health care applications and so on. However due to the resource limitations of sensor nodes along with the wireless communication nature of sensor networks, traditional security solutions (suitable for wired network and based on public key cryptography) can not be applied here[3,4,5]. Most of the security protocols of sensor networks are based on symmetric key cryptography and the number of those protocols is not many. Several security protocols such as SPINS, LEAP and TinySec have been built to provide security (i.e. confidentiality, authentication, integrity and so on) for sensor networks. However, all of them are built under some assumptions and following their own models. For example, the basis of SPINS [6] to provide security is to assure asymmetry though symmetric key cryptography by delaying the key disclosure time; it also assumes that all the sensor nodes are loosely synchronized. In contrast, LEAP [7] mainly relies on its key distribution techniques to ensure security for sensor applications. LEAP considers that different level of sensor nodes need different level of security and it is very much suitable for a sensor application with a hierarchical architecture. On the other hand, the approach of TinySec [8] is totally different from the previous two. It tries to ensure security for sensor applications in link layer level. Furthermore their capability to fulfill security demands also differs with each other. For example SPINS can not guard against DoS attacks where as LEAP can; LEAP can be used in large area where as SPINS can not; TinySec can successfully guard against many routing attacks in sensor networks where as SPINS cannot.

Since all those security protocols are built based on different assumptions it is really very difficult to compare all those security protocols and say which security protocol is suitable for which sensor application. If formally defined security environments (or security classes) were present for sensor networks, it would be very easy to trace which security protocols are suitable for which security environments and thus for which sensor applications. Research community of sensor networks has realized the urgency of having a defined security environment [9], but up to now, no effort has been made to formally define the security environment for sensor networks. In this paper, we try to fulfill this urgency of WSNs by introducing a generic framework to define all the security environments of WSNs. From now on we use the words security environment and security class interchangeably.

The rest of the paper is organized as follows. We discuss our proposed framework in section 2; then we demonstrate the usage of our framework in section 3 and

978-1-4244-2014-8/08/$25.00 (c)2008 IEEE

finally we draw the conclusion along with an indication of future work in section 4.

II. Our proposed framework

The basis for this work is to find out all the key factors those have significant impacts on sensor networks as well as on sensor applications. In other words, we try to identify all the key factors those are linked with security demands of sensor applications and sensor networks (e.g. confidentiality, integrity, authentication, availability, robustness, scalability, guard against attacks and so on), and their surroundings (e.g. network and routing structure, mobility of sensor nodes, and so on) as well. After all the key factors are identified, we analyze their outcomes for categorizing them into different groups. We also consider the interrelationship among those key factors during this grouping process. Then we find the possible outcomes of each group by analyzing its key factors' outcomes. Finally the outcomes of those groups are arranged in all possible ways to find all the security classes of sensor networks. The following figure clearly shows our framework in an abstract level:

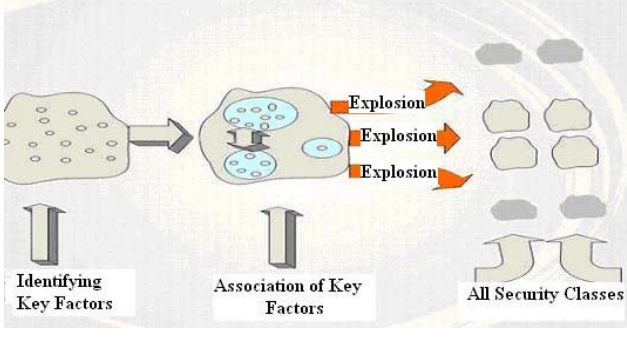


Fig. 1: Proposed framework in an abstract level

Basically, we follow the following three steps to build our framework.

1. Identify all the key factors
2. Association of key factors
3. Arranging groups in all order to find all security environments

In the following three subsections, we discuss theses three steps of our framework one by one.

A. Finding out the key factors in sensor networks

In order to find out the key factors we attentively analyze the existing security protocols of sensor networks [6,7,8,10], the demand of sensor applications[11], the set up infrastructure of sensor networks [its network and routing structure], the surroundings of sensor networks as well as sensor applications and so on. We finally come up with the following key factors list:

a. Nature of the deployment Area of the network
b. Availability of base stations in the deployed networks
c. Number of nodes
d. Mobility of nodes
e. Node properties
f. Probability of attacks
g. Network and Routing protocol
h. Confidentiality requirements for sensor applications
i. Authentication and Integrity requirements for sensor applications
j. Availability requirement for sensor applications
k. Data freshness requirement for sensor applications
l. Robustness requirement for sensor applications
m. Fault tolerance requirement for sensor applications
n. Scalability requirement for sensor applications

To be more precise, the first seven items (a to g) of the above list come from our analysis on the existing security protocols and the surroundings of sensor networks as well as sensor applications. The rest of the items (h to n) of the above list basically portraits all possible requirements of different sensor applications.

Identifying the key factors is the vital part of our framework. However due to space limitations here we give a very short argument to justify our selection of key factors. The nature of deployment area of sensor applications varies from application to application. Some application may require a huge deployment area where huge number of sensor nodes will build the network; this type of sensor networks may ask for sufficient number of base stations as well. In addition, some sensor networks (or sensor applications) ask for the facility of mobility of nodes (where a newly deployed sensor node can attach into the live network or a node can change its location) from its security protocols. Moreover a security protocol has to consider the network and routing structure of the sensor networks. Furthermore different sensor applications have different security demands- some ask for confidentiality where as some ask only authentication and integrity. Almost most of the sensor applications demand availability, robustness, scalability and data freshness from its security protocols. In addition with this, sensor networks are susceptible to various attacks mainly physical attacks (node capturing) and routing attacks. So, attacker's motivation as well as the likelihood of the occurrence of an attack (i.e. probability of attacks) is also a very dominating factor.

B. Association of key factors*:*

The main purpose of this section is to discover the relationships among the identified key factors for the security environment of sensor networks. Though all the key factors have unique impact on sensor networks, they are also interrelated with each other. This interrelationship among those key factors has been used to associate them into groups. After that, a qualitative measurement has been taken to find all the possible impacts of each group on wireless sensor network's security classes. It is very important to state that some key

factors have significant impacts on more than one group. After analyzing all the key factors, their impacts on security environments, and their interrelationships among each other, we group those key factors into the following categories:

- Security Demands
- Network and Routing Structures
- Mobility of Nodes

1. Security Demands:

Security demands are one of the important characteristics of a security environment of sensor networks. It comes as the requirements of sensor applications. The variation of the extent of security (e.g. security demand: high or medium) can be used as a means to find the difference between two different security classes for sensor networks. Several key factors have profound impacts on the process of finding out security demands for sensor applications. Some of them are helpful to find out what kind of security a sensor application ask for and the rest of them can be useful to determine in what extent they demand it. Confidentiality, integrity, authentication, and data freshness represent the first case; and availability, robustness, probability of physical attacks, scalability, and node property can portrait the latter case.

Selection of security demands: In sensor networks, assuring authentication automatically ensures the integrity and the purpose of data freshness is just to provide the assurance of proper authentication and integrity by guarding against replay attacks [12]. It is observed that almost all the sensor applications ask for authentication, and integrity, and there are some applications (such as military application) where secrecy or confidentiality is a necessity. Taking these security demands of sensor applications into considerations, we divide security environments of sensor networks into two categories:

- CIA based Security Environments (security environment with confidentiality, integrity, availability) and
- IA based security environment (security environment with integrity and availability).

Measuring the extent of security demands: Different sensor applications demand different levels of security. For example, the demand of security in battlefield application is higher than that of environment monitoring application. In addition, applications like machine monitoring, health and diagnostics monitoring, industries' production monitoring ask for high degree of accuracy and robustness. We consider availability, robustness, scalability, fault tolerance, probability of attacks, and node property in order to measure the extent of security demands for all security environments of sensor networks; no matter whether it is CIA based security environments or IA based security environments. In this case, the demand of availability and robustness, and scalability determines the extent of security demands for all security classes in sensor networks. Table 1 shows our technique to find out the overall extent of security demands for security environments of sensor networks.

Table 1 Finding out overall security demands

Assumptions made: 1.Availability & Robustness: High(H) |Moderate(M) 2. Scalability: High(H) | Moderate(M) 3. Basis of decision making: HH/MH/HM => H, MM => M		
Availability & Robustness	**Scalability**	**Overall security demands**
H	H	H
H	M	H
M	H	H
M	M	M

To summarize the above findings, it can be concluded that, when a sensor application asks for high degree of availability, robustness and scalability, it actually demands the optimal security. In general, availability and robustness are deeply connected with each other. We can link this two as follows. To remain robust a system must assure its availability in any circumstances. Most of the time, robustness and availability depend on the capability of the sensor networks to guard against different attacks and to make sure that the entire network survives with the presence of some compromised nodes(i.e. fault tolerance) [13]. From the existing research in sensor networks [14, 15], it is found that probability of physical attacks and nodes' capacity also play an influential factor to measure the extent of the demand of availability and robustness in sensor network. For instance, most of the nodes are not tamper proof and some security solutions ask for tamper proof nodes [16]. So if in a sensor application, the probability of attacks is high and nodes are not tamper proof, the demand for availability and robustness will definitely be high. Here we try to formally measure the demand of availability and robustness by considering *Probabilities of attacks*, *Nodes' capacity* and *Fault tolerance* which is shown in Table 2.

Table 2 Demand of Availability & Robustness

Assumptions made: 1. Probability of attacks: High(H) |Moderate(M) 2. Node's capacity: High (H) | Moderate (M). Here high means nodes are tamper proof and moderate means existing sensor nodes. 3. Basis of decision making: HH/MH/MM => M, MH => H		
Probability of Attacks or Fault tolerance	**Scalability**	**Demand for availability and robustness**
H	H	M (Future application)
H	M	H
M	H	M (Future application)
M	M	M

However, we observe that when probability of attacks is high, the demand for fault tolerance is also high. That is why, probabilities of attacks or fault tolerance in addition with nodes' capacity will provide the same solution.

Considering the above steps, we finally show our technique to measure the overall security demands for sensor networks' security environment in figure 2.

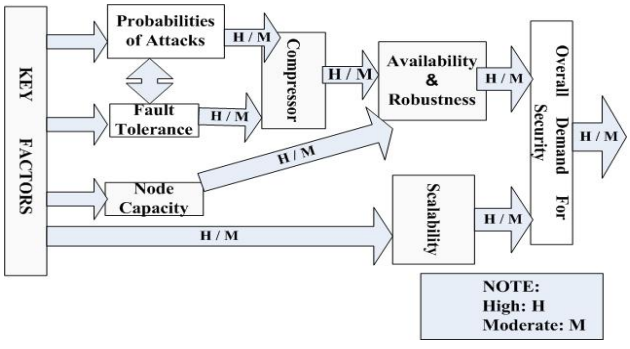


Fig. 2 Our model to find overall security demands

2. Network and Routing Structure:

We identify network and routing structure as a unique characteristic for a security environment for sensor networks. Most of the time, it can be used as a criteria to distinguish two different security environments in sensor networks. There are several key factors which can influence the behavior of network and routing structure in sensor networks. In this case, we consider the following three as main factors:-

- Nature of deployment area,
- Availability of base stations, and
- Number of nodes

However, along with main factors, mobility of nodes, probability of physical attacks, and some security demands are also considered as secondary factors to determine all the network and routing possibilities. We split the job into two parts: at first, we determine all the possibilities for network structure, and routing structure and then we combine those possibilities with some realistic assumptions in order to get all the possible outcomes for network and routing structure.

Network structure:

Most of the time, sensor network does not have any defined network structure [17]. Sometimes, a network structure is assumed to find a better routing technique. For small deployment area with adequate presence of base stations, and moderate number of nodes, we suggest star topology as the best network structure. However when deployment area widens and / or number of nodes increases and it is possible to cluster the nodes, hierarchical architecture can be considered in this case. Most of the time, it is found that sensor applications deployed in closed places can easily choose either of this two (for example hierarchical, Star) [18]. We finally propose the following options for network structure: star, hierarchical and undefined.

Routing structure:

For all most every sensor applications, multi-hop routing is chosen due to the radio link communication. However probability of physical attacks and /or the demand for top level security can force routing techniques to give more. For this type of situation, we suggest multi-path routing to make the network robust. However, multi–path routing may demand for some extra (which requires redundant path to reach to the destination). It is worth mentioning that when network area is small, and probability of physical attacks is low, multi–hop routing should be chosen instead of multi-path routing. In addition most of the time sensor applications those requires high degree of security may use multi-path routing.

Finally based on the above discussions, we reach to the following possibilities for network and routing structure of sensor networks' security environment.

(Star, multi-hop),
(Hierarchical, multi-hop),
(Undefined, multi-hop),
(Hierarchical, multi-path), and
(Undefined, multi-path).

Assumptions made:

- Most of the time, star topology does not require multi-path routing.
- Probability of physical attacks is considered very low.

3. Mobility of Nodes:

We identify mobility of nodes as a dominant factor which can uniquely distinguish several security environments. Most of the sensor applications demand that node should be mobile so that they can be joined into the network without any external help [19]. However the static nature of nodes can be assumed for some sensor applications [20].

C. Defining possible security environments for sensor networks

At first, we divide the security environments of sensor networks in two broad categories, namely CIA based security environments and IA based security environments. Except the demand for confidentiality, both of them are same. Then we narrow it down to find out all the security classes for both categories. Since the above mentioned groups of key factors can uniquely distinguish different security environments, we try to define all the security environments of sensor networks by just arranging all the groups in all possible orders. As mentioned earlier, there are three groups: security demands, network and routing structure and mobility of nodes. Their possible outcomes are summarized below:

Security demands:
High (H)| Medium (M).

Network and routing structure:
(Star, multi-hop),
(Hierarchical, multi- hop),
(Undefined, multi-hop),
(Hierarchical, multi-path), and
(Undefined, multi-path).

Mobility of nodes:
mobile | immobile

These three groups can independently make a border line between two different security classes for sensor applications. In addition, their different outcomes can indicate different characteristics of the security environments of sensor networks. So the arrangement of their different outcomes should show different types of security classes for sensor networks. Now by just doing the combination of all the possible values of those groups, we can get all possible security classes of sensor networks. The framework for defining all the security classes is shown in Figure 3.

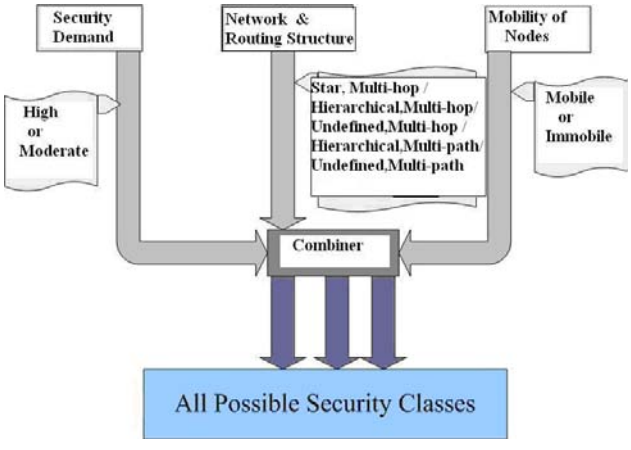


Fig. 3 Framework for finding all possible security classes of sensor networks.

From the following figure, it is clearly understood that the number of all possible security classes (or security environments) is equal to the summation of all the combinations of those groups. So there are about 20 (2*5*2) security classes. It is important to note that, to get the entire security classes, only 5 possibilities of network and routing structures are considered. In reality, there may have more than 5 possibilities. However, this framework correctly finds out all the possible security classes if the possible values of groups are correctly identified.

III. Our Framework in Action

Our framework shows biasness towards sensor applications. To be more precise, we assume the fact that a sensor application always wants to find available security protocols for its security class. To achieve so, a sensor application needs to identify its own security class as well as all the security protocols belongs to that class; after that it will compare all the security protocols and find the best one. We consider these facts while we are developing our framework.

To show the usage of our framework, we first find some typical security classes which are very common in real sensor applications. After that we show how a sensor application can identify its security class using this framework. Once a sensor application identifies its security class, it automatically knows what it wants from its security protocols. Finally we show how a security protocol can find its security class.

A. Finding out some typical security classes

In order to get some typical security classes (for both CIA and IA based security environments), we basically take a subset from all possible security classes where we make the assumption of "undefined" network structure. In this case the possibilities for network and routing structure would be: **(__, multi-hop) and (__, multi-path)**. In table 3, we show 4 typical security classes of sensor networks. All of these classes are very common and are applicable for both CIA and IA based security environments.

Table 3 Four typical security classes

Environment name	Security demands	Network and routing structure	Mobility
WSN_Env_A	High	Multi-path	Mobile
WSN_Env_B	Moderate	Multi-hop	Mobile
WSN_Env_C	High	Multi-path	Immobile
WSN_Env_D	Moderate	Multi_hop	Immobile

B. Finding out the security class for a sensor application

No matter whether a sensor application exists at the moment or not, it will find its security class using our framework given that it successfully identifies all the outcomes of the key factors. For example a sensor application with the requirements of *moderate* security demand, (undefined, *multi-hop)* network and routing structure and the support of mobility of nodes, will fall under the security class **WSN_Env_B**.

C. Finding out the security class for a security protocol

All the security classes defined by our framework are applicable for CIA and IA based security environments. A security class got from our framework implicitly express the requirements of its security protocols. For instance, WSN_Env_B security class ask for moderate security demand, *(undefined, Multi-hop)* network and routing structure and the support of mobility of nodes from its

security protocol. By further exploration of its security demand (moderate in this case), it can also enlist the requirements of availability, robustness, scalability and other important factors from its security protocol. Here we try to find the security class of a security protocol called SPINS. For this we need to know the outcomes of those groups: security demand, network and routing structure and mobility of nodes.

SPINS [6] is a security protocol which can ensure confidentiality along with authentication. It is made for resource constrained environment. However it is prone to DOS attacks. Furthermore it is not very much suitable for large scale of sensor nodes deployment. In addition, it can not show a high degree of robustness as well as fault tolerance. Along with this, SPINS can not be implemented for an environment where possibility of attacks is very high. Moreover, SPINS is made for normal sensor nodes, having the average node capacity. So it can be concluded that SPINS can fulfill the ***moderate*** level of security demands. In addition, SPINS uses multi-hop structure instead of multi-path as multi-path requires extra resources which are not the goal of SPINS. Further more SPINS supports mobility of nodes. So we can conclude that SPINS is a protocol for security class **WSN_Env_B**. In the same way, we can find the security class of any security protocol using our framework.

IV. Conclusion & Future work

.Our proposed framework is a very generic one. It can correctly find out all the security classes of sensor networks given that the outcomes of all key factors are properly identified. Using this framework, a security protocol will always find the security class which it belongs to and can trace all the security protocols of its security class if there is any. This framework can be used as the baseline for the further development of sensor networks; formally defined security classes can provide the formal specification of its security protocol which will definitely help to build good security protocols.

Correctly identifying the outcomes of key factors are the key for the success of our framework. Since our work is the first attempt to formally define all the security classes of WSNs, a future research on re-evaluating our identified key factors and their outcomes is always encouraged. In addition, we have a future plan to introduce a metrics to find out the best security protocol in a particular security class.

References

[1] Thomas Haenselman, *Sensornetworks*, GFDL, 2006.

[2] Diane Cook, and Sajal Das, Smart Environments: Technology, Protocols, and Applications, Wiley, 2004.

[3] Adrian Perrig, John Stankovic, David Wagner, "Security in Wireless Sensor Networks," Communications of the ACM, vol. 47, no. 6, pp. 53-57, June 2004.

[4] Jian Wang, Z Y XIA, and Lein HARN, "Storage optimal key sharing with authentication in sensor networks," in Proc. ISPA Workshops, 2005, pp. 466–474, 2005.

[5] Xiao Chen, Jawad Drissi, "An Efficient Key Management Scheme in Hierarchical Sensor Networks," in Proc. MASS, 2005, 2005.

[6] Adrian Perrig, Robert Szewczyk, J.D Tygar,Victor Wen and David E. Culler, "SPINS: Security Protocol for Sensor Networks, networks," in Proc. 7th International Confenrecne on Mobile Networking and Computing 2001, vol. 8, no. 5, pp. 189-199, 2001.

[7] Sencun Zhu, Sanjeev Setia, and Sushil Jajodia, "LEAP-Efficient Security Mechanisms for Large scale Distributed Sensor Networks," in Proc. ACM Conference on Computer and Communications Security (CCS '03), Washington D.C., October 2003, 2003.

[8] Chris Karlof, Nadim Sastry, and David Wagner, "TinySec: A Link Layer Security Architecture for Wireless Sensor Networks," in Proc. ACM SenSys, 2004.

[9] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," in Proc. IEEE Communications Magazine, vol. 40, no. 8, pp. 102-116, 2002.

[10] Donggang Liu, Sensun Zhu, and Sushil Jajodia, "Practical Broadcast Authentication in Sensor Networks," in Proc. MobiQuitous 2005, 2005.

[11] Jan Blumenthal, Matthias Handy, Frank Golatowski, Marc Haase, Dick Timmermann, "Wireless Sensor Networks-New Challenges in Software Engineering," in Proc. IEEE conference on ETFA'03, vol. 1, pp. 551-556, 2003.

[12] C. Karlof, and D. Wagner, "Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures, " in Proc. 1st IEEE Workshop on Sensor Network Protocols and Applications, May 2003, pp. 113-127, 2003.

[13] Bhaskar Krishnamachariy, Yasser Mourtada, and Stephen Wicker, "The Energy-Robustness Tradeoff for Routing in Wireless Sensor Networks," in Proc. IEEE Conferenece on Communications, 2003, vol. 3, pp. 1833-1837, 2003.

[14] Xun Wang, Wenjun Gu, Sriram Chellappan, Kurt Schosek, and Dong Xuan, "Lifetime Optimization of Sensor Networks under Physical Attacks," in Proc. IEEE International Conference on Communications, 2005, vol. 5, pp. 3295-3301, 2005.

[15] H. S. Ng, M. L. Sim, and C. M. Tan, "Security issues of wireless sensor networks in healthcare applications," BT Technology Journal, 2006, vol. 24 , no. 2, pp.138-144, April 2006.

[16]Anthony D. Wood, and John A. Stankovic, "Denial of Service in Sensor Networks," IEEE Computer, vol. 35, no. 10, pp. 54-62, October 2002.

[17] Basil Etefia, "Routing Protocols for Wireless Sensor Networks,".

[18] F. L. LEWIS. *Wireless Sensor Networks: Smart Environments - Technologies, Protocols, and Applications*, ed. D.J. Cook and S.K. Das, John Wiley, New York, 2004.

[19] Aman Kansal, Mohammad Rahimi, William J Kaiser, and Mani B Srivastava, "Controlled Mobility for Sustainable Wireless Networks," in Proc. IEEE SECON 2004, pp. 1-6, 2004.

[20] Ugur Cetintemel, Andrew Flinders, and Ye Sun, "Power-Efficient Data Dissemination in Wireless Sensor Networks," in Proc. 3rd ACM international workshop on Data Engineering for Wireless and Mobile Access San Diego, CA, USA, 2003, 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effect of 3 wt.% Bi in Sn-Zn solder on the interfacial reaction with the Au/Ni metallization in Microelectronic Packaging

Md. Obaidul Haque and Ahmed Sharif

Department of Materials and Metallurgical Engineering (MME), Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh.
E-mail: " Dr. Ahmed Sharif " asharif@mme.buet.ac.bd

Abstract - The work presented in this paper focuses on the role of 3 wt.% Bi in the base Sn–9%Zn solder on the shear strengths and the interfacial reactions with Au/Ni/Cu pad metallization in ball grid array (BGA) applications. Sn-Zn based Pb-free solder alloys were kept in molten condition (240°C) on the Au/electrolytic Ni/Cu bond pads for different time periods ranging from 1 minute to 60 minutes. Cross-sectional studies of the interfaces were also conducted to correlate with the fracture surfaces. Sn-Zn-Bi solders showed better results in terms of shear strength on liquid state annealing than Sn-Zn solders. Two failure modes, ball cut and interfacial intermetallics/pad separation are assessed for the different solders and reflow times. The consumption of Ni in the Sn-Zn solder was larger than that in the Bi-containing solder. By the addition of 3% Bi in the eutectic Sn-Zn solder, the formation of Ni-Zn compound is reduced which in turn increase the reliability of the solder joint to the higher extent.

I. Introduction

Due to recent legislative, environmental, public sentiment and market drivers around the world, traditional Pb-containing (SnPb) solders in electronic industry are under strict restriction. Extensive investigations are on-going over the last few years to find an acceptable Pb-free solder for various electronic attachment applications [1-3]. The common alternatives to the standard eutectic tin-lead solder investigated so far are based on tin alloys in combination with copper, silver, antimony, bismuth or zinc. A key issue affecting the integrity and reliability of solder joints for high Sn containing alloys is the fast interfacial reactions between the molten solder and the under bump metallization (UBM). Recently, Sn–Zn solder has become highly recommended as a substitute for Sn–Pb eutectic solder due to its lower melting point [4]. Sn–Zn solder can also be used without replacing the existing manufacturing lines or electronic components [5]. Moreover, Sn–Zn is advantageous from an economic point-of-view because Zn is a low cost metal. However, Sn–Zn eutectic solder is difficult to handle practically due to its highly active characteristics [5]. The basic microstructure of Sn-9Zn binary alloys and their interfaces with Cu and Ni were investigated [5,6]. Date et al. developed a ductile-to-brittle transition mode of Sn-9Zn joints on Cu metallization by impact test after aging, in which Cu_5Zn_8 thickening and void formation accounted for the brittleness of solder/Cu test coupon with increasing aging time [7]. It was found that the addition of Bi in Sn-Zn near eutectic solder improved the soldering properties [8-10]. Shohji et al. also confirmed that the Sn-8Zn-3Bi showed inferior ductility to Sn-37Pb but double tensile strength of it at room temperature [11]. However, few studies have concentrated on the effect of long time reflow [12], despite the fact that, BGA solder joints may undergo several cycles of reflow during component manufacturing, soldering and rework processes. Hence, this study aims to investigate the effect of extended reflow on the reliability of Sn-9Zn and Sn-8Zn-3Bi (wt.%) solder joints on Au/Ni metallization.

II. Experimental Procedures

A solder mask defined copper bond pad on the flexible substrate of a BGA package was used as a base for electrodeposition of Ni and Au. The solder mask-opening diameter was 0.6 mm with 7μm thick Ni at the ball pad. The average thickness of Au layer was 0.5 μm. The compositions of the solder alloys were Sn-9Zn and Sn-8Zn-3Bi (wt.%). Differential scanning calorimetry analysis showed the solidus and liquidus temperatures of the Sn-8Zn-3Bi solder at 187^0C and 197^0C, respectively. The melting temperature of the eutectic Sn-9%Zn solder was around 199^0C. Lead-free solder balls with a diameter of 0.76 mm, were placed on the prefluxed Au/Ni/Cu bond pad of the substrates and reflowed at a temperature of 240°C for 1 minute in a convectional oven. The flux used in this work was a commercial rosin activated (RA) flux. The assembled packages then reflowed isothermally at 240^0C for times (t) of 5, 10, 30, and 60 minutes in a high temperature oven to examine the interfacial reactions between the solder and the thin-film UBM.

Shear tests were performed on both the as-reflowed (1 minute of reflow) and extended-reflowed samples using a Dage Series 4000 Bond Tester. The shear tool height and the test speed of the shear test in this work were about 40µm and 300µm/s, respectively. About 40 randomly chosen solder balls were sheared to obtain the average and the extent of deviation. The fracture surfaces after the ball shear tests were investigated thoroughly by SEM in the secondary electron mode as well as by EDX.

To investigate the microstructures, the as-reflowed and extended-reflowed samples were mounted in resin, cured at room temperature, mechanically ground and then polished in order to obtain the cross sections of the solder/UBM interfaces. The chemical and microstructural analyses of the gold-coated cross-sectioned samples were obtained using a Philips XL scanning electron microscope (SEM) equipped with an energy dispersive X-ray spectrometer (EDX). The accuracy of the compositional measurement was about ±5%. To find out the formula composition of the IMCs, the chemical analyses of the EDX spectra were corrected by standard ZAF software. The back-scattered electron (BSE) imaging mode of the SEM was used for the interfacial study.

III. Results and Discussion

The interfacial reactions between the Pb-free solders and electrolytically deposited Ni/Cu bond pads were conducted at 240 °C. The mechanical strength of the interface was measured for each reflow condition. It is noted in Fig. 1 that the average shear strength of the Bi-containing solder joints is higher than that of the Sn-Zn solder joints. Initial average shear load of the solder joints was around 1.95 kgf and 1.9 kgf for Sn-Zn-3Bi and Sn-Zn solders, respectively. Fig. 1 also shows that the solder ball shear load for both the solders during extended reflow increases with the increase of reflow time upto 10 minutes and then turns to decrease. It is also interesting to observe nearly the same value of the minimum shear strength for both the joints. The Sn-Zn-3Bi solders gave relatively better ball shear load at about 1.78 to 2.08 kgf over the whole duration of reflow.

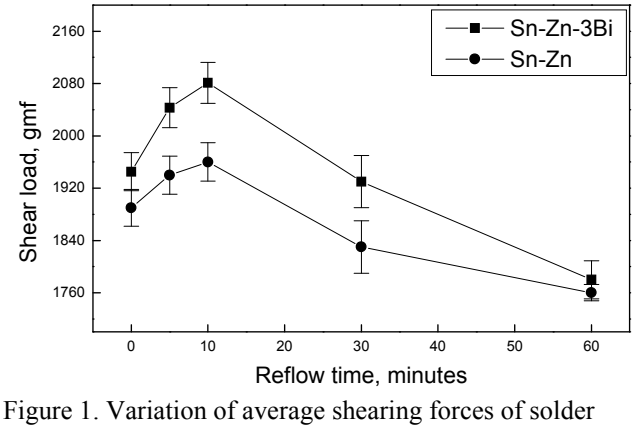


Figure 1. Variation of average shearing forces of solder joints with respect to the different times of reflow.

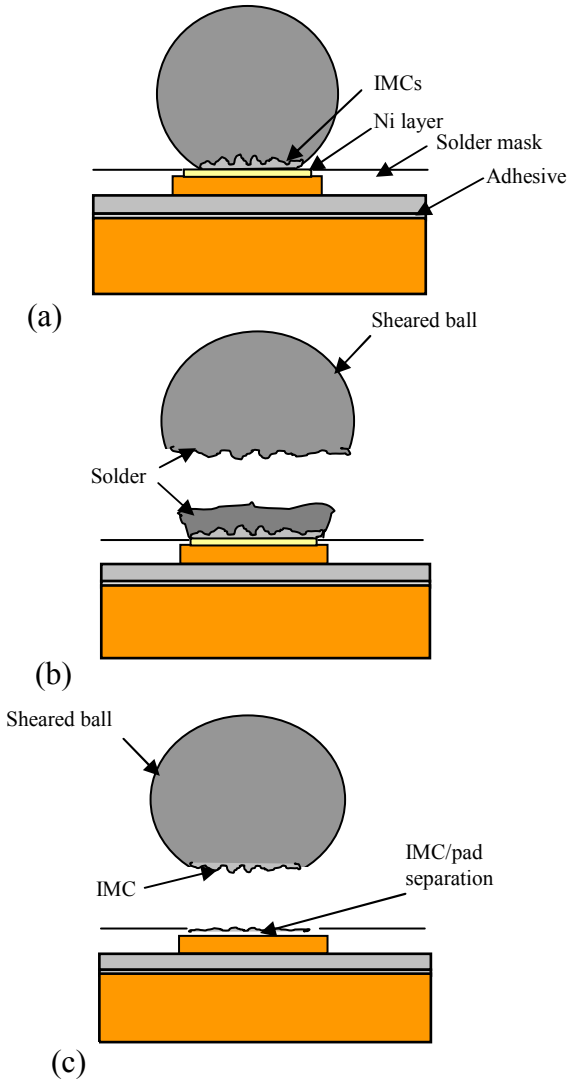


Fig. 2. Schematic drawing of shearing fracture modes of BGA solder joints: (a) shearing setup; (b) mode I; (c) mode II.

After measuring the shear loads, fractured surfaces of the residue pads and sheared balls were immediately studied by SEM. The results reported here (from fracture surface and cross-sectional studies) are based on the highest number of similar occurrences from each readout point of shear strength data. There were two main failure modes, I and II, observed after the shearing tests, as schematically shown in Fig 2. Mode I was identified to be complete cut through the bulk solder balls as illustrated in Fig. 2(b). This fracture occurred at a location near but lower than the shearing height (40µm), leaving a thick layer of solder on the pad. This indicates that the solder/pad bond is much stronger than the shear strength of the bulk solders. It was confirmed that for up to 10 minutes of reflow, the dominant shearing behavior for both the Sn-Zn and Sn-Zn-3Bi solder joints were mode I. The fracture surfaces of the two solder joints were more or less similar, showing a large ductile

deformation of the solder. Fig. 3 shows the fracture surface of the interface formed between the molten Sn–Zn-3Bi solder alloy and the electrolytic Ni/Cu bond pad at 240 °C for 10 minutes. EDX analysis from these surfaces proves that fracture occurs through the solder alloys. Similar trend were also noticed for the Sn-Zn solders. It is consistent with the high value of their shear strengths. It is suggested that extended reflow enlarge the contact area and lower the height of the solder joints. As a result, the forces become higher because of the larger sheared area. A schematic diagram of the extent of deformation of the bulk solder is shown in Fig.4. Due to the solder mask, the contact area between the solder and the bond pad remains the same. However, the area along the shearing height is increased after long time molten condition (Fig. 5b).

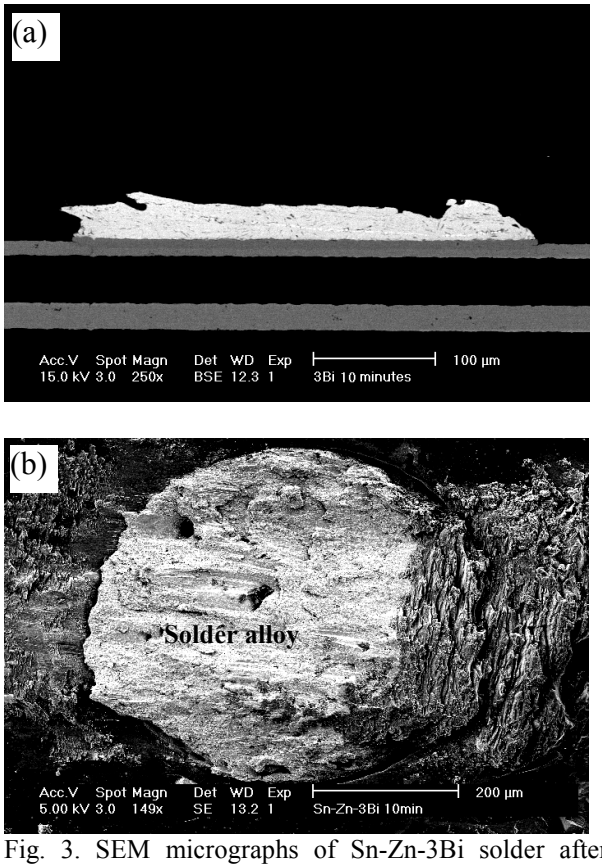


Fig. 3. SEM micrographs of Sn-Zn-3Bi solder after 10 minutes of reflow (a) cross-section of a fractured joint and (b) top view of the pad side fracture surface.

Mode II is entirely different, in that the fracture went through the IMC/pad interface, as depicted in Fig. 2(c). For long time molten reaction, mostly brittle fracture occurred within the IMCs for all the solder alloys. The brittleness of the solder joints increased with the increase of reflow time. Fig. 5 shows typical fracture surfaces of the interface of the Sn-Zn alloy with the electrolytic Ni bond pad for 60 minutes of reaction time. Fig. 5b shows that the fracture surface is almost flat. Failure mode was very brittle. EDX analysis revealed that the brighter layer on the pad side consisted of Ni-Zn IMCs—which might be Ni_5Zn_{21}. EDX results on the black region of the pad showed only Ni-which might be the unreacted Ni of the substrate. However, at the ball side, Ni and Zn were detected under EDX, which might be Ni_5Zn_{21}.

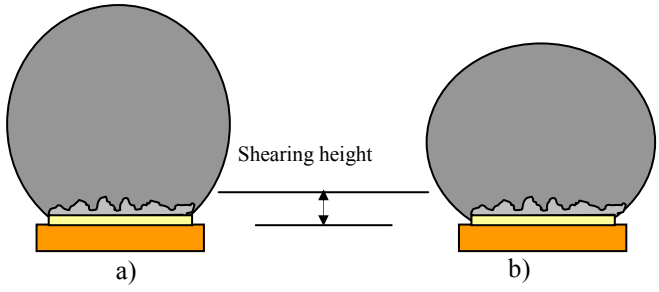


Fig. 4 Schematic drawings of BGA solder joints of (a) as-reflowed; and (b) after 60 minutes of reflow.

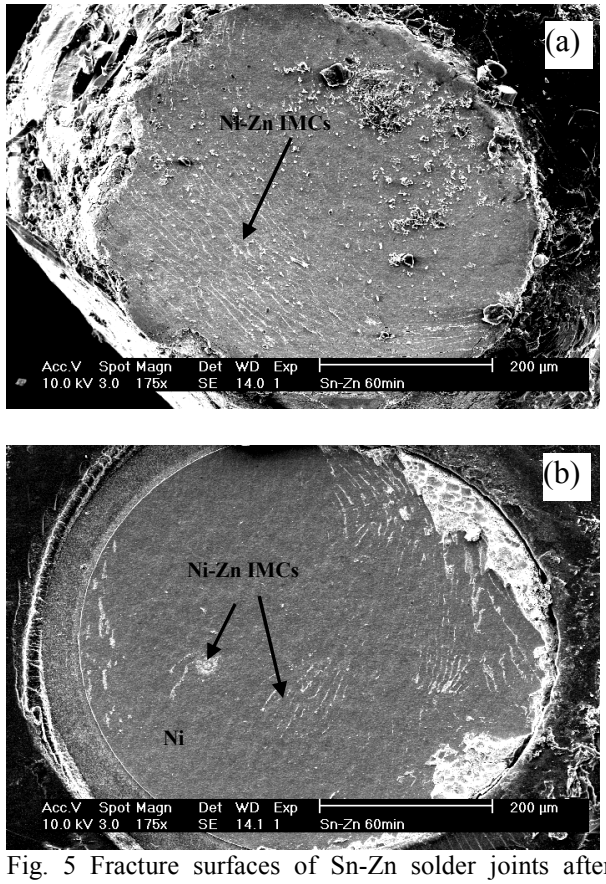


Fig. 5 Fracture surfaces of Sn-Zn solder joints after 60 minutes of reflow (a) ball side and (b) pad side.

Detailed cross-sectional studies were carried out to investigate the relationship between the shear strength and the interfacial morphologies of the solders with the Au/electrolytic Ni/Cu pads. These are cross-sectional micrographs with the section plane perpendicular to the interfaces. All interfaces, more or less, reveal similar features—the solidified solder, reaction zone, original electrolytic Ni layer, and Cu pad. Interestingly, Au did not dissolve into the solders. Instead, Au formed an

intermetallic compound at the interface. EDX analysis of the reaction zone revealed that the IMC is composed of Au and Zn. EDX results shows that the Au percentage of this layer is about 25 at. %. This observation implies that the IMC layer is the $AuZn_3$ compound. In the electrolytic Ni/Sn-Zn-Bi solder joint, the average thickness of intermetallics was 2.3µm. As the initial IMC contained no Ni, the consumption of the original Ni layer in all the Sn-Zn(-Bi) solder joints was negligible. Most interestingly, layer-type spalling was observed in all the solder systems from the initial reflow. At the interface, a very thin layer of IMC was noticed. Fig. 6 shows the microstructures of the solder interfaces after 5 minutes of reflow. Even after 5 minutes of reflow, the interfacial IMC was still very thin.

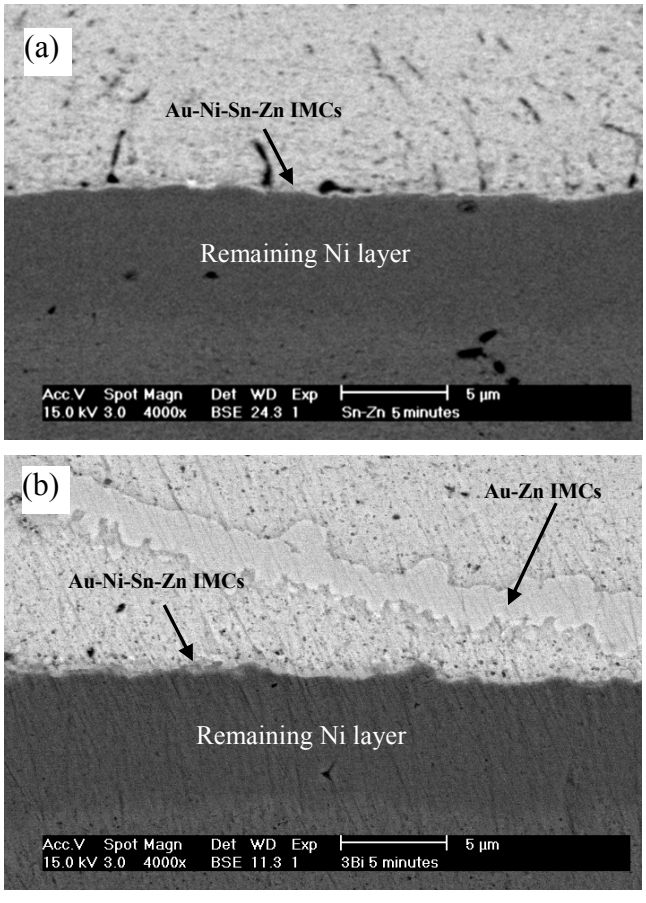


Fig. 6 SEM micrographs showing the interfaces of the solder joints after 5 minutes of reflow of (a) Sn-Zn and (b) Sn-Zn-3Bi.

After 10 minutes reaction in molten condition, it was seen that IMCs thicknesses of all the samples grew gradually with the increasing of time. According to EDX, the IMC at the interface after 10 minutes reflow consisted of Au, Ni, Sn and Zn. The average composition of the interfacial IMCs layer near the substrate side was determined to be 52-59 Zn, 15-20 Au, 4-8 Ni and 6-17 Sn (at.%). For Sn-9Zn solder, the amount of Sn in the interfacial IMC was in the higher side of the range with less amount of Au. Over the original Ni layer, a thin new type of IMC was observed. EDX analysis revealed that the new layer was composed of 78-80 Zn and 22-20 Ni (at.%). This observation confirms the presence of Ni_5Zn_{21} compound. Another interesting thing to note that the spalled Au-Zn IMC layer was situated near the interface for the Bi-containing solders. For Sn-Zn solder, the spalled IMC layer was found little deep in the bulk solder (i.e. away from the interface) even after the initial reflow (Fig. 7a).

After 30 minutes of reaction in molten condition, it was seen that IMCs thickness increased whereas original Ni layer thickness decreased for both the solder systems. It was noticed that the Ni-Zn IMC formed a layer over the original Ni layer and this IMC layer was larger in the Sn-Zn solder system than that in the Bi-containing solder system. However, the interfacial Au-Ni-Sn-Zn IMC layer was larger in the Bi-containing solders than that in the Sn-Zn solder. After 30 minutes of extended reflow, the spalled Au-Zn IMC layer were found as a row of discrete islands in the Bi-containing solder system (Fig. 7b).

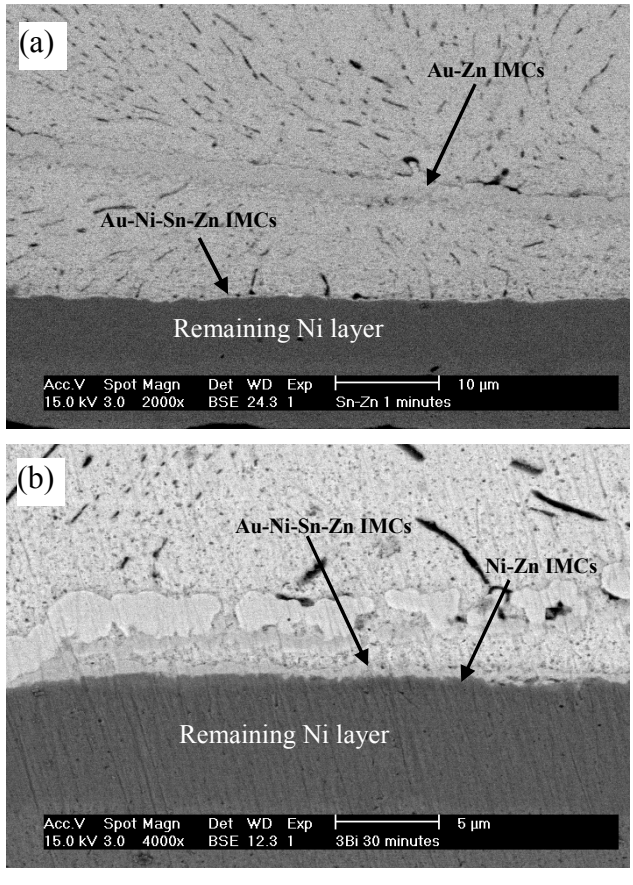


Fig. 7 Backscattered electron micrographs illustrating the interfaces of a) Sn-9Zn solders after 1 minute (magnification X2000) and b) Sn-8Zn-3Bi solders after 30 minutes (magnification X4000) of reflow at 240^0C.

The bright interfacial Au-Ni-Sn-Zn IMC thickness was about 0.7–0.9 µm and 0.9–1.2 µm in Sn-Zn and Sn-Zn-Bi solders respectively after 60 minutes of molten reaction

(Fig.8). The Ni-Zn IMC layer thickness was about 1.4–1.9 μm and 1.1–1.3 μm in Sn-Zn and Sn-Zn-Bi solders respectively. By measuring the remaining Ni thickness from SEM micrographs and by subtracting this from the initial thickness, the consumed Ni thickness was deduced. Fig. 9 shows a comparison between the two different solder alloys of the thickness reduction due to consumption of the Ni from the substrate at 240^0C. Only about 1.5-2.0 μm of the Ni layer takes part in the reaction with both the Zn-containing molten solders after 60minutes. The more participation of the Ni later on with the interfacial IMC is the source of the consumption of the electrolytic Ni layer.

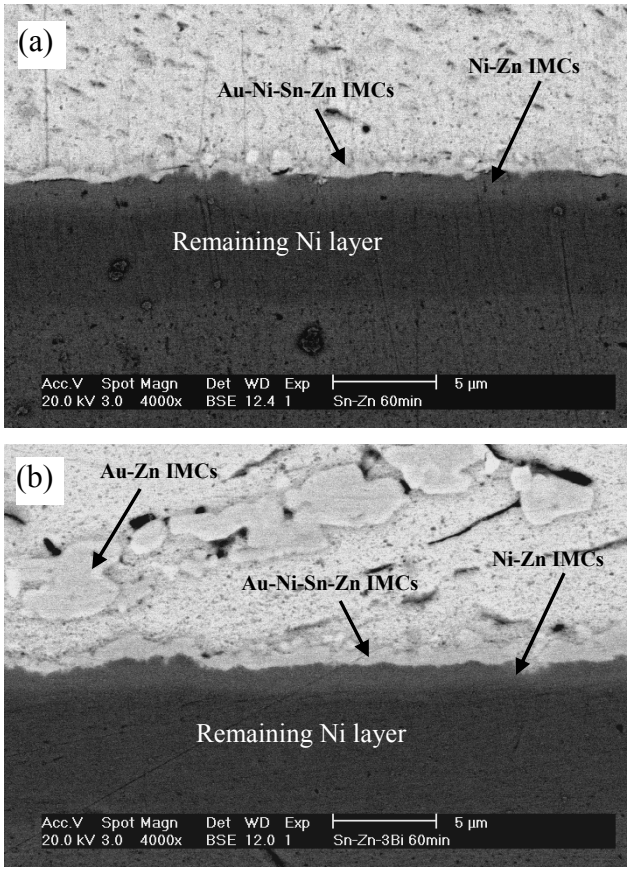


Fig. 8 Backscattered electron micrographs illustrating the interfaces after 60 minutes at 240^0C of a) Sn-9Zn and b) Sn-8Zn-3Bi solders.

From Fig.7 it was also evident that the consumption of Ni in the Sn-Zn solder was larger than that in the Bi-containing solders. For Bi-containing solder, supply of Au for the formation of the interfacial IMC was easier, as the spalled Au-Zn was situated closer to the interface. Thus the interfacial Au-Ni-Sn-Zn IMC was bigger in the Sn-Zn-Bi solder than that in the Sn-Zn solder. The thick interfacial IMCs might hinder the supply the Zn atoms to form the Ni-Zn IMC which in turn slower the consumption of the Ni layer. In the initial stage of reflow, since the thickness of the IMCs for all the solder systems was small, it did not influence the dissolution of the Ni too much. In the initial IMCs, around 4-6at.% Ni was present, whereas in the formed later Ni-Zn IMCs, the atomic percentage of Ni was around 22-20. Thus, the consumption of Ni from the substrate to produce a 1μm thick binary Ni-Zn IMC layer is much higher than that needed to produce the same thickness of Au-Ni-Sn-Zn quaternary compounds. As the Ni-Zn layer thickness was bigger in the Sn-Zn solder, the consumption of Ni from the bond pad was also larger.

In flip-chip packaging, several re-flows, often up to seven or eight times are needed. Each of them brings the solder alloys above the melting temperature for a period of about 30 s to 1 minute [14]. The collective effect of such multiple reflows on the reaction between molten solder and UBM has been a reliability issue. Although the IMC spalling started very early in the reflow, it did not influence greatly the shear strength of the solder joints. The highly reactive nature of the Zn confirms an instant IMC formation at the interface with the spalling of the Au-Zn compound layer. The interfacial IMC together with the unreacted Ni provides the adhesion between the solder and the substrate. With the increase of reaction time, the thick Ni-Zn compound layer creates the weakest link with original Ni layer. By the addition of 3% Bi in the eutectic Sn-Zn solder, the formation of Ni-Zn compound is reduced which in turn increase the reliability of the solder joint to the higher extent.

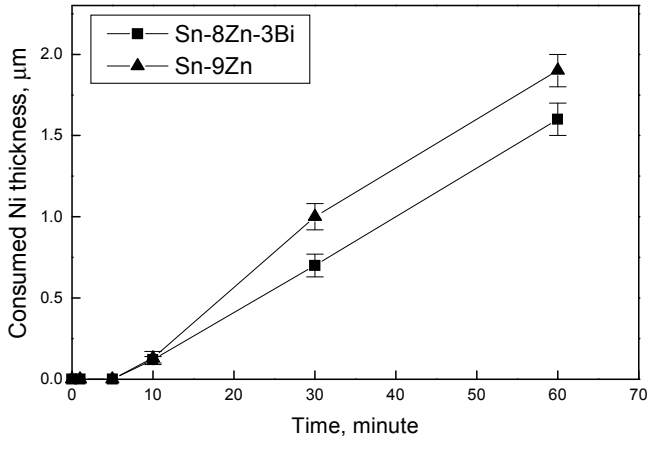


Fig. 9 The consumed thickness of Ni vs. reflow time at 240^oC for Sn-9Zn and Sn-8Zn-3Bi solder systems.

IV. Conclusions

The effect of extended reflow on the shearing behavior of three types of BGA solder joints on Au/Ni metallization was investigated. Two failure modes were identified for the different solders and time of reflow. Up to 10 minutes of reflow, the main fracture mode for all the Sn-Zn(-Bi) solder joints were cut through of the bulk solder balls. With long time of reflow, the solder joints failed at the interface of the IMC/original Ni layer. It was confirmed that during the initial reflow, the IMC layer started to spall-off from the

interface. However, this had little effect on the shear load of the solder joints. It was noticed that the formation of Ni-Zn intermetallics at the interface was decreased just by adding 3% Bi in Sn-9%Zn solder alloy with less overall interfacial reaction at the solder joint.

References

1. K. Suganuma, "Preface-special issue on lead-free soldering in electronics" Materials Transactions. vol. 45, pp. 605-605, 2004.
2. K. Zeng and K. N. Tu, "Six cases of reliability study of Pb-free solder joints in electronic packaging technology" Material Science and Engineering R. vol. 38, no. 2, pp.55-105, 2002.
3. M. Abtew and G. Selvaduray, "Lead-free solders in microelectronics", Material Science and Engineering R, vol. 27, no 5-6, pp. 95-145, 2000**.**
4. J. M. Song and Z. M. Wu, "Variable eutectic temperature caused by inhomogeneous solute distribution in Sn-Zn system", Scripta Materialia, vol. 54, no. 8, pp. 1479-1483, 2006.
5. K. S. Kim, J. M. Yang, C. H. Yu, I. O. Jung and H. H. Kim, "analysis on interfacial reactions between Sn–Zn solders and the Au/Ni electrolytic-plated Cu pad", Journal of Alloys and Compounds. vol. 379, no. 1-2, pp. 314-318, 2004.
6. L. L. Duan, D. Q. Yu, S. Q. Han, H. T. Ma and L. Wang, "Microstructural evolution of Sn-9Zn-3Bi solder/Cu joint during long-term aging at 170°C", Journal of Alloys and Compounds. vol. 381, no. 1-2, pp. 202-207, 2004.
7. M. Date, T. Shoji, M. Fujiyoshi, K. Sato and K. N. Tu, "Ductile-to-brittle transition in Sn-Zn solder joints measured by impact test", Scripta Materialia. vol. 51, pp. 641-645, 2004.
8. P. C. Shih and K. L. Lin, "Interfacial bonding behavior with introduction of Sn–Zn–Bi paste to Sn–Ag–Cu ball grid array package during multiple reflows", Journal of Materials Research. vol. 20, no. 1, pp. 219-229, 2005.
9. Y. S. Kim, K. S. Kim, C. W. Hwang and K. Suganuma, "Effect of composition and cooling rate on microstructure and tensile properties of Sn–Zn–Bi alloys", Journal of Alloys and Compounds. vol. 352**,** no. 1-2, pp. 237-245, 2003.
10. I. Shohji, T. Nakamura, F. Mori and S. Fujiuchi, "Interface reaction and mechanical properties of lead-free Sn–Zn alloy/Cu joints", Materials Transactions. vol. 43, no. 8, pp. 1797-1801, 2002.
11. I. Shohji, C. Gagg and W. J. Plumbridge, "Creep properties of Sn-8mass%Zn-3mass%Bi lead-free alloy". Journal of Electronic Materials. vol. 33, no. 8, pp. 923-927, 2004.
12. Y. Chonan, T. Komiyama, J. Onuki, R. Urao, T. Kimura and T. Nagano, "Influence of P content in electroless plated Ni-P alloy film on interfacial structures and strength between Sn-Zn solder and plated Au/Ni-P alloy film". Materials Transactions. vol. 43, no. 8, pp. 1887-1890, 2002.
13. M. N. Islam, Y. C. Chan, A. Sharif and M. O. Alam, "Comparative study of the dissolution kinetics of electrolytic Ni and electroless Ni–P by the molten Sn3.5Ag0.5Cu solder alloy", Microelectronics Reliability. vol. 43, no. 12, pp. 2031-2037, 2003.
14. C. Y. Liu, C. Chen, A. K. Mal and K. N. Tu, "Direct correlation between mechanical failure and metallurgical reaction in flip chip solder joints", Journal of Applied Physics. vol. 85, no. 7, pp. 3882-3886, 1999.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Design, Simulation and Application of a Novel Compound MOSFET for Emerging CMOS Technology

Shuza Binzaid, John O. Attia

Department of Electrical and Computer Engineering, Prairie View A&M University
Prairie View, Texas 78446, USA
E-mail: salmashuza@yahoo.com

Abstract - An Active-Region-Cutout (ARC) technique was developed and applied to an enclosed poly MOS device to overcome shorted source-drain issue. This novel transistor is named as Active-Region-Cutout-Transistor (ARCT). This transistor is simulated and found to be very tolerant to inner-device interferences in harsh environments such as radiation. ARC technique has an advantage of making MOSFETs with more than three terminals and thus forming a compound MOSFET. The gate poly extension was made through the unipotential electrode of the active region of drain to further reduce effects of interferences even further. Two types of CMOS circuits have been studied by replacing transistors with the compound ARCTs.

I. Introduction

Tolerance to interferences of sub-micron and nano-meter electronic circuits and devices is the primary concern for operating in harsh environments like space radiation. The circuits and devices affect in normal operation, parameters and also reliability of circuit becomes degraded considerably. Many approaches have been taken over the years to make devices hardened-by-design (HBD). Many HBD transistors are being used today such as ELT (Enclosed Layout Transistor) and H-Gate MOSFET [1, 2]. Both ELT and H-Gate take more space on the silicon than a 2-edge standard MOSFET. Figure 1 shows the layout of these devices for their size comparisons. Devices are becoming more intolerant to interferences as they get smaller in size in advanced silicon processes.

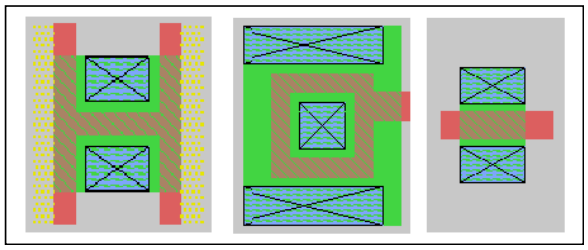


Figure 1 From left NMOS H-Gate, an ELT and a 2-edge standard transistor.

The design approach in this work is applied to an enclosed poly MOS device in figure 2. This device has an enclosed gate and a smaller device aspect ratio, W/L, compared to those HBD devices.

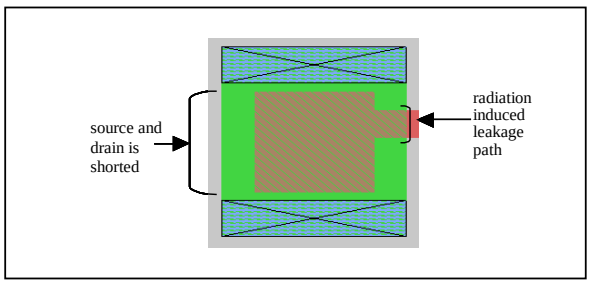


Figure 2 An enclosed poly NMOS device.

An active-region-cutout (ARC) technique was developed to overcome the short between the drain and the source. ARC technique removes a small part of active short area, and thus removing the short in the device. Also the poly extension out of the device was made through two electrodes that have the same electrical potential. Thus the device can have significantly reduced current leakage path. These electrodes were placed on the same active region of either source or the drain of the device. It is called equipotential electrodes [3]. Figure 3 shows the layout of the ARCT.

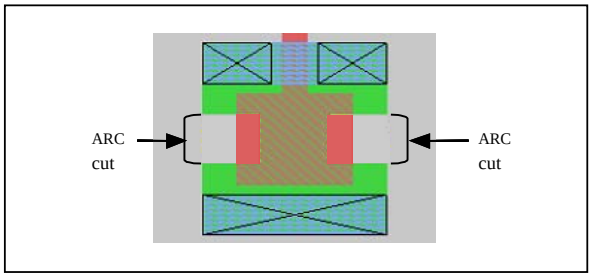


Figure 3 The ARCT showing ARC regions.

The gate is I-shaped. But it has some features similar to H-Gate transistor like the active regions have longer gate

edges. On the other hand, ARCT was found to be much smaller than other HBD transistors. An integrated circuit's (IC) device density can be improved, if built using the ARCT [3].

Simulation results show that it has very low radiation induced-leakage current than 2-edge standard transistor. Also these low values are similar to H-Gate and ELT [4].

II. EFFECTS ON MOSFETS

Current leakage mechanism under the radiation environments is illustrated in figure 4 for a 2-edge standard MOSFET. In the MOSFET, device operating threshold voltage can change due to accumulation of charges in oxide. First, change occurs at the interface corners.

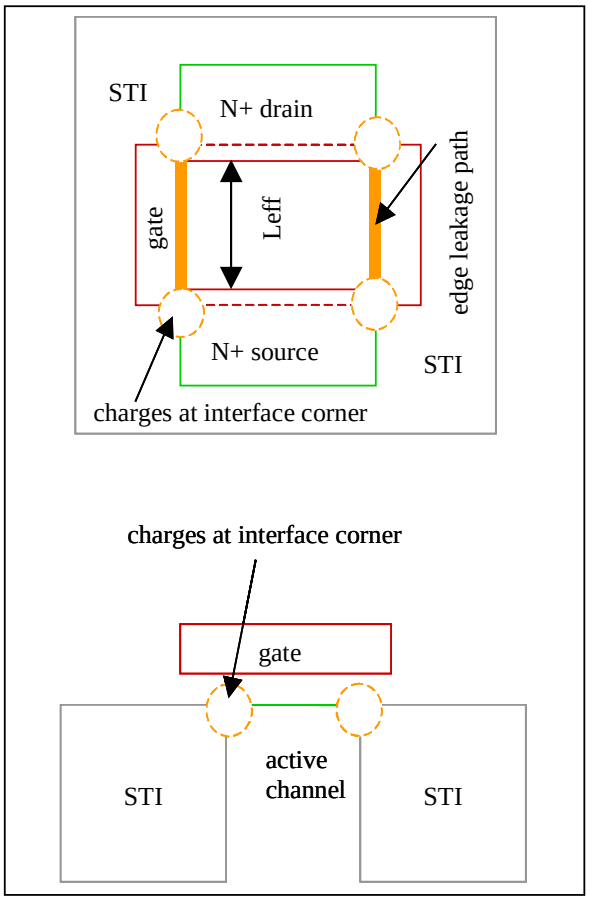


Figure 4 Upper figure shows top-view and lower figure shows the side-view of the 2-edge standard MOSFET.

Channel inversion occurs at these corners due to these accumulated charges. If the latter are large enough, then accumulated charges affect STI edges along the channel near the interface. Then device creates an off-state leakage path [5]. This is caused by channel edge inversion due to accumulated charges. Continuing this inversion in smaller device affects the entire channel and thus it completes a high current path. This radiation mechanism affects the 2-edge transistor very quickly and turns it on.

Figure 5 shows the interface corners of the H-Gate and ARCT. The H-Gate and ARCT have very different shapes of STI edges. These shapes keep STI corners away from active channel edges. Arrows in figure 6 show them. A complete channel edge inversion does not occur easily. So a direct channel leakage path is not present in these two MOSFETs [3].

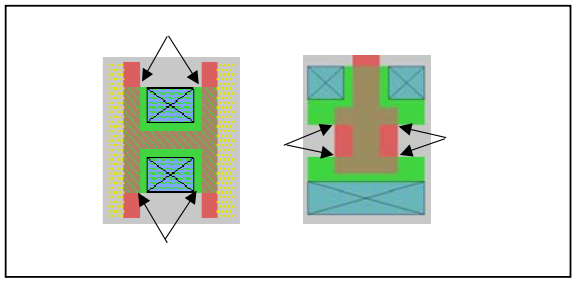


Figure 5 Arrows showing STI interface corners of H-Gate and ARCT.

III. DEVICE SIMULATIONS AND RESULTS COMPARISON

Figure 6 shows the log value of charge density at the interface with respect to the induced leakage current of the ARCT. The leakage current is 2.23E-08 amps at 1E12 charges/cm^{-2} of silicon mid-band gap. The leakage current of a simple MOSFET is shown in the figure 8.

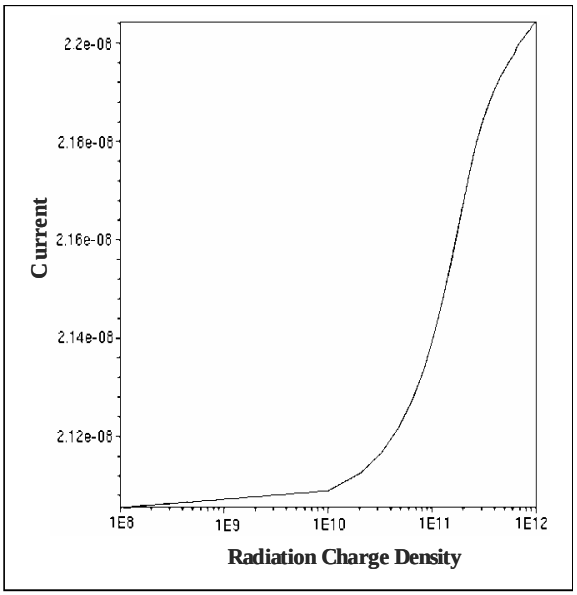


Figure 6 Leakage Current vs. Interface charge density of the ARCT.

All simulations were completed keeping the electrical parameters same for all devices. The plot in figure 7 shows that the 2-edge standard MOSFET has leakage current around 22E-05 amps at 1E12 charges/cm^{-2}. So ARCT has much lower leakage current.

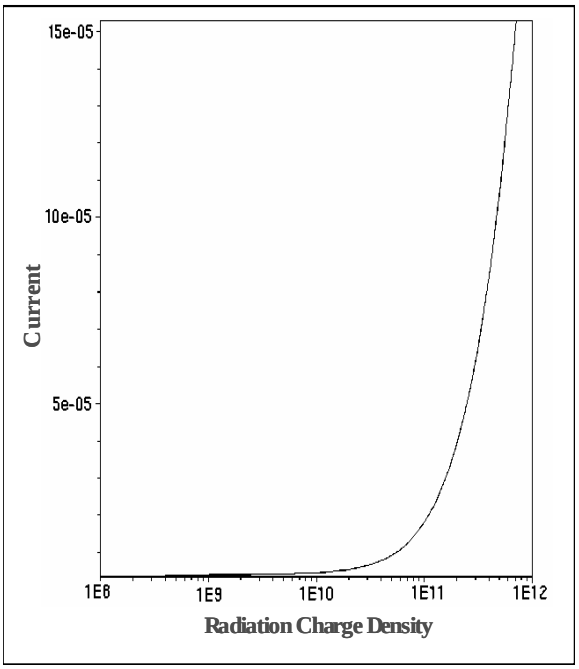


Figure 7 Leakage Current vs. Interface charge density of a 2-egde standard MOSFET.

Simulation results of these MOSFETs are given in Table 1. From these values, it is found that the 2-edge standard MOSFET has leakage current 1.13E4, 9.5E3 and 9.8E3 times higher than those of H-Gate, ELT and ARCT. All these devices were simulated using electrical parameters of 0.5um silicon process technology of AMI Semiconductor Inc. Design rules for this process technology were maintained for all transistors.

Table 1
Induced Leakage Currents at Charge Density for Transistors

Device Name	Leakage Currents @ charge density				
	1.00E+08	1.00E+09	1.00E+10	1.00E+11	1.00E+12
Standard MOSFET	1.00E-05	1.05E-05	1.10E-05	2.20E-05	2.20E-04
ARCT	2.00E-08	2.10E+08	2.11E-08	2.14E-08	2.24E-08
ELT	2.25E-08	2.26E-08	2.27E-08	2.28E-08	2.30E-08
H-Gate	5.16E-07	2.17E-07	2.18E-07	2.41E-07	1.95E-08

Normally, the leakage current increases as the size of the transistor increases. The leakage current of the 2-edge standard MOSFET would even larger if area was same size of ARCT. It is seen from the table 1 that H-Gate, ARCT and ELT have very low leakage currents. Thus it can easily be concluded that ARCT is a HBD transistor, similar to the H-Gate and ELT.

IV. COMPOUND ARCT WITH MULTIPLE ELECTRODES

The ARCT has one more advantage over other HBD transistors. It can have more than three terminals made possible by using ARC techniques to increase the number of source and drain electrodes. This makes ARCT to be more than one MOSFET. Current flow path in the device can determine the position of the transistor. Figure 8 shows that it is possible to replace three MOSFETS (M1, M2 and M3) using one 3-electrode compound ARCT in the sense amplifier [6]. Figure 9 shows that compound ARCT can be used at output stage of a high-speed sample and hold circuit [7]. This is a very good example how a compound ARCT can be used to replace two or more transistors in circuits. Also these ARCTs provide immunity to interferences of these circuits. All MOSFETS in these circuits can be replaced by both normal ARCTs and compound ARCTs.

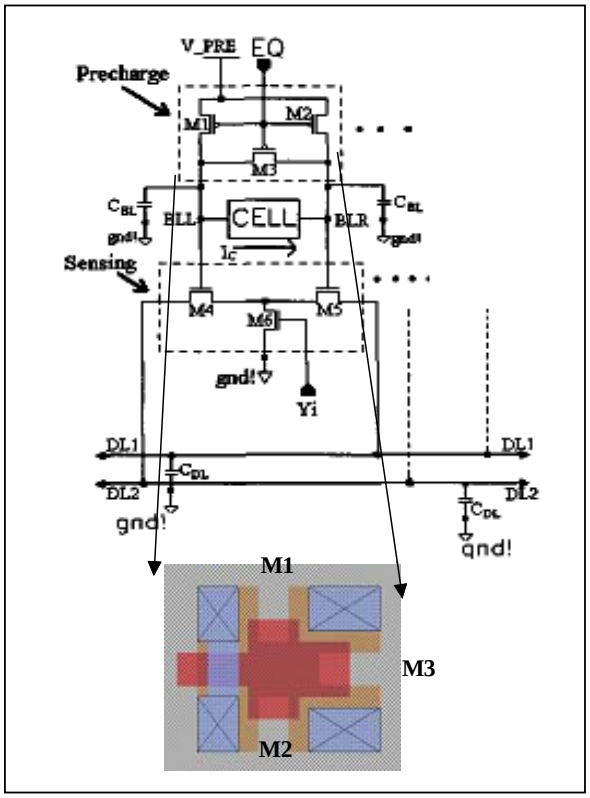


Figure 8 A compound PMOS ARCT is having three-electrodes that can replace three MOSFETS in the sense amplifier's pre-charge circuit (upper dotted box).

So the compound ARCT can easily be configured for replacing two and more MOSFETs in circuits. These circuits can also be compacted with the compound ARCT, because initially they have separately placed transistors in circuits. Therefore using compound ARCTs can increase device density of an IC and thus it can be fabricated in a smaller silicon space. Other HBD devices used in such circuits take larger silicon space because they cannot be

as small as compound ARCT. So area of ICs can be reduced in two steps: using ARCT, as it is smaller in size of ARCT and then using the compound ARCT, whenever applicable in CMOS circuits.

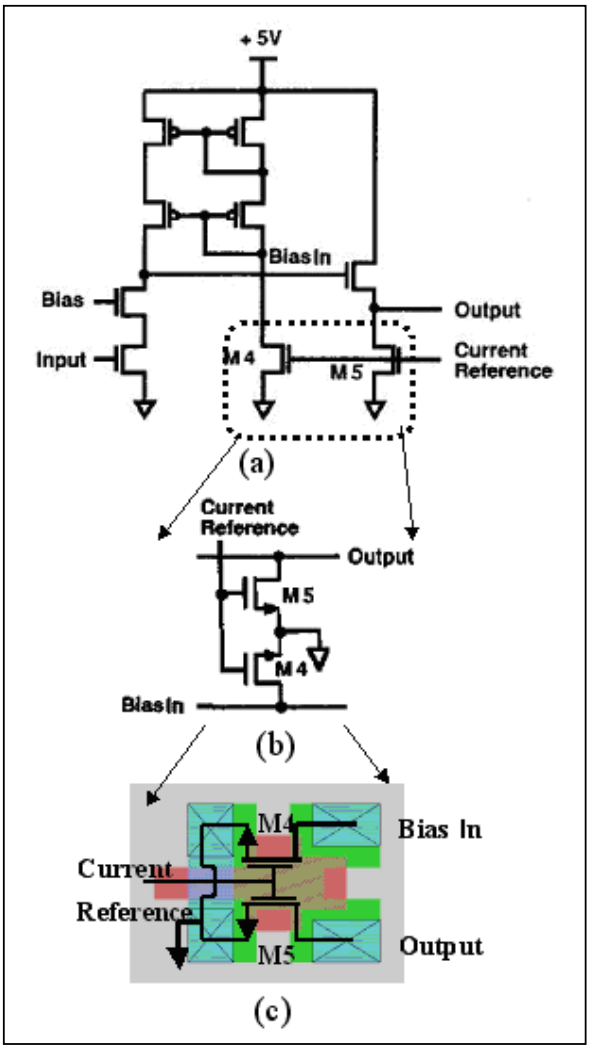


Figure 9 (a) the high-speed sample and hold circuit, (b) the output stage of the circuit and (c) a compound ARCT replaces two transistors in the output stage.

V. CIRCUIT LAYOUT EXTRACTION, SPICE SIMULATIONS AND RESULTS

First a layout needs to be extracted to an intermediate script format before it can be converted to a SPICE circuit description format. A compound device cannot be extracted directly from the layout, because it has equipotential electrodes and more than two active electrodes. Basically, the extraction tool looks for two gate edges that separate two active regions. Thus the tool concludes a device to be a transistor. The ARCT has two poly edges with two active regions. It extracts fine if it does not have equipotential electrodes. But incase of compound ARCT, it finds more than two active regions with poly edges. The extraction tool is setup today such that soon as it finds two poly edges separated by two active electrodes, it concludes as a transistor and ignores the other active electrode. Thus the tool misinterprets and fails to extract any compound MOSFET.

An equivalent layout of compound transistor can be created such that the extraction tools can see it as separate transistors [8]. Figure 10 the equivalent device of the compound ARCT with four terminals. All MOSFETs in this device have a common gate.

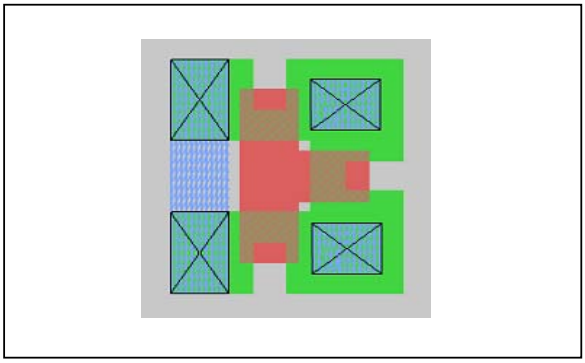


Figure 10 The equivalent device of compound ARCT used for circuit extraction for SPICE in figure 10.

After the layout was extracted, it was converted to a SPICE format. Then appropriate transistor model parameters and also the signals were assigned in the SPICE circuit's definition file. After the simulation was run, the plot with the probe data was produced. The plot of SPICE simulation of the compound ARCT is presented in figure 11. The plot in figure 11 corresponds to V(3) = Output = 5V, V(4) = Bias In = 0.5V and V(5) = Current Reference = 5V in figure 9.

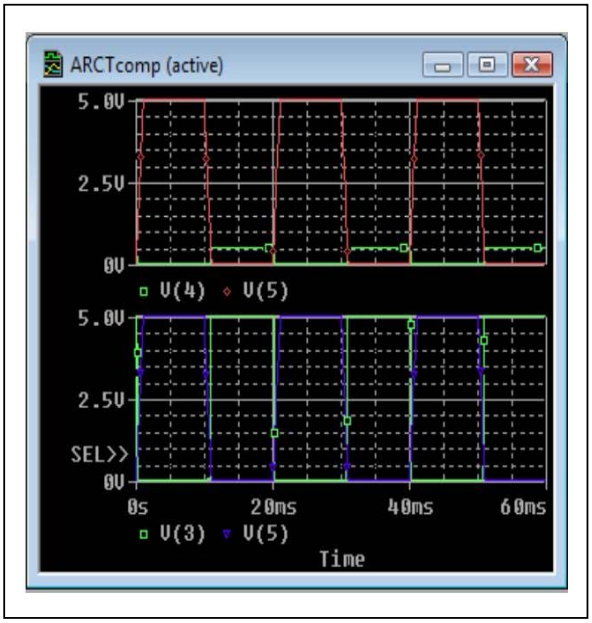


Figure 11 The plot of SPICE simulation using the compound ARCT in the circuit of figure 10.

VI. ADVANCES IN STATE-OF-THE-ART

A novel MOS transistor, called ARCT that can be used as HBD transistor. It can be implemented as a compound ARCT and it can be used for circuits to replace more than two transistors. This compound transistor can replace a number of MOSFETs where they share a common gate signal and a common electrode. Examples of such circuits were presented in the section IV.

ARCT does not have any process dependency and thus it does not require any process changes. Also it can use any available common single poly process of silicon fabrication. Any commercial foundries can fabricate IC designed using ARCT.

ARCT is a smaller HBD device. Device density is improved in circuits. W/L of ARCT is much lower than ELT and H-Gate. So overall cost of circuits designed with ARCT and compound ARCT in fabricated IC is improved.

VII. CONCLUSIONS

ARCT is compared with other HBD devices such as the ELT and the H-Gate. Simulation results show that ARCT behaves similar to those HBD devices. Also ARCT has much less induced-leakage current than a 2-edge standard MOSFET. ARCT has another advantage over other HBD MOSFETs in that it can be configured as more than one transistor making it a compound ARCT. Compound ARCT has equipotential electrodes to further reduce leakage currents. A layout extraction technique of making MOS devices that are equivalent to the compound ARCT was presented. SPICE simulation confirms that the compound ARCT works normally in circuits. The compound ARCT can be used in the design of sense amplifier and comparator circuits. The resulting circuit occupies less silicon area and it is tolerant in harsh environment like space.

VIII. ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (NSF), under award number 0531507. Any opinions, findings, conclusions and recommendations expressed by the authors do not necessarily reflect those of the NSF.

Authors also like to thank Dr. Ron. D. Schrimpf at Vanderbilt University, Tennessee, USA, for his advice about effects on HBD MOSFETs in this work.

IX. REFERENCES

[1] S. S. Chen, S. H. Lu, T. H. Tang, "A Comparison of Floating-Body Potential in H-Gate Ultrathin Gate Oxide Partially Depleted SOI PMOS and NMOS Devices Based on 90-nm SOI CMOS Process" Electron Device Letters, IEEE, Volume 25, Issue 4, Apr. 2004, Page(s): 214 – 216.

[2] W.J. Snoeys, T.A.P. Gutierrez, G.A. Anelli, "New NMOS Layout Structure for Radiation Tolerance**",** IEEE Transactions on Nuclear Science, Volume 49, Issue 4, Part 1, Aug. 2002, Page(s): 1829 – 1833.

[3] S. Binzaid, J. O. Attia, and R. D. Schrimpf, "An Active-Region-Cutout-Transistor (ARCT) Apparatus for Minimizing Leakage Current in Radiation Environments", US Provisional Patent, US 61/004,429, Accepted and Recorded on 28 Nov. 2007.

[4] S. Binzaid, J. O. Attia and R. D. Schrimpf, "Biased Active Region Cutout Electrode Transistor (BARCET) Apparatus for Ultra-Low Leakage Currents in Radiation Environments", US Provisional Patent, US 61/062,116, Accepted and Recorded on 24 Jan. 2008.

[5] G. Niu, J. Suraj, G. Banerjee, J. D. Cressler, S. D. Clark, M. J. Palmer, S. Subbanna, "Total Dose Effects on the Shallow-Trench Isolation Leakage Current Characteristics in a 0.35um SiGe BiCMOS Technology", IEEE Transactions on Nuclear Science, Volume 46, Issue 6, Dec. 1999, Page(s): 1841 – 1847.

[6] Y. Tsiatouhas, A. Chrisanthopoulos, G. Kamoulakos, T Haniotakis, "New memory sense amplifier designs in CMOS technology", The 7th IEEE International Conference on Electronics Circuits and Systems, 2000. Volume 1, 17-20 Dec. 2000 Page(s): 19 - 22.

[7] Peter J. Lim, *and* Bruce *A.* Wooley, "A High-speed Sample-and-Hold Technique Using a Miller Hold Capacitance", IEEE Journal of Solid-State Circuits, Volume 26. No. 4, Apr. 1991, Page(s) 643-651.

[8] S. Binzaid, J. O Attia, "Enclosed Layout Transistor with Active Region Cutout", IEEE Region 5 P-Basics2 Conference, Apr. 2008, Pages(s): 22-26.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Active-Region-Cutout-Enclosed-Layout-Transistor Device Applications in Electronics

Shuza Binzaid, John O. Attia

Department of Electrical and Computer Engineering, Prairie View A&M University
Prairie View, Texas 78446, USA
E-mail: salmashuza@yahoo.com

Abstract -ARC technique was developed and latter a HBD ELT was modified with ARCs. The new MOSFET was called ARCELT, which contains three electrodes and thus it became a compound MOSFET having three transistors within it. Relationship between transistors in ARCELT and ARCs is established. The effective area of each transistor reduced to almost 33% of ELT. Some special application techniques of ARCELT in electronic circuits are studied and implemented. Technique to calculate and control device aspect ratio, W/L is found to approximate the value. Special technique is needed to make equivalent device of the ARCELT for SPICE circuit simulation. ARCELT reduces electronic circuit's size; thus saves silicon space and improves IC's design density.

***Index Terms*—Active-region-cutout (ARC), enclosed-layout-transistor (ELT), Active-region-cutout-enclosed-layout-transistor (ARCELT), hardened-by-design (HBD), integrated circuit (IC).**

I. Introduction

A. Prior Works

Sub-micron and nano-meter electronic circuits are the primary concern for operating within the normal tolerance in harsh environments like space. One of the design approaches for making devices hardened-by-design (HBD) device is Encosed-layout-Transisor (ELT) [1]. ELT takes more space on the silicon than a 2-edge standard MOSFET. Figure 1 shows layout of these two devices and their size comparison.

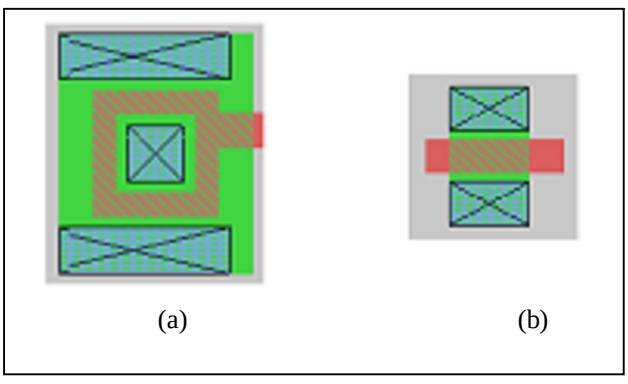


Figure 1 (a) an ELT and (b) a 2-edge standard MOSFET.

As the technologies for fabricating electronic circuits and devices move into the sub-micron levels, devices need to be smaller, more compact and interference-free. New devices and techniques to meet challenges are being sought of newer technologies.

B. Effects of Radiation Interferences

Current leakage mechanism under harsh radiation environments is illustrated in figure 2(a) and 2(b) for a 2-edge standard MOSFET. In this MOSFET, the device operating threshold voltage can change due to accumulation of charges. First, change occurs at the interface corners shown in figure 2(a).

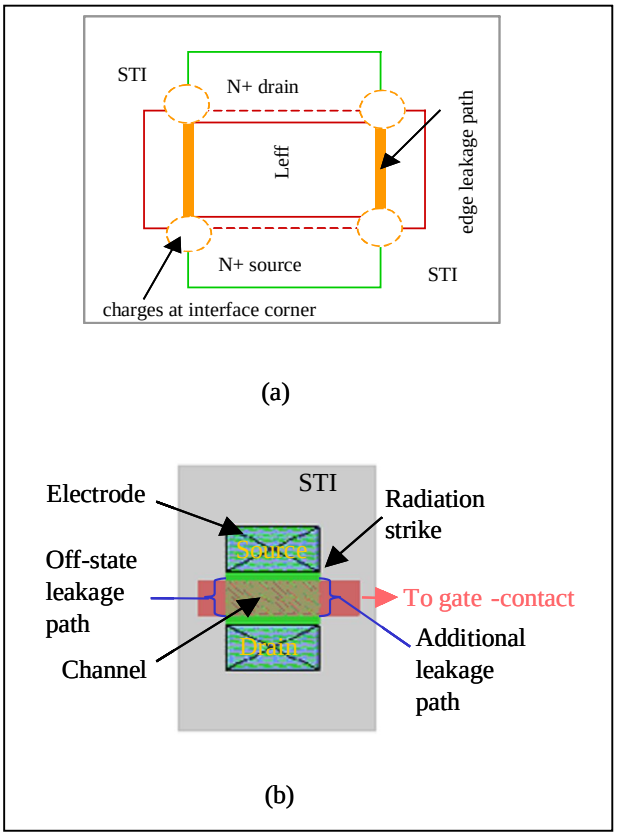


Figure 2 (a) radiation effects mechanism and (b) illustration of off-state leakage path in the 2-edge standard MOSFET.

When radiation strikes near the MOSFET, the channel inversion occurs at these corners due to accumulation of radiation induced-charges. If charges are strong enough, these accumulated charges affect STI edges along the channel near the interface. Then the device channel inversion occurs along the channel edges. Thus device creates an off-state leakage path [2]. The channel has a high current path immediately after and thus transistor turns on. This mechanism affects the 2-edge transistor severely due to generation of larger induced-leakage currents and device degrades reliability.

C. The ARCELT

In order to make the device smaller and flexible aspect ratio W/L, the design concept of ARC was applied to an enclosed poly MOS device. Also the poly extension out of the device was made through two electrodes that have the same electrical potential. Thus the device does not have any current leakage path at poly extension. These electrodes were placed on the same active region of either source or the drain of the device. It is called equipotential electrodes. This device is a HBD transistor [3]. The ARC technique was applied to ELT that is recognized as a HBD MOSFET. The latter device is named as ARCELT [4]. Figure 3 shows the design and layout technique of ARCELT. It has three electrodes and four transistors. Figure 4(a) shows the single transistor and 4(b) shows the four-transistor ARCELT. Also figure 4(c) shows ARCELT with four ARCs where it creates total of eight transistors. Current path in this NMOS device determines the configuration as NPN transistors within.

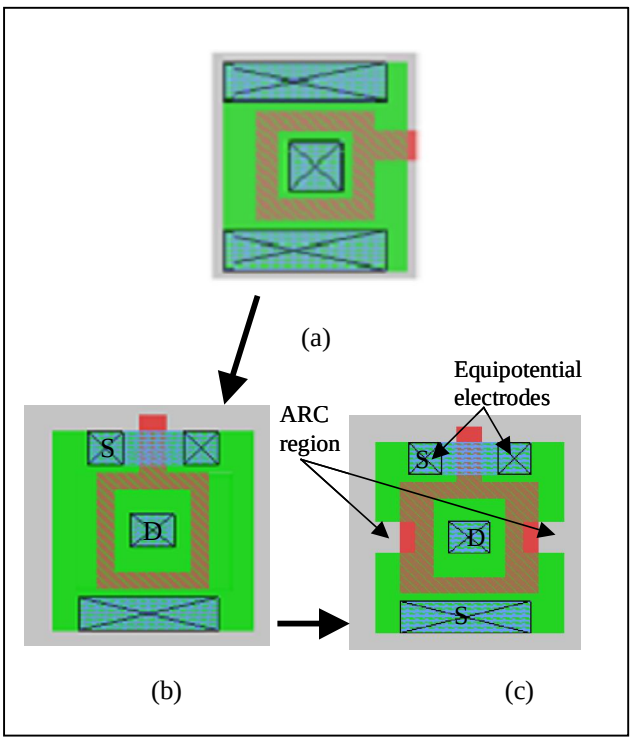


Figure 3 (a) the ELT (b) modified ELT with equpotential electrodes and (c) the ARCELT having two ARCs and three active electrodes (S-D-S).

ARCELT has two ARC regions and a pseudo-enclosed active area that surrounds the gate like an ELT. There is an electrode at the center of the device enclosed by the gate poly. Also there are two electrodes outside of the gate poly separated by two ARCs. By having this enhancement to the ELT, the ARCELT became truly a three-electrode MOSFET. This MOSFET still contains very similar characteristics of immunity to radiation as the ELT [4].

With these three electrodes, ARCELT has enhanced the ELT into a compound MOSFET**.** ARCELT's source (S) and drain (D) can be configured as S-D-S, D-S-D, S-S-D and D-D-S for three-electrode MOSFET. For example, figure 3(c) is configured as S-D-S. Also it can be configured as a normal ELT with only one source and one drain when outer electrodes are connected with metal wire in layout. This feature makes the ARCELT a very flexible HBD transistor for designing electronic circuits.

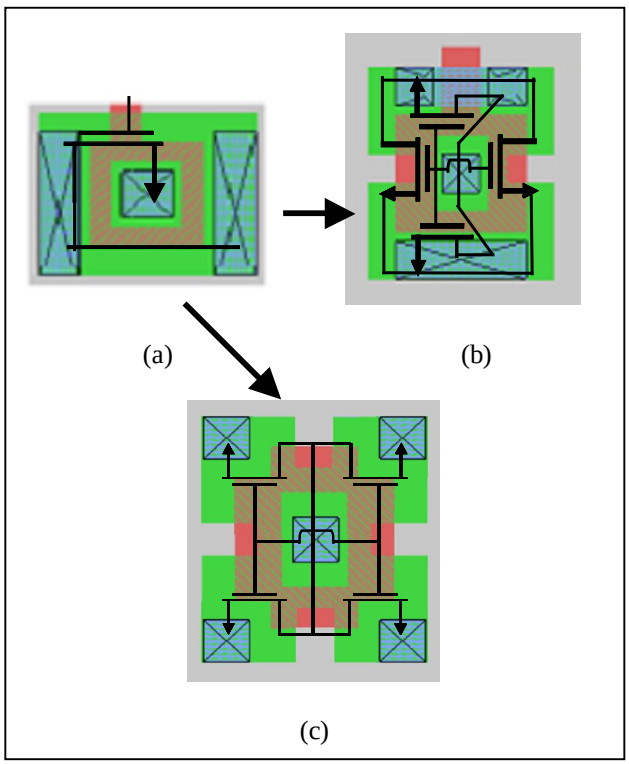


Figure 4 (a) one transistor in the ELT and (b) four transistors in the three-electrode ARCELT as figure 3(c) and (c) eight-transistor ARCELT where transistors created by ARCs not shown here.

The relationships between number of ARC, number of electrodes and transistors in ARCELT are found by:

$$n_{electrodes} = n_{ARC} + 1 \qquad \text{when} \quad n_{ARC} = 2,3,4\ldots$$

and also number of transistors is:

$$n_{transistors} = 2n_{ARC} \qquad \text{when } n_{ARC} = 2,3,4\ldots$$

So a compound ARCELT creates transistors twice the number of applied ARCs.

II. The ARCELT in Electronic Circuits

The ARCELT is a compound MOSFET whose current flow path in the device can determine the position of the transistor. The ARCELT also provide immunity to radiation-induced leakage current in harsh radiation environments. All MOSFETS in circuits, illustrated in this section, can be replaced by compound ARCELTs. So they provide radiation immunity to interferences when circuits are designed. Also circuits using ARCELT are compacted, because the initial designed circuits had all transistors separated. Therefore using compound ARCELTs can increase device density of HBD ICs and thus it can be fabricated in a smaller silicon space. As a result, the yield in commercial production technologies will be increased with the use of ARCELT.

A. Application Circuits

Many circuits can replace a number of individual transistors using ARCEL. Figure 5(a) shows circuit of a sense amplifier [5]. Simple three-electrode compound ARCELT can replace three transistors shown in Figure 5(b). Total of six transistors can be replaced in this circuit with two ARCELTs shown here.

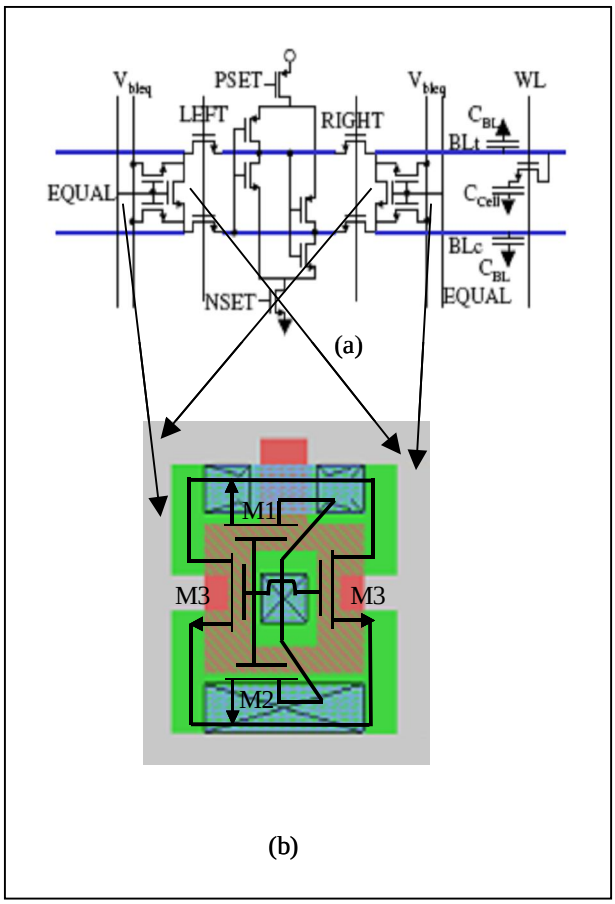


Figure 5 (a) circuit diagram of the sense amplifier and (b) three-electrode ARCELT that replaces transistors M1, M2 and M3 of sense amplifier.

The ARCELT in figure 5(b) has four transistors where two of them are in parallel. They are marked as M3 in the figure 5(b). So functionally it is a three-transistor compound MOSFET. The average area of each transistor became 1/3 of ELT's one transistor, effectively. So the area of a transistor reduced to about 33% in ARCELT when compared to that of the ELT.

Figure 6 shows that it is possible to replace two MOSFETS using a three-electrode compound ARCELT in the circuit. Figure 6(a) is the input stage of a comparator circuit. Both transistors, M4 and M5 connect circuit outputs to ground [6]. Only two transistors in ARCELT play this major role shown in figure 6(b). The other two transistors in ARCELT stay benign in the circuit. It is a very good example of how exceptions to application of ARCELT can be made for its transistors within.

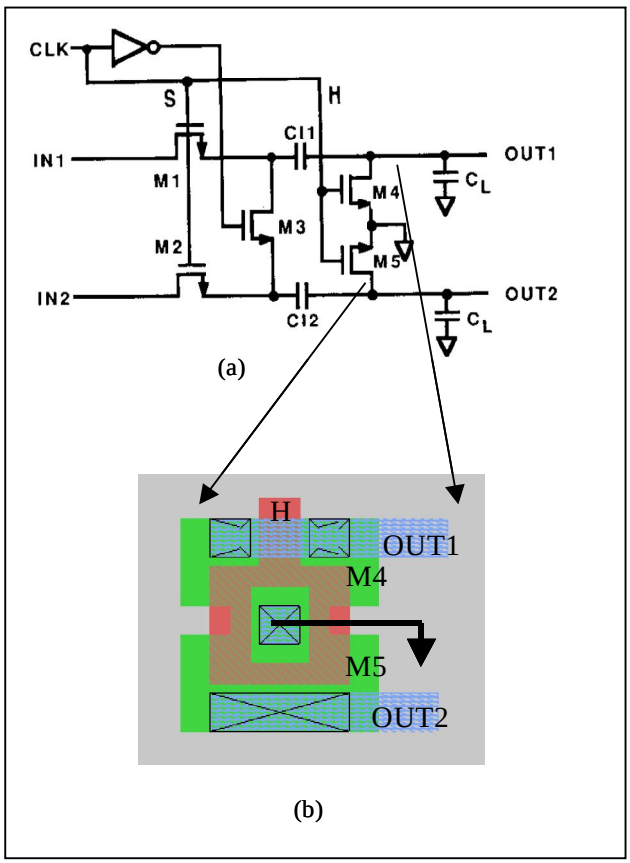


Figure 6 (a) input stage of the comparator circuit and (b) the 3-electrode compound ARCELT. It replaces two transistors M4 and M5.

B. Circuit Layout Extraction, SPICE Simulations and Results

A layout is first extracted to an intermediate script format before it can be converted to a SPICE circuit description format. A compound device cannot be extracted directly from the layout. An equivalent layout of the compound transistor can be created where the layout extraction tool can see it as separate transistors. Figure 7(a) shows layout of the compound ARCELT and figure

7(b) shows its equivalent device. Functionally they both are equivalent as they have similar current conducting paths.

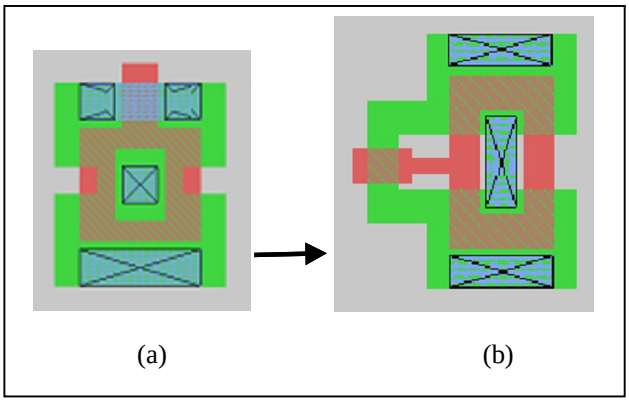


Figure 7 (a) the ELT and (b) its equivalent device.

After the layout is extracted using the equivalent device of the compound ARCELT shown in figure 7(b), appropriate transistor model parameters and also the signals were assigned in the SPICE circuit definition file. After the simulation was run, the plot with the probe data was produced. The plot of SPICE simulation of the compound ARCELT is shown in figure 8. The plot corresponds to V(3) = Out1, V(4) = Out2 and V(5) = H. Also V(1) = IN1 = 5V and V(2) = IN2 = 0.5V, shown in figure the comparator circuit in figure 6(a). This circuit produces equal potential for both outputs when signal from the clock is at logic '1'. The plot of circuit with ARCELT verifies the proper operation of the circuit.

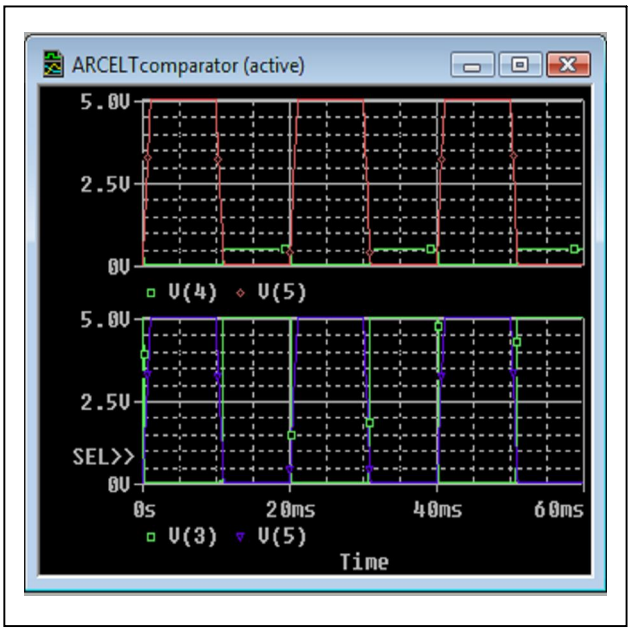


Figure 8 the data plot of SPICE simulation of the compound ARCELT for the comparator circuit shown in the figure 6(a).

Layout extraction tools today in both Cadence and Magic, cannot perform its tasks for special designs where a MOSFET has separated active regions with same potential. So the extraction tool fails to recognize the equipotential electrodes. Again, the tool is set such that electrically separated regions by channel i.e. active regions can be only two for a MOSFET. Basically, the extraction tool looks for two gate edges that separate two active regions, like a simple 2-edge standard transistor. Thus the tool concludes a device to be a transistor. The ELT has two poly edges with inner and outer active regions. The extraction algorithm works well. But in the case of a compound ARCELT, it finds more than two active edges with same poly gate. The current tool is setup such that as soon as it finds two poly edges with two active regions it concludes a transistor, but it ignores the other active regions. Thus the tool misinterprets and fails to extract a compound transistor [5,7]. Thus the ARCELT could not be extracted without making its equivalent device shown in figure 7(b). Also it is clear that there is a need for a new revision of extraction tool which should be able to analyze circuits thoroughly, justify placements of transistors for compound MOSFETs and also various complex and unconventional designs where W/L variation is combined within a compound MOSFET.

III. The ARCELT in IC Design

A. MOSFET Area Comparisons

Area is one of the major concerns for making ICs cost effective. When comparing a single 2-edge standard MOSFET in figure 1(b), with two-electrode ELT in figure 1(a), at first the ET seem too inefficient in silicon space due to its larger sizes. As the ELT was made compound, it becomes very efficient in area. Circuits can be drawn compact for ICs. Also many specific circuits provide increased reliability under radiation when they are designed using the ARCELT. The table 1 shows the areas of ARCTELT, described in this work and it also shows percentage of silicon space compared to 2-edge standard MOSFET.

TABLE I. MOSFETS AREA (λ^2) COMPARISON AND PERCENT INCREASE TO A 2-EDGE STANDARD MOSFET.

No of Device	Areas of MOSFETs (λ^2)			
	2-edge STD MOSFET	% of STD MOSFET	ARCELT Areas	% of STD MOSFET
1	108	100%	288	267%
2	180	167%	360	333%
3	252	233%	360	333%
4	324	300%	400	370%
5	396	367%	400	370%
6	468	433%	400	370%
7	540	500%	400	370%
8	612	567%	400	370%

Table 1 presents the minimum area required by number transistor using layout standard unit of λ

(lambda). Areas for ELT and ARCELT are given for compound MOSFET when number of devices are greater than 1. The table shows a definite advantage by saving silicon area when the number of devices increases for these compound devices. For example, when the ARCELT has 8 transistors within, the average area of a transistor is 50 sq-λ compared to 108 sq-λ of the 2-edge standard MOSFET. It is a 53% reduction of area for ARCTELT compared to that of 2-edge standard MOSFET. Thus ARCELT makes significant improvement of device density in designing and compactness of ICs.

B. Aspect Ratio, W/L

Circuits can easily be implemented with compound devices when a number of transistors share a common source or drain and also the same signal applied their gates. The aspect ratio, W/L is not so easy to determine in these compound MOSFETs. To find it accurately is necessary to run device simulations for driven currents and internal resistances of the current path of transistors. The simple compound MOSFET presented here is symmetric. The device size depends on both width W and length L of the channel of the ARCELT. It has one exception that value of W depends on value of L. So basically L is the limiting factor for this MOSFET. There are process dependent minimum values for L, along with other layout parameters that are required during IC design. A close approximation of W/L can be calculated as:

$$\text{Channel area } A = \sum \text{ Area of each rectangle x } \lambda^2$$

$$W = \sum \text{ Length of each sides of channel x } \lambda$$

$$\therefore \text{Aspect ratio} = \frac{W}{L} = \frac{W^2}{A}$$

Applying this equation of aspect ratio, at W = 2 λ for the ARC region, L of ARC is found to be 3.22 λ. Then again using these values, the W of electrodes is found to be 16 λ.

Both W and L are important to know for circuit simulations. Gate capacitance depends on these values for simulation. Circuit's frequency response, switching delays, dynamic currents etc., depend on these values.

IV. Conclusions

Applying ARCs in ELT produces compound MOSFET named ARCELT. Applications are presented how this compound MOSFET can be used in the design of circuits. Extraction technique for equivalent device of ARCELT is described. Plot of SPICE simulation result show that ARCELT functioned properly in the circuit. Analysis of area is completed and compared with 2-edge standard MOSFET. Design density is found to improve in IC's and thus cost of silicon fabrication can be reduced significantly. Relationship between number of ARCs and the number of electrodes is found for the ARCELT. Also the number transistors produced by the number of ARCs in the compound ARCELT is formulated in this paper. With 4 ARC cuts, the ARCELT can have 8 transistors that can reduce transistor size by more than 50% compared to that of 2-edge standard MOSFET. This paper also discusses importance of aspect ratio for effective cost of IC. Also a technique to approximate closely how to find the aspect ratio is described in this paper.

V. Acknowledgments

This work was supported by the National Science Foundation (NSF), under award number 0531507. Any opinions, findings, conclusions and recommendations expressed by the authors do not necessarily reflect those of the NSF.

Authors also like to thank Dr. Ron. D. Schrimpf at Vanderbilt University, Tennessee, USA, for his advice about effects on HBD MOSFETs in this work.

VI. References

[1] W.J. Snoeys, T.A.P. Gutierrez, G.A. Anelli, "New NMOS Layout Structure for Radiation Tolerance**",** IEEE Transactions on Nuclear Science, Volume 49, Issue 4, Part 1, Aug. 2002, Page(s): 1829 – 1833.

[2] G. Niu, J. Suraj, G. Banerjee, J. D. Cressler, S. D. Clark, M. J. Palmer, S. Subbanna, "Total Dose Effects on the Shallow-Trench Isolation Leakage Current Characteristics in a 0.35um SiGe BiCMOS Technology", IEEE Transactions on Nuclear Science, Volume 46, Issue 6, Dec. 1999, Page(s): 1841 – 1847.

[3] S. Binzaid, J. O. Attia, and R. D. Schrimpf, "An Active-Region-Cutout-Transistor (ARCT) Apparatus for Minimizing Leakage Current in Radiation Environments", US Provisional Patent, US 61/004,429, Accepted and Recorded on 28 Nov. 2007.

[4] S. Binzaid, J. O. Attia, "Enclosed Layout Transistor with Active Region Cutout", IEEE Region 5 P-Basics2 Conference, Apr. 2008, Pages(s): 22-26.

[5] J. Vollrath, "Signal margin analysis for DRAM sense amplifiers", The First IEEE International Workshop on Electronic Design, Test and Applications, 2002 Proceedings, 29-31 Jan. 2002, Page(s): 123 – 127.

[6] J. T. Wu, B. A. Wooley, "A 100-MHz Pipelined CMOS Comparator", IEEE Journal of Solid-State-Circuits, Voume 23, Issue 6, Dec. 1998, Page(s): 1379-1385.

[7] S. Binzaid, J. O. Attia, "Configurable Active-Region-Cutout-Transistor for Radiation Harened Circuit Applications", IEEE Canadian Conference on Electrical and Computer Engineering, May 2008, Page(s) 1215-1218.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

1.55 μm Laser Using InN-Based Quantum Well Heterostructure

Md. Tanvir Hasan[1], Md. Azim Ullah[2], Md. Asaduzzaman[2], and Ashraful G. Bhuiyan[2]

[1]Dept. of Electronics & Telecommunication Engineering, Faculty of Science & Information Technology, Daffodil International University (DIU), Dhaka – 1207, Bangladesh

[2]Dept. of Electrical & Electronic Engineering, Khulna University of Engineering and Technology (KUET), Khulna-9203, Bangladesh

E-mail: tan_vir_bd@yahoo.com

Abstract - The 1.55 μm semiconductor laser has recently been the subject of much research effort around the world. In this work a novel InN-based 1.55 μm quantum wll laser has been proposed. In order to reach the threshold current a very small voltage of 1.10 V is required. A very small threshold current of 7.3 mA is required to emit the light. The calculated values of threshold current and threshold voltage are lower than the reported values of the conventional 1.55 μm laser. The spectral range of the optical material gain is also found to be narrow which compensate to all optical losses. The above study indicates that the proposed InN-based 1.55 μm laser is very promising for the fabrication of future high performance laser.

I. Introduction

With the age of information technology well upon us, there is a growing demand for high-performance low-cost communication systems that can enable higher data rates and span longer distances. Fiber optic networks are the most promising solution for transmission of large amounts of information due to their low attenuation and large bandwidth capabilities. A basic fiber optic system consists of a transmitter, a communication channel, and a receiver. For high speed applications, semiconductor lasers are ideal candidates as transmitters. Several advantages of semiconductor lasers are their high efficiency, low-cost fabrication, high reliability, and direct modulation capability. The 1.55μm lasers have recently been the subject of much research effort around the world. It is a promising solution for low attenuation, high performance, long distance communication by fiber optic link.

The existing materials used to design the optical fiber lasers are ternary, quaternary and five elements including GaInNAsSb/GaAs [1], GaInNAsSb/GaNAs [2], InGaAs/InGaAsP [3] and AlGaInAs [4]. These materials are difficult to grow with constituent element which leads to degrade the quality of the material, creates lasing problem, and reduces the device performance and lifetime. To overcome these difficulties, researchers are looking for alternative candidates. Recently, it is reported that InN has band gap energy around 0.7-0.9eV and its alloy with GaN extended the energy band gap up to 3.4eV [5], which is compatible with the wavelength of the optical fiber. Good quality film of InN or its alloy with GaN is now routinely obtained. Therefore, it has very important potential to fabricate high speed laser diodes due to its low electron effective mass and high electron saturation velocity which can replace the conventional semiconductor laser technology in the optical communication system. However, compared with the commercialized 1.55 μm lasers diode, InN-based laser diode are very much immature. Attention should be given in the study and applicability of InN-based semiconductor laser for optical communication. The main goal of this work is to theoretical design of 1.55 μm lasers using InN-based quantum well heterostructures. These include the theoretical analysis and calculation of light output power, threshold current, optical gain and efficiency.

II. Device Structure

The simple typical schematic structure of a InN-based 1.55 μm laser, which is considered in this study, is shown in Fig. 1. The 35 Å active layer of InN or its alloys (1.55 μm), single quantum well (SQW) is surrounded on both sides by 70 Å $In_{0.15}Ga_{0.85}N$ guiding layer that are embedded in a GaN cladding layer and finally a 0.6 μm-thick $In_{0.15}Ga_{0.85}N$ contact layer. Thickness of active layer is chosen on the basis of optimization of threshold current and optical confinement factor. A 300 μm long cavity

p-$In_{0.15}Ga_{0.85}N$ contact layer (0.6μm)
p-GaN cladding layer (0.1μm)
p-$In_{0.15}Ga_{0.85}N$ guiding layer (70Å)
Active layer (InN or alloys of 0.8 eV)(35Å)
n-$In_{0.15}Ga_{0.85}N$ guiding layer (70Å)
n-GaN cladding layer (0.1μm)
n-$In_{0.15}Ga_{0.85}N$ contact layer (0.6μm)
C-sapphire substrate (0.4μm)

Fig. 1 Schematic layer structure of InN-based 1.55 μm laser.

Table 1: Design parameters of the proposed InN based 1.55 μm laser

Spontaneous emission coupling factor, β_{sp}	10^{-5}
Volume of active layer, V_A	2.5 μm^3
Photon lifetime, $\tau_r = \tau_n$	1 ns
Temperature, T	300 K
Confinement factor, Γ_a	0.2
Refractive index, η_r	2.7
Sum of a contact and a bulk resistance	4 Ω

length, two mirrors with reflectivity 35% and 100%, 1st and 2nd, respectively, are used in this device structure. The other required parameters which are considered for theoretical designing and performance evaluation are also shown in Table 1.

III. Results and Discussions

The light output power, threshold current, optical confinement factor, optical gain and efficiency are important parameters to specify the device performance of the laser. Figure 2 shows the shows predicted output injection current-voltage (I-V) characteristic of the newly proposed InN-based 1.55 μm laser at room temperature. In order to reach the threshold current a very small voltage of 1.10 V is required, as shown in Fig. 2. While other 1.55 μm laser such as GaInNAsSb/GaNAs-based quantum well laser has threshold current 12.64 mA and required threshold voltage 6.90 V [2].

The output power from the newly proposed InN-based 1.55 μm laser with the injection current for different temperature is shown in Fig. 3. The carrier concentration in the active layer is enhanced with increasing in the injection current into lasers. When the carrier concentration exceeds the threshold carrier concentration,

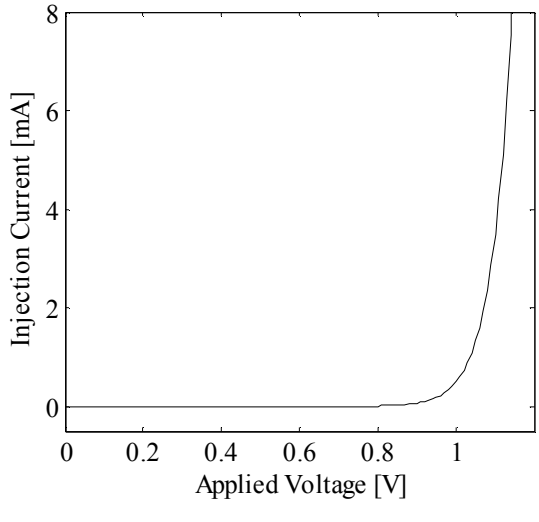


Fig. 2 Injection current vs. voltage (V-I) characteristics of the proposed InN-based 1.55 μm laser at 300 K.

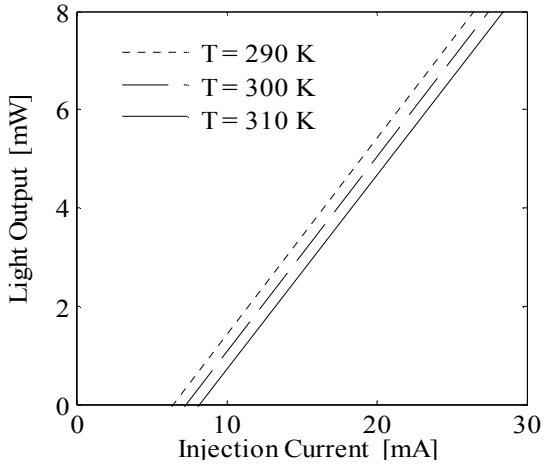


Fig. 3 The dependence of light output power on injection current for the proposed InN-based 1.55 μm laser at different temperatures.

laser oscillation starts and the light output drastically increases compared to below the threshold. The threshold current density increases and the external differential quantum efficiency decreases with increasing the temperature which is shown in Fig. 3. By increasing injection current above threshold, output power increase linearly in microwatt range. A very small threshold current of 7.3 mA is required to emit light.

If electrons are pumped in the conduction band and holes in the valance band, the electron-hole recombination process (photon emission) can be stronger than the reverse process of electron-hole generation (photon absorption). In general, the gain coefficient is defined by difference between the emission and absorption coefficient. The gain coefficient and the confinement factor of the quantum-well active layer are dependent on

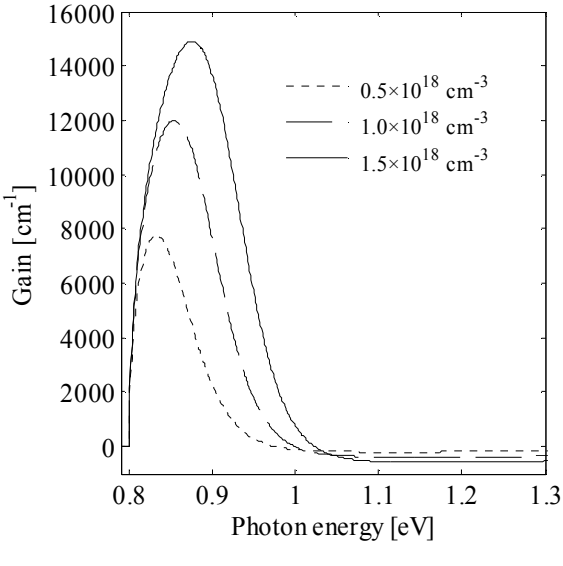


Fig. 4 Gain spectra for different carrier concentrations at 300 K.

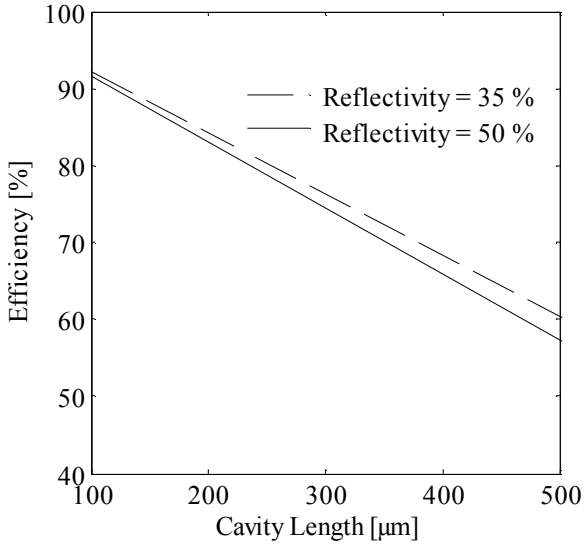


Fig. 6 The efficiency vs. cavity length for different reflectivity.

the carrier density. Figure 4 shows the variation of optical gain with photon energy with different carrier concentration at room temperature. The spectral range of the optical material gain is found to be narrow which compensate to all optical losses.

The effects of cavity length on the efficiency have also been studied. With increase of the cavity length from 100 to 500 μm efficiency decreases from 90 to 60% for reflectivity at 35% of one mirror and 100% shown in fig.6. Efficiency also increases with decreasing the reflectivity.

IV. Conclusion

We have theoretically designed InN-based 1.55μm Laser and evaluated its performance. In order to reach the threshold current a very small voltage of 1.10 V is required while other 1.55 μm laser such as GaInNAsSb/GaNAs-based quantum well laser has threshold current 12.64 mA and required threshold voltage 6.90 V. The threshold current density increases and the external differential quantum efficiency decreases with increasing the temperature. A very small threshold current of 7.3 mA is required to emit light. The spectral range of the optical material gain is found to be narrow which compensate to all optical losses. With increase of the cavity length from 100 to 500 μm efficiency decreases from 90 to 60% for reflectivity at 35% of one mirror and 100%. Efficiency also increases with decreasing the reflectivity. The above calculated results indicate that the proposed InN-based 1.55μm lasers are very promising for the fabrication of high performance and low cost devices. Therefore, more attention should be given on it.

References

[1] L. L. Goddard, S.R. Bank, M.A. Wistey, H.B. Yuen, J. S. Harris, "High Performance GaInNAsSb/GaAs Lasers at 1.5μm," Proc. SPIE, 5738, p. 180, April 2005.

[2] Robart P. Sarzala and Wlodzimierz Nakwaski, "GaInNAsSb/GaNAs quantum-well VCSELs: Modeling and physical analysis in the 1.50−1.55 μm wavelength range", J. Appl. Phys. vol.101, no. 7, p. 073103, April 2007.

[3] K. Uomi, M. Aoki, T. Tsuchiya, A. Takai**,** "Dependence of High-speed Properties on the Number of Quantum Wells in 1.55 μm InGaAs-InGaAsP MQW λ/4-shifted DFB Lasers", J. of Quantum Electronics, vol. 29, no. 2, pp. 355-360, February 1993.

[4] L. L. Goddard, S.R. Bank, M.A. Wistey, H.B. Yuen, Z. Rao, J. S. Harris, "Recombination, Gain, Band Structure, Efficiency, and Reliability of 1.5 μm GaInNAsSb/GaAs Lasers," J. Appl. Phys., vol. 97, no. 8, p. 83101, April 2005.

[5] Ashraful G. Bhuiyan, K. Sugita, K. Kasashima, A. Hashimoto, A. Yamamoto, V. Yu. Davydov, "Single-crystalline InN films with an absorption edge between 0.7 and 2 eV grown using different techniques and evidence of the actual band gap energy", Appl. Phys. Lett., vol. 83, no. 23, pp. 4788-90, December 2003.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effects of Phosphorus Doping on J-V and C-V Characteristics of Pulsed Laser Deposited Camphoric Carbon/P-Silicon Heterojunction Device

M. Zahurul Islam[1,2] and Sharif Mohammad Mominuzzaman[1]

[1]Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh,
[2]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada
E-mail: momin@eee.buet.ac.bd

Abstract - The current density-voltage (J-V) and capacitance voltage (C-V) characteristics of a p-n heterojunction device are studied. The device was fabricated by depositing phosphorus (P) doped carbon (C) thin film on p-type silicon (Si) substrate by pulsed laser deposition (PLD) technique at room temperature. Camphor ($C_{10}H_{16}O$), a natural source, was used as the starting precursor for the carbon layer of the heterostructure. Carbon layers of the device were obtained using target containing different amount of P. Both the J--V and C-V characteristics reveal that the device behaves as a successful p-n junction device up to a certain P content of the target material for C layer, and this is near about 5% of P by mass. At higher P content of the target, the material properties of the grown C layer on Si change and the device performance deteriorates. The built-in potential of the device with varying P content of the target material for C is also estimated and found them to agree well with the experimentally obtained device characteristics.

I. Introduction

Heterojunctions and multi planer structures in VLSI technology created a wide range of possibilities for device development. Some compound semiconductor materials are used with Si to create these structures in practice. But low cost, stable and highly efficient semiconductor devices are yet to be commercially realized due to their high material and production costs. However, with the emergence of carbon (C) as a semiconductor material the situation is expected to be changed. The enthusiasm in carbon is for the following reasons, such as, it is abundantly available in the earth in crystalline and numerous amorphous forms and it exhibits unique properties in favour of its application in the fabrication of heterojunction devices in conjunction with Si. The most important and interesting feature of carbon is that its optical gap (and hence other opto-electronic properties) can be tuned over an unusual wide range from that of semi metallic graphite (~0.0 eV) to that of insulating diamond (~5.5 eV) by varying the ratio of sp^3 and sp^2 hybridized bonds in its structure. So, now-a-days C has attracted the attention of the researchers for its application in electronic devices. C based heterostructures such as, metal insulator semiconductor (MIS) diodes, Schottky diodes, heterojunction diodes [1-3] on silicon have already been reported and thereby demonstrate the potentiality of application of C materials in electronic devices.

Previous study reveals that the energy of the C species generated by different deposition method varies and plays an important role in controlling the sp^3/sp^2 ratio and hence the optical gap of the carbon. The population of sp^3 and sp^2 bonds in precursor material also dictates the sp^3/ sp^2 ratio. So, the properties of grown carbon film depend on the method of deposition and the precursor material used [4]. Pulsed laser deposition (PLD) technique used for thin film deposition has become popular for its simplicity, versatility and capability to generate highly energetic carbon species with large sp^3 hybridized bonds, which enhances the synthesis of high quality films with good mechanical and opto-electrical properties. However the film remains amorphous and structure is complex. Again, usually graphite was used as the precursor of carbon in PLD technique. But, carbon thin film obtained by using PLD technique with camphor, a natural source, as the precursor of carbon is reported to have better optoelectrical properties [4,5]. The presence of huge sp^3 sites in camphor molecule helps in depositing high quality diamond like carbon (DLC) film and the presence of hydrogen ions reduces the defects of the film by passivating the dangling bonds. Undoped carbon is reported to be lightly p-type in nature [6]. Veerasamy *et al.* reported n-type doping in C using P powder [5] and nitrogen (N) gas [7]. We have used PLD technique to make *n*-C/*p*-Si device by depositing carbon films on Si (boron-doped) at room temperature using camphor as the starting precursor of carbon and P powder as the dopant source. In the present paper, we have investigated the experimentally obtained current-voltage (J-V) and capacitance-voltage (C-V) characteristics of this device to find the effect of P doping.

II. Experimental Details

Camphor has been used as the source of carbon for the carbonaceous thin film of the heterostructure. Details of

978-1-4244-2014-8/08/$25.00 (c)2008 IEEE

the chemical structure of camphor, camphor burning system and the target preparation method have been described elsewhere [5]. In brief, camphor was burnt in a 1-metre-long and 11-cm-diameter quartz tube.

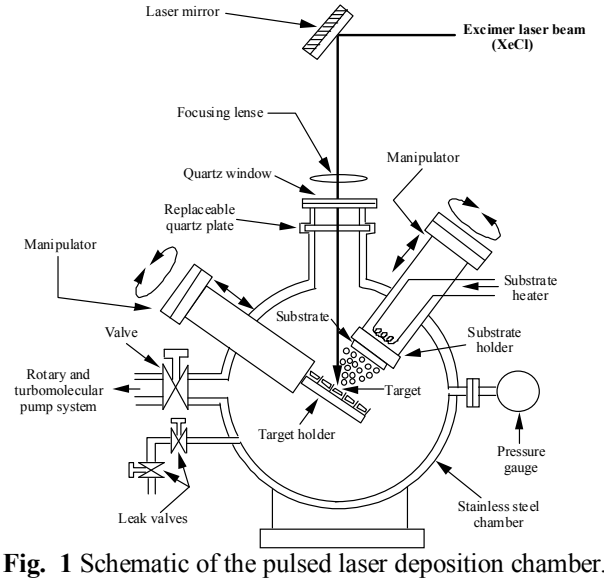


Fig. 1 Schematic of the pulsed laser deposition chamber.

The soot deposited along the walls of the tube was collected, dried in the oven for an hour. In order to dope, the soot was mixed with varying amount of red P powder (1%, 3%, 5% and 7% by mass) and compressed into pellets. These pellets were used as targets for P doped carbonaceous thin films in PLD Chamber. Carbon thin films were deposited on boron doped Si substrate by excimer laser (NISSIN 10X, XeCl, λ= 308 nm, τ= 20 nsec, repetition rate = 2 Hz, spot size = 5.5 mm^2), which is focused on the target at an incident angle of 45° to the target normal. The substrate was mounted parallel to the target at a distance of 45 mm. The films were deposited at room temperature at a base pressure of 10^{-6} Torr. The laser pulse energy was 150mJ on the window. The schematic of the PLD chamber is shown in Fig. 1. Gold electrode of about 15 nm was deposited on carbon film and that of about 100 nm is deposited on Si by conventional electron beam evaporation method. The contacts are found to be ohmic. The schematic of the n-C/p-Si heterostructure is shown in Fig. 2. The current density-voltage (J-V) and the capacitance-voltage characteristics of the device are obtained by using standard measurement techniques.

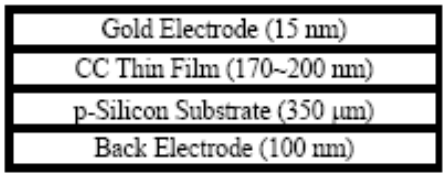


Fig. 2 Schematic of n-C/p-Si heterojunction device.

III. Results and Discussion

The forward and reverse J-V characteristics of the n-C/p-Si heterojunction device, where the C layer was deposited from the target containing 1%, 3%, 5% and 7% of P by mass, are shown in Fig. 3. It is evident from Fig. 3 that the fabricated devices are rectifying in nature, because the characteristics seem to be similar to the typical diode J-V characteristics. It also shows that for a fixed applied

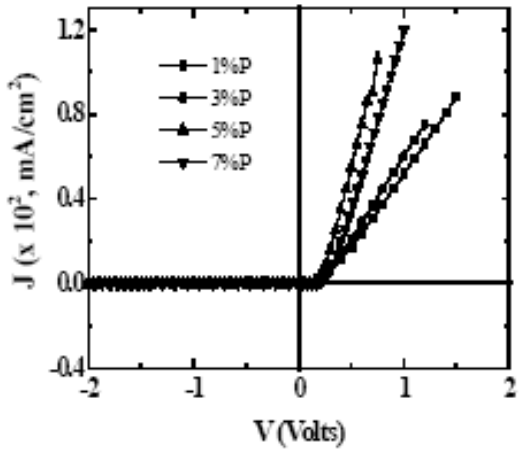


Fig. 3 J-V characteristics of *n*-C/*p*-Si, where C layer is grown from target containing different amount of P (1%, 3%, 5% and 7% by mass).

voltage in the forward bias region, the value of current density, J increases gradually with the increase in P content up to 5% by mass. This signifies that, the rectifying quality improves up to this point of dopant concentration and deteriorates thereupon. This phenomenon may be associated with the fact that the optical gap of the carbon layer decreases for 7% of P (reported earlier) due to graphitisation of C during the growth process [8].

The C-V characteristics of the device, as a function of %P content in the target of C layer (1%, 3%, 5% and 7% by mass) are shown in Fig. 4. It is seen that for a fixed applied reverse bias, the value of capacitance decreases with increase in the P content from 1% to 5% and then interestingly increases for 7% P. This may be attributed to

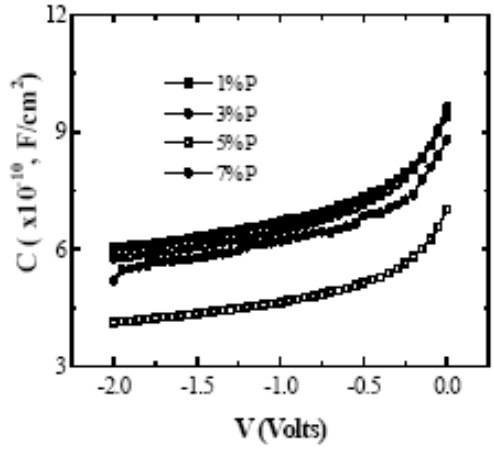


Fig. 4 Capacitance vs. voltage characteristics for n-C/p-Si heterostructure where carbon film is grown from target containing 1%, 3%, 5% and 7%phosphorus.

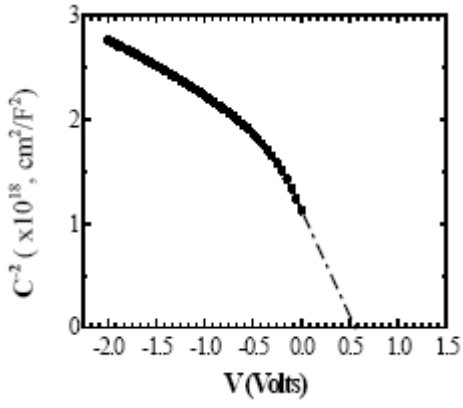


Fig. 5 Extrapolation of the low voltage linear region of the C^{-2}-V curve to the C^{-2}=0 line for the device where carbon layer is grown from the target containing 1% of P.

the same reason, i.e. for 7% P the optical gap of the grown C layer of the heterojunction decreases [8], so the properties of the C layer (becomes more graphitic) and hence the device characteristics of this sample are different from that of other samples.

For a *p*-*n* heterojunction device, the junction depletion capacitance per unit area is given by (1) [9],

$$C = \sqrt{\frac{qN_{D1}N_{A2}\varepsilon_1\varepsilon_2}{2(\varepsilon_1 N_{D1} + \varepsilon_2 N_{A2})(V_{bi} - V)}} \quad (1)$$

where, ε_1 is the dielectric constant of *n*-type material, ε_2 is the dielectric constant of *p*-type material, q is the elemental charge, N_{D1} is the donor concentration, N_{A2} is the acceptor concentration, V_{bi} is the built-in potential of the heterojunction and V is the applied bias. For a particular sample of the *p*-*n* heterojunction device, equation (1) can be written as,

$$C^{-2} = K(V_{bi} - V) \quad (2)$$

where, K is constant. Fig. 4 shows the C^{-2}-V curves for the device that was fabricated using 1% P. The value of C^{-2} increases linearly with the reverse bias voltage. But, at lower voltage levels, the C^{-2}-V plot shows a non-linear behaviour. This is usually attributed to the existence of deep levels [10]. To estimate the built-in potentials of the four samples of *n*-C/*p*-Si heterostructures, C^{-2}-V curves at low voltage linear regions are extrapolated to the C^{-2}=0 line, as suggested by the equation (2). Fig. 5 includes this extrapolation for the device where C layer is grown from the target containing 1% of P.

The values of the built-in potential obtained by extrapolating the C^{-2}-V curves of the four samples are shown in Fig. 6. It shows that the built-in potential of the device decreases with the increase in % P content except for the sample with the 7% P content. This anomaly for 7% P may be due to the same reason as stated earlier, i.e. the change in the optical gap of the carbon layer [8] and thereby changing the properties of the grown carbon layer and hence the built-in potential of the device. The validity of the relative values of the obtained built-in potentials can further be enhanced by a close inspection of the J-V characteristics (Fig. 3) near the knee regions.

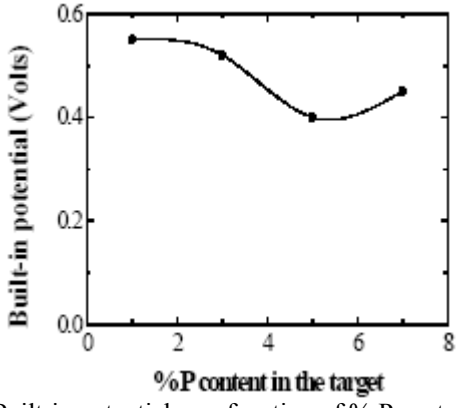


Fig. 6 Built-in potential as a function of % P content in the target from where the C layer is grown.

IV. Conclusions

P doped C thin films (*n*-type) are deposited on *p*-type Si substrate by pulsed laser deposition technique using a camphoric carbon soot target with varying amount of P. Both the J-V and C-V characteristics of the device, as a function of % P content in the target for carbon layer, can be understood well on the basis of the previously reported results of the structural and optical studies of the device. The obtained built-in potential values are also in good agreement with experimental observations. The results indicate that the characteristics of the heterojunction diode improve with the increase in P content of the target material for carbon up to 5% by mass and deteriorate thereupon. The results found are expected to be useful for understanding the behaviour of P doped C thin films and its practical implementation in electronic device application.

V. Acknowledgements

The authors would like to thank Prof. Tetsuo Soga, Nagoya Institute of Technology, for providing the data to make analyses to understand the device characteristics in the present work.

References

[1] V. S. Veerasamy, G. A. J. Amartunga, J. S. Park, H. S. Mackenzie and W. I. Milne, "Properties of n-Type Tetrahedral Carbon (ta-C)/p-Type Crystalline Silicon Heterojunction Diodes" IEEE Trans. on Electron Devices, vol. 42, pp- 577, 1995

[2] N. Konofaos and C. B. Thomas, "Characterization of heterojunction devices constructed by amorphous diamondlike films on silicon," J. Appl. Phys., vol 81, no. 9, pp. 6238-6245, 1997.

[3] S. M. Mominuzzaman, M.Rusop, T. Soga, T. Jimbo and M. Umeno, "Nitrogen Doping in Camphoric Carbon Films and its Application to Photovoltaic Cell", Solar Energy Materials and Solar Cells, vol. 90, no. 18-19, pp.3238-3243, 2006.

[4] S. M. Mominuzzaman, T. Soga, T. Jimbo and M. Umeno, "Camphoric carbon soot: a new target fordeposition of diamond-like carbon films by pulsed laser ablation," Thin Solid Films, vol. 376, pp. 1-4, 2000.

[5] S. M. Mominuzzaman, H. Ebishu, T. Soga, T. Jimbo and M. Umeno, "Phosphorus doping and defect studies of diamond like carbon films by pulsed laser deposition using camphoric carbon target," Diamond and Related Materials, vol. 10, pp. 984-988, 2001.

[6] Veerasamy, V. S., Amaratunga, G. A. J., Davis, C. A., Timbs, A. E., Milne, W. I. and Mackenzie, D. R., "n-Type Doping of Highly Tetrahedral Diamond-Like-Amorphous Carbon," J. Phys.: Condens. Matter 5, L169-L174, 1993.

[7] Veerasamy, V. S., Yuan, J., Amaratunga, G. A. J., Milne, W. I., Gilkes, K. W. R., Weiler, and M., Brown, L. M., "Nitrogen Doping of Highly Tetrahedral Amorphous Carbon," Phys. Rev. vol. B, no. 48, pp. 17954-17959, 1993.

[8] Mominuzzaman, S.M., Krishna, K. M., Soga, T., Jimbo, T. and Umeno, M., "Optical Absorption and Electrical Conductivity of Amorphous Carbon Thin Films from Camphor: a Natural Source," Jpn. J. Appl. Phys., vol. 38, no. 2A, pp. 658-663, 1999.

[9] S. M. Sze, Physics of Semiconductor Devices, 2nd edition, John Willy and Sons, pp. 74-96, 1981.

[10] G. I. Roberts and C. R. Crowell, Solid State Electron, vol. 16, no. 29, 1973.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Effect of Gate Bias on Channel in Depletion-All-Around Operation of the SOI Four-Gate Transistor

Shafat Jahangir[1], Quazi Deen Mohd Khosru[1], and Anisul Haque[2]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh.
E-mail: sj_7117@yahoo.com
[2]Department of Electrical and Electronic Engineering, East West University, Dhaka 1212, Bangladesh.

Abstract– In depletion-all around (DAA) operation of SOI four-gate transistor (G^4-FET), the conducting channel can be surrounded by depletion regions induced by independent vertical MOS gates and lateral JFET gates. This enables majority carriers to flow through the volume of the silicon film far from both silicon/oxide and p+ gate/n-channel interfaces. A numerical model using FEMLAB with MATLAB is developed to obtain the potential distribution solving 2-D Poisson equation using finite element method. This model is extendable to fully depleted (FD) structure. Using this model, effect of gate bias on the location and size of the conducting channel is studied. Gradual change of the size of the conducting channel from drain to source is also studied when drain is positively biased. Under appropriate gate bias voltages, the cross-section of the channel may be made sufficiently narrow to invoke quantum mechanical effects.

I. Introduction

The G^4-FET is a silicon-on-insulator (SOI) transistor, which combines an SOI MOSFET and a lateral double-gate JFET [1]. G^4-FET comprises four gates in a single transistor. No extra fabrication steps are needed to incorporate G^4-FET-biased circuits with a standard partially or fully depleted (FD) SOI process. In G^4-FET, conduction parameters such as threshold voltage, trans-conductance, sub-threshold slope, etc. related to a single gate can be adjusted by the biases on the remaining gate when the device is driven from one gate keeping the other gates at constant voltages. When G^4-FET is driven from multiple gates simultaneously, multiple-input circuits with much reduced transistor count is obtained compared to the standard CMOS procedure. Thus multiple independent gates of the device enhance circuit design flexibility.

In G^4-FET, both surface and volume conduction modes can be achieved applying appropriate bias on the four gates. This paper is focused on a particular operation mode of the G^4-FET providing volume conduction, which is called depletion-all-around (DAA) [2]. Because of the use of extra MOS gates, this volume conduction mode gives rise to some better features than that of a single JFET. These include some enhanced electrical characteristics such as high intrinsic dc gain [3], low-noise operation [4], [5], and radiation hardness [6], [7]. Hence it is very attractive for analog circuits, where low-noise and/or radiation-hard operation is required.

So far, most of the works on G^4-FET have been done observing experimental data and using commercial software like Silvaco/Atlas. In some works, analytical models are proposed [8]. In this paper we develop a physics-based numerical model for G^4-FET to obtain potential distribution for the modeling of the electrostatic parameters and current-voltage characteristics. We determine the potential distribution by solving 2-D Poisson equation numerically using finite element method. We also investigate the influence of gate bias on the size and location of the conducting channel.

II. G^4-FET Structure and DAA operation

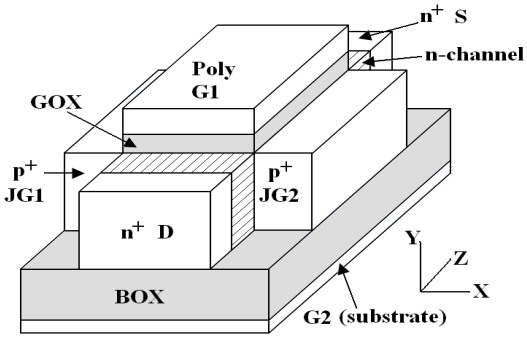
The G^4-FET is a double-gate MOSFET comprising two lateral junction-gates (JG1 and JG2) or reciprocally it is a lateral double-gate JFET comprising two vertical MOS gates (poly silicon top-gate G1 and back-gate G2) (Fig. 1).

Normally the junction gates are reverse biased with respect to the channel and the drain-current I_D comprises of majority carriers. The structure of an n-channel G^4-FET is the same as that of a p-channel inversion-mode SOI MOSFET with two body contacts on each side of the channel (Fig.1): These body contacts and source/drain of the inversion-mode MOSFET are used in the G^4-FET as source/drain and junction gates, respectively.

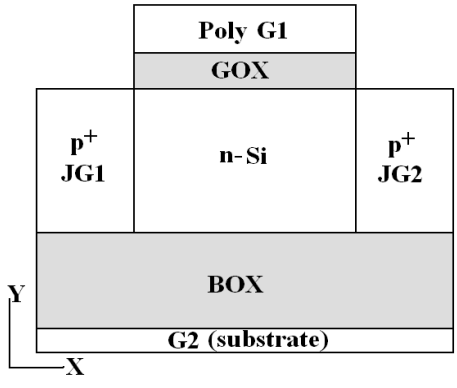
In DAA operation, all four gates are negatively biased in case of an n-channel G^4-FET inducing depletion at both silicon/oxide and n-silicon/p+ junction interfaces. Consequently, I_D flows through the center of the silicon film, and the conducting channel is surrounded by depletion regions. If these negative biases are sufficiently strong, the device performance and properties in DAA mode are remarkably affected by the presence of inversion layers under MOS gates because then both the junction gates get interconnected through the inverted silicon film.

In our work 2-D simulation is done on a partially depleted DAA device structure with a channel width of w_si = 90 nm, a film thickness of t_si = 60 nm, a gate oxide thickness of t_gox = 15 nm and a buried oxide

thickness of t_box = 40 nm. The channel doping density is $N_D = 5\times10^{17}$ cm^{-3} in the n-type device. The doping density in p+ junction gates is $N_A = 2\times10^{20}$ cm^{-3}.



(a)



(b)

Fig. 1: (a) Schematic structure of the n-channel G^4-FET, (b) Cross section of the device (drain and source are not shown in this case)

III. Finite element solution

We want to calculate potential profile at any cross-section between drain and source from depletion charge distribution in case of n-channel G^4-FET. 2-D Poisson equation is given in equation (1).

$$-\in_0\in_r\left[\frac{\partial^2V(x,y)}{\partial x^2}+\frac{\partial^2V(x,y)}{\partial y^2}\right]=\rho_{depl} \qquad (1)$$

Where, ρ_{depl} = Depletion charge density (qN_D) within the depletion region and zero within the neutral channel. $\in_0$ = Permittivity of free space, $\in_r$ = Relative dielectric constant.

We have used FEMLAB with MATLAB to solve Eq. (1) using finite element analysis. FEMLAB provides a powerful, interactive environment for modeling and solving scientific and engineering problems using finite element method [9].

Classical PDE in multi-physics mode is used for solving 2-D Poisson equation which is given in coefficient form in FEMLAB as,

$$-\nabla\cdot(c\nabla u)=f \qquad (2)$$

Comparing equation (2) with (1),

$c \equiv \in_0\in_r$, $u \equiv V(x,y)$, and $f \equiv \rho_{depl}$.

Since the profile of the depletion region for any given set of gate bias voltages is not known initially, we start with an arbitrary depletion charge distribution as input to the 2-D Poisson solver. In FEMLAB we do not define the depletion charge distribution in terms of sub-domains, rather we define depletion charge node by node i.e. any node within the depletion region contains charge qN_D and that in the channel contains zero charge. This is why our developed model can accurately work with all modes of operation. We reach the actual voltage profile through Gauss-Seidel iterative technique. We cannot use the usual analytical semi-classical equations of depletion width for reverse biased p+-n junction and MOS capacitor in this work because those equations cannot take into account 2-D effects such as charge sharing between the junction gates and the MOS gates.

At the external boundaries of the four gates we use Dirichlet i.e. fixed voltage boundary condition. Newmann i.e. continuous electric flux boundary condition is used at all internal boundaries. It is assumed that *flat band voltage* (V_{FB}) is zero.

As we consider heavily doped p+ junction gates ($N_A = 2\times10^{20}$ cm^{-3}) and since the applied bias voltages to poly silicon gates are quite small, it is assumed that depletion regions of reverse biased p+ junction/n-channel do not extend into p+ gates i.e. poly silicon *depletion effect* is neglected.

IV. Results and discussion

We present 2-D simulation results for n-channel G^4-FET. When gates are biased with sufficient negative voltage, depletion regions from all gates overlap and we get a *fully depleted* structure leaving no conducting channel from drain to source and the device goes to *off-state* (Fig. 2).

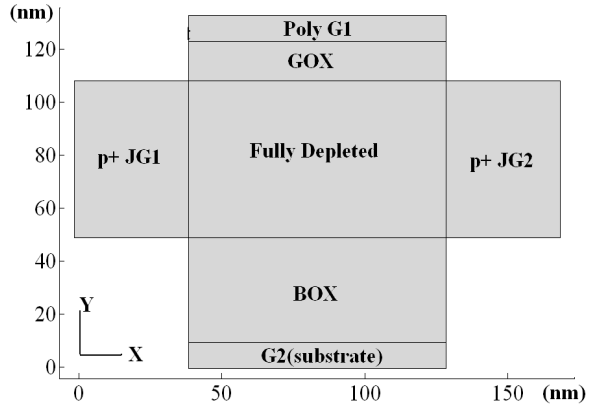


Fig. 2: 2-D simulation result showing fully depleted (FD) structure at any cross-section between drain and source with bias conditions: V_pg = -0.9 V, V_bg = -0.9 V, V_jg1 = -0.9 V, V_jg2 = -0.9 V and V_D = 0 V, V_S = 0 V.

By applying appropriate bias on all four gates, we can confine the channel at the center of the silicon film, thereby avoiding unwanted surface scattering and improving carrier mobility. We show such a structure when both drain and source are at 0 V in Fig. 3.

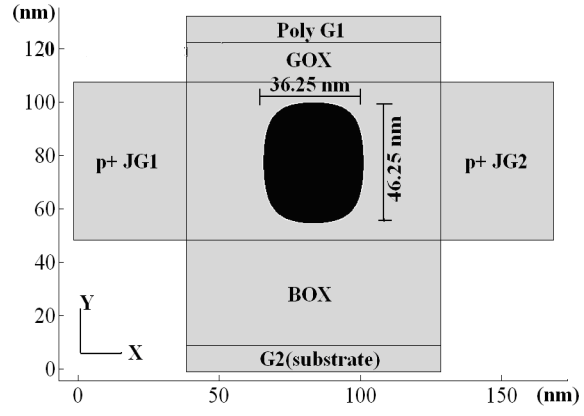


Fig. 3: 2-D simulation result showing position of channel (black region) at any cross-section between drain and source with bias conditions: V_pg = -0.28 V, V_bg = -0.55 V, V_jg1 = -0.28 V, V_jg2 = -0.28 V and V_D = 0 V, V_S = 0 V. Channel is at 0 V.

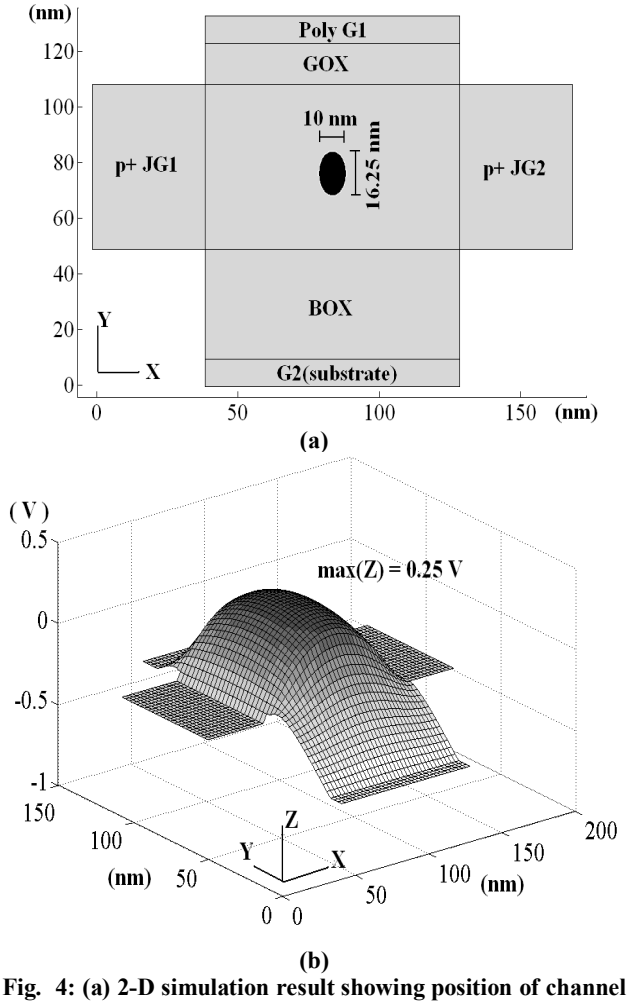


Fig. 4: (a) 2-D simulation result showing position of channel (black region) at drain terminal with bias conditions: V_pg = -0.28 V, V_bg = -0.55 V, V_jg1 = -0.28 V, V_jg2 = -0.28 V and V_D = 0.25 V, V_S = 0 V, (b) Corresponding potential distribution where the channel is at drain voltage (0.25 V).

Similar channel structure as Fig. 3 is observed in Fig. 4 at the drain end of the channel with V_D = 0.25 V and V_S = 0V. In this case the channel is contracted than before at the drain end because the n-channel/p+ junction at the drain side is more reverse biased than before. Increasing V_D leads to a state where the drain side is *pinched-off*. Moving forward towards source from drain terminal, keeping the bias conditions as stated above, we get gradually expanding channel at each subsequent cross-section and finally reach the structure as Fig. 3 at source terminal. This is because the voltage applied at the drain is dropped gradually as we travel from drain to source and becomes zero at the source.

Keeping the bias conditions as above, Fig. 5 presents a depletion-all-around channel structure in between drain and source where the channel is at 0.15 V. It is noticeable that this time the channel is larger than that at the drain but smaller than that at the source.

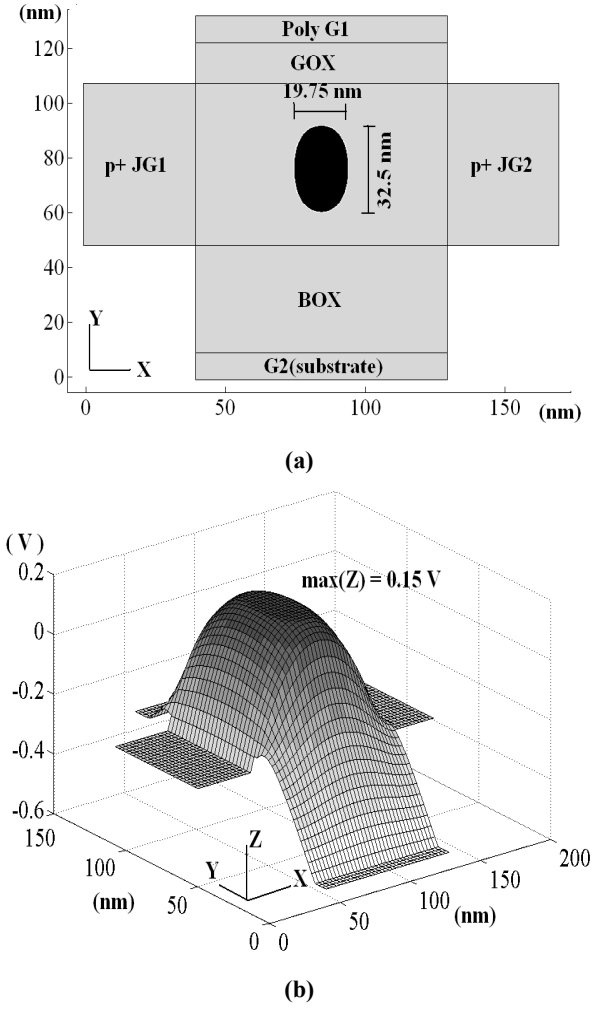


Fig. 5: (a) 2-D simulation result showing position of channel (black region) at a cross-section between drain and source with bias conditions: V_pg = -0.28 V, V_bg = -0.55 V, V_jg1 = -0.28 V, V_jg2 = -0.28 V and V_D = 0.25 V, V_S = 0 V, (b) Corresponding potential distribution where the channel at the cross section is at 0.15 V.

The position of the channel within the silicon film can be adjusted by adjusting the different gate bias

voltages. Fig. 6 shows such a scenario where the channel is shifted to the left.

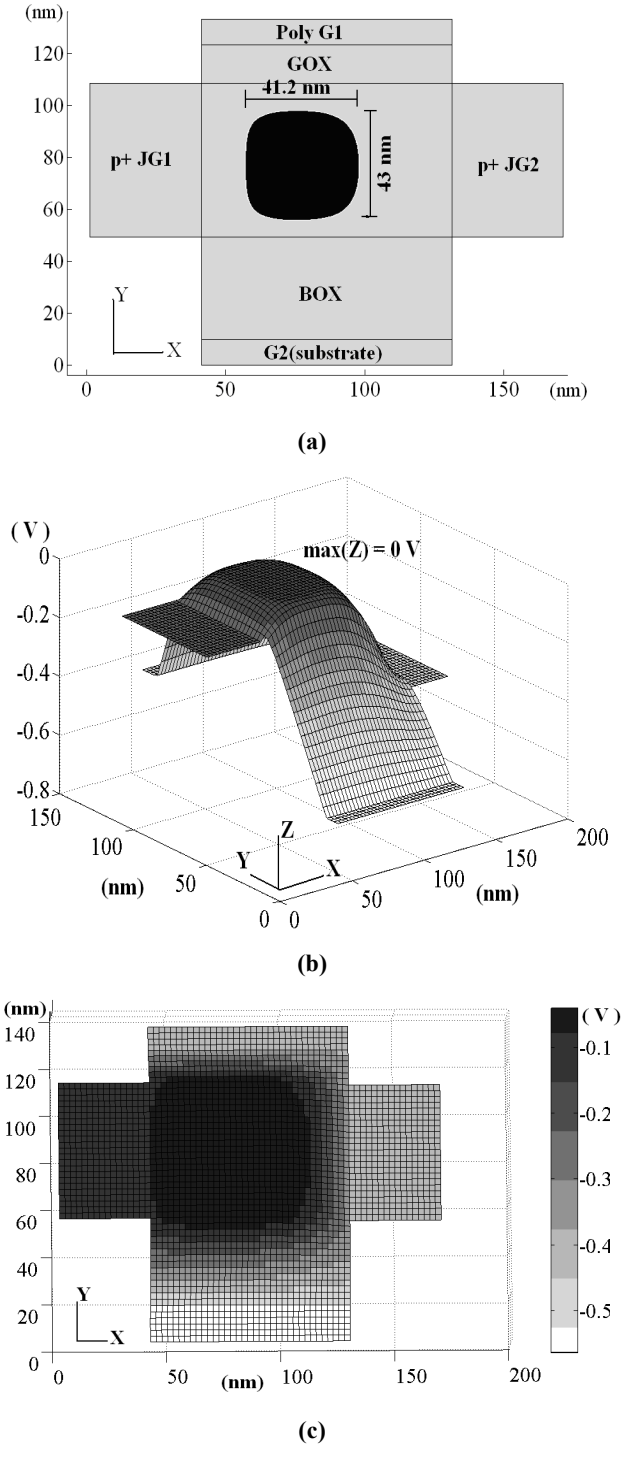


Fig. 6: (a) 2-D simulation result showing position of channel (black region) at any cross-section between drain and source with bias conditions: V_pg = -0.4 V, V_bg = -0.6 V, V_jg1 = -0.1 V, V_jg2 = -0.4 V and V_D = 0 V, V_S = 0 V, channel is shifted to the left and it is at 0 V, (b) Corresponding potential distribution, (c) 2-D surface plot of the same potential profile

Using our model in DAA operation, we can control the size and location of the conducting channel through appropriate gate bias to increase carrier mobility, thereby enhancing device performance. Besides we are able to confine the dimensions of the channel to less than 20 nm. For such dimension, electron energy gets quantized which gives rise to various new and interesting quantum mechanical effects. Study of those effects will be reported elsewhere.

V. Conclusion

We have developed a numerical model using FEMLAB with MATLAB to obtain the potential distribution solving 2-D Poisson equation using finite element method. We have studied the influence of different gate bias voltages on the location and size of the conducting channel, which also includes fully depleted condition for certain gate bias voltages. The developed model is used to investigate the gradual change of the size of the conducting channel from drain to source, when drain is positively biased. It is also found that by applying appropriate bias voltages on the four gates, the conducting channel can be narrowed down to less than 20 nm yielding various new quantum mechanical effects in carrier transport.

References

[1] B. J. Blalock, S. Cristoloveanu, B. M. Dufrene, F. Allibert, and M. Mojarradi, "The multiple-gate MOS-JFET transistor," *Int. J. High Speed Electron. Syst.*, vol. 12, no. 2, pp. 511–520, 2002.

[2] K. Akarvardar, S. Cristoloveanu, P. Gentil, R. D. Schrimpf, B. J. Blalock. "Depletion-all-around operation of the SOI four-gate transistor," *IEEE Trans. Electron Devices*, vol. 54, no. 2, pp. 323–331, Feb. 2007.

[3] K. Akarvardar, S. Cristoloveanu, P. Gentil, B. J. Blalock, B. Dufrene, and M. M. Mojarradi, "Depletion-all-around in SOI G^4-FETs: A conduction mechanism with high performance," in *Proc. 34th ESSDERC*, 2004, pp. 217–220.

[4] K. Akarvardar, B. Dufrene, S. Cristoloveanu, P. Gentil, B. J. Blalock, and M. M. Mojarradi, "Low-frequency noise in SOI four-gate transistors," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 829–835, Apr. 2006.

[5] J. A. J. Tejada, A. L. Rodríguez, A. Godoy, J. A. L. Villanueva, F. M. Gomez-Campos, and S. Rodriguez-Bolivar. "A low-frequency noise model for four-gate field-effect transistors," *IEEE Trans. Electron Devices*, vol. 55, no. 3, pp. 896–903, Mar. 2008.

[6] K. Akarvardar, S. Cristoloveanu, B. Dufrene, P. Gentil, R. D. Schrimpf, B. J. Blalock, J. A. Chroboczek, and M. M. Mojarradi, "Evidence for reduction of noise and radiation effects in G^4-FET depletion-all-around operation," in *Proc. 35th ESSDERC*, 2005, pp. 89–92.

[7] K. Akarvardar, S. Cristoloveanu, R. D. Schrimpf, B. Dufrene, P. Gentil, B. J. Blalock, and M. M. Mojarradi, "Total-dose radiation hardness of the SOI 4-gate transistor (G^4-FET)," in *Proc. ECS SOI Conf.*, 2005, vol. 2005-3, pp. 99–106.

[8] K. Akarvardar, S. Cristoloveanu, and P. Gentil. "Analytical modeling of the two-dimensional potential distribution and threshold voltage of the SOI four-gate transistor" *IEEE Trans. Electron Devices*, vol. 53, no. 10, pp. 2569–2577, Oct. 2006.

[9] *COMSOL Multiphysics homepage.* http://www.comsol.com

[10] K. Akarvardar, S. Chen, B. J. Blalock, S. Cristoloveanu, P. Gentil, and M. M. Mojarradi, "A novel four-quadrant analog multiplier using SOI four-gate transistors (G^4-FETs)," in *Proc. 31th ESSCIRC*, 2005, pp. 499–502.

[11] K. Akarvardar, S. Chen, J. Vandersand, B. J. Blalock, R. D. Schrimpf, B. Prothro, C. Britton, S. Cristoloveanu, P. Gentil, and M. M. Mojarradi, "Four-gate transistor voltage-controlled negative differential resistance device and related circuit applications," in *Proc. IEEE Int. SOI Conf.*, 2006, pp. 71–72.

[12] B. Dufrene, K. Akarvardar, S. Cristoloveanu, B. J. Blalock, P. Gentil, E. Kolawa, and M. M. Mojarradi, "Investigation of the four-gate action in G^4-FETs," *IEEE Trans. Electron Devices*, vol. 51, no. 11, pp. 1931–1935, Nov. 2004.

[13] S. C. Terry, S. Chen, B. J. Blalock, J. R. Jackson, B. M. Dufrene, M. M. Mojarradi, S. K. Islam, and M. N. Ericson, "Temperature compensated reference circuits for SOI," in *Proc. IEEE Int. SOI Conf.*, 2004, pp. 112–114.

[14] S. Chen, J. Vandersand, B. J. Blalock, K. Akarvardar, S. Cristoloveanu, and M. M. Mojarradi, "SOI four-gate transistors (G^4-FETs) for high voltage analog applications," in *Proc. 31th ESSCIRC*, 2005, pp. 311–314.

[15] B. Dufrene, B. J. Blalock, S. Cristoloveanu, K. Akarvardar, T. Higashino, and M. Mojarradi, "Subthreshold slope modulation in G^4-FET transistors," *Microelectron. Eng.*, vol. 72, no. 1–4, pp. 347–351, May 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Range Key Query Scheme for Multidimensional Databases

K. M. Azharul Hasan[1], Tatsuo Tsuji[2] and Ken higuchi[2]

[1]Department of Computer Science and Engineering, Khulna University of Engineering and Technology, Bangladesh
[2]University of Fukui, Fukui shi 910-8507, Japan.
E-mail: azhasan@cse.kuet.ac.bd, tsuji@pear.fuis.fukui-u.ac.jp, higuchi@pear.fuis.fukui-u.ac.jp

Abstract - In this paper, a new implementation scheme of range key query for multidimensional databases is proposed and evaluated. The scheme implements a multidimensional database by employing an extendible multidimensional array. By using multidimensional arrays, fast random addressing functions for element access can be invoked by knowing a tuple of subscripts of an array element. However these kinds of multidimensional arrays suffer from some problems. In our scheme, these problems are solved by an efficient scheme of record encoding based on the notion of extendible array. The scheme shows good retrieval performance for range key query compared with conventional implementation of RDMS.

I. Introduction

Today's organizations frequently make business decisions based on statistical analysis of their high dimensional enterprise data. This large-scale data needs to handle efficiently which is promoting extensive research themes on organization or implementation schemes for multidimensional array [1]-[3]. The multidimensional arrays must be extended in terms of the range of column values of each dimension because the business model changes frequently. For example adding new column values or adding new dimension to an existing multidimensional array needs the array to be extended.

In this paper, we propose and evaluate a new implementation scheme for range key query for multidimensional databases based multidimensional extendible array. Multidimensional arrays are good to store dense data, but most datasets are sparse which wastes huge memory because a large number of array cells are empty and thus are very hard to use in actual implementation. In particular, the sparsity problem increases when the number of dimensions increases. This is because the number of all possible combinations of dimension values exponentially increases, whereas the number of actual data values would not increase at such a rate. Efficient storage schemes are required to store such sparse data for multidimensional arrays. More over conventional schemes for storing multidimensional arrays do not support dynamic extension of an array and hence addition of a new column value is impossible if the size of the dimension overflows. In this paper the notion of extendible array is employed [4][5] to solve the dynamic extension problem. An extendible array is extendible in any direction without relocation of the data already stored. Such advantages make it possible for an extendible array to be applied into wide application area where necessary array size cannot be predicted and can be varied according to the dynamic environment during operating time of the system.

The data allocation is a key performance factor for multidimensional databases. This holds especially for data warehousing environments where huge amounts of data have to be dealt with. In this research a new data allocation scheme is used based on the notion of subarray The scheme stores only the effective array elements that exists as actual record which solves the sparsity problem above. It can be effectively applied to the implementation of multidimensional database systems [6][7], or data warehouse systems [8][9] for multidimensional analysis. There are some works [12]-[15] that can be found for multidimensional idea but those are not extensible.

II. Employing Multidimensional Extendible Array

The conventional storage allocation scheme for arrays is either row major or column major ordering. Though the allocation technique provides optimal storage utilization but the extension of the dimensions lacks in all but one dimension. Such asymmetry in extendibility is not inevitable. It is devised schemes [4][5] for multi dimensional storing arrays, which are readily extendible in all directions. An extendible array, however, does not store an individual array; rather, it is storing an array and all its potential extensions. The scheme is an *n* dimensional rectangular array that grows by adjoining blocks, which are subarrays of dimension *n-1*. Within which each subarray storage allocation is in row-major or Lexicographic order.

An *n* dimensional extendible array *A* has a history counter *h* and three kinds of auxiliary tables for each extendible dimension $i\,(i=1,...,n)$. See Fig. 1. These tables are history table H_i, address table L_i, and coefficient table C_i. The history tables memorize extension history *h*. If the size of *A* is $[s_n, s_{n-1},..., s_1]$ and the extended dimension is *i*, for an extension of *A* along dimension *i*, contiguous memory area that forms an $n-1$ dimensional subarray *S* of size $[s_n, s_{n-1},..., s_{i+1}, s_{i-1},..., s_2, s_1]$ is

dynamically allocated. Then the current history counter value is incremented by one, and it is memorized on the history table H_i, also the first address of *S* is held on the address table L_i. An element $< i_n,...,i_1 >$ in an n dimensional array of size $[s_n, s_{n-1},...., s_1]$ is allocated on memory using the addressing function like this:

$$f(i_n, i_{n-1},..,i_2, i_1) = s_1 s_2 ... s_{n-1} i_n + ... + s_1 i_2 + i_1$$

Here, we call $< s_1 s_2 ... s_{n-1}, s_1 s_2 ... s_{n-2},..., s_1 >$ as *coefficient vector*. Since a subarray of *A* is of $n-1$ dimensional fixed size, its coefficient vector is $n-2$ dimensional. If A's dimension is greater than 2, its coefficient table C_i's are prepared and each slot of C_i holds the corresponding coefficient vector.

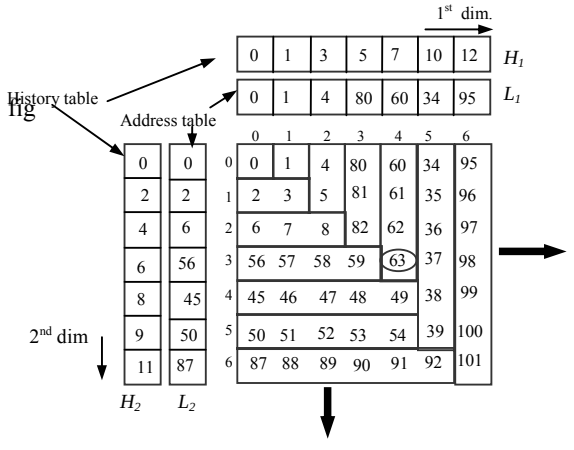


Fig. 1: Realization of 2 dimensional extendible array

Consider the element <4,3> in Fig. 1. Compare $H_1[4] = 7$ and $H_2[3] = 6$. Since $H_1[4] > H_2[3]$, it can be proved that the element <4,3> is involved in the extended subarray *S* occupying the address from 60 to 63. The first address of *S* is known to be 60, which is stored in $L_1[4]$. Since the offset of <4,3> from the first address of S is 3, the address of the element is determined as 63. Note that we can use such a simple computational scheme to access an extendible array element only by preparing small auxiliary tables.

The superiority of this scheme in element accessing speed and memory utilization is shown in [5] comparing with other schemes such as hashing [10].

III. The Implementation Model

The model is based on the extendible array and is presented in detail in [17]. Here in this paper the efficiency for range key query for multidimensional databases is presented. This section gives a very short over view of the model. Each column of a relational table corresponds to a dimension of the extendible array and each column value of a record is uniquely mapped to a subscript of the dimension. A subarray is constructed for each distinct value of a column. For a relational table *R* with *n* columns, the model logically corresponds to the pair (*M, A*). *A* is an *n* dimensional extendible array created for *R* and *M* is the set of mappings. Each m_i $(1 \le i \le n)$ in *M* maps the *i*-th column values of *R* to the subscripts of the dimension *i* of *A*.

Name	Age
Azam	21
Khair	22
Yamin	22
Hasan	26
Kamal	24

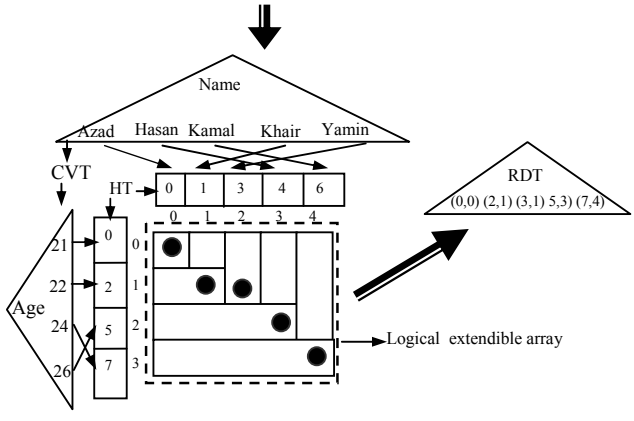


Fig. 2: The implementation model

In this technique, we specify an element using the pair of the history value and offset value. Since a history value is unique and has a one-to-one correspondence with the corresponding subarray, the subarray including the specified element of an extendible array can be referred to uniquely by its corresponding history value. Moreover, the offset value (i.e., logical location) of the element is also unique in the subarray. Therefore each element of an *n* dimensional extendible array can be referenced specifying the pair (*history value, offset value*).

From the logical structure (*M*, *A*) above, each mapping m_i in *M* is implemented using a single B^+-tree called CVT(key subscript ConVersion Tree), and the logical extendible array *A* is implemented using a single B^+-tree called RDT(Real Data Tree) and a single table (HT) for each dimension corresponding to the three auxiliary tables of an extendible array. An example of the physical implementation of the model is shown in Fig. 2. For example, the column value "Hasan" is mapped to the subscript 4 as the insertion order, though in the sequence set of CVT, the key "Hasan" is in position 2 due to the property of B^+-tree. Each column value in the record is inserted to the corresponding CVT as a key value. If the key value does not exist then the logical extendible array A is extended by one along the dimension and initialized.

The set of the pairs (*history value, offset value*) for all of the effective elements in the extendible array are inseted

into a B^+ tree as the key value called RDT. Here, the effective elements mean the ones that correspond to the records in the relational table. We assume that the key occupies the fixed size storage and the *history value* is arranged in front of the *offset value*. Hence the keys are arranged in the order of the history values and keys that have the same history values are arranged consecutively in the sequence set of RDT.

IV. The Range Key Retrieval Scheme

A range key query [11] has a single predicate of the form (*column name* < *value*) or (*column name* > *value*) or (*column name between value1 and value2*). Here range key query means to search records, some of whose column values are specified for a specific dimension. Let the specified range involve the known column values $v_1, v_2, ..., v_{NRQ}$ of dimension *k*. The subscripts of $v_1, v_2, ..., v_{NRQ}$ are determined as $I_{rq} = \{g_k(v_1), g_k(v_2), ..., g_k(v_{NRQ})\}$. The $v_1, v_2, ..., v_{NRQ}$ values are organized consecutively in the sequence set of CVT_k. Let $h_1, h_2, ..., h_{NRQ}$ be the history values that correspond to the subscripts $g_k(v_1), g_k(v_2), ..., g_k(v_{NRQ})$ and the minimum history value be $h_{\min} = \min(h_1, h_2, ..., h_{NRQ})$. Assumed that the candidate range of the converted subscripts of the corresponding dimension *k* has *NRQ* subscripts.Here $g_i(v_i)$ denotes the mapped subscript for the column value v_i in CVT_i. Figure 3 shows the candidate range (bold dotted line) of a range key query for a 2 dimensional extendible array.

Key matching : Let <*h,o*> be a key in the sequence set of RDT. From <*h,o*>, the corresponding tuple of the subscripts $< i_n, i_{n-1}, ..., i_1 >$ of the logical extendible array can be uniquely computed. For the specified key <*h,o*> the subscripts can be simply computed by repeated divisions by knowing the coefficient vector stored in in the auxilary table one for each dimension. See the addressing function described in Section 2. If the computed subscript of dimension *k* is i_k and $i_k \in I_{rq}$ then <*h,o*> proves to match condition for the range key retrieval.

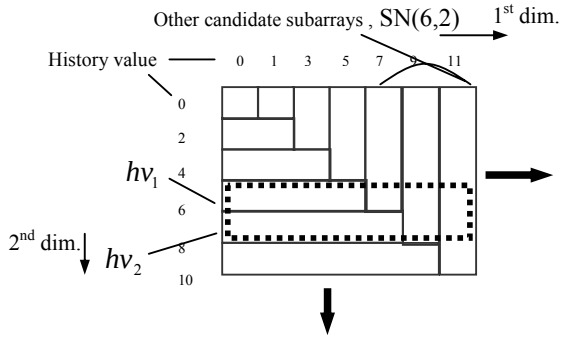


Fig. 3: Candidate subarrays

The Search Scheme: Among all the subarrays the candidate subarrays are searched only.

Definition 1 (Candidate subarray): Let h_{min} be the $h_{\min} = \min(h_1, h_2, ..., h_{NRQ})$. The candidate sub arrays that are to be searched are $h_1, h_2, ..., h_{NRQ}$ in the dimension k and the subarrays that have history value greater than $h_{\min}$ and do not belong to the known dimension *k* are the candidates. See Fig. 3. The subarrays corresponding to history values $h_1, h_2, ..., h_{NRQ}$ are termed as *pricipal subarrays.*

All the records belonging to the *principal subarrays* are included in the retrieval results for the query. Hence if the history values that correspond to the *principal subarrays* are encountered while the sequential traverse of the sequence set, then all the records of the subarray are retrieved. When other candidate subarrays are encountered in the traversing, they are searched as described bellow.

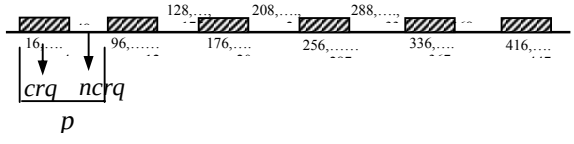


Fig. 4: Candidate and non candidate offsets

Example 1 : Consider a 5 dimensional subarray of size [2,3,5,4,4] and assume that the keys in RDT to be searched are corresponding to the elements $< x_5, x_4, a_3, x_2, x_1 >$ of the logical extendible array. Where $x_1 = 0,1,2,3$, $x_2 = 0,1,2,3$, $a_3 = 1,2$ (*NRQ*=2), $x_4 = 0,1,2$ and $x_5 = 0,1$. There are total $s_5 \times s_4 \times s_2 \times s_1 \times NRQ = 96$ offsets {16,17,…,47, 96,97,…,127, 175,177,…,207, 256,257,…,287, 336,337,…,367, 414,415,…,447 } are the candidates for searching in the subarray and they are arranged periodecally. Fig. 4 shows the ***candidate range of offsets*** (**crq**) and ***non candidate range of offsets*** **(ncrq)** .

The 5 dimensional offset in example 1 total 96 offsets are candidates with candidate and non candidate range of offsets are shown in Fig. 4. The period for the known subscript a_3 is

$p = s_1 \times s_2 \times s_3 = s_1 \times s_2 \times NRQ + s_1 \times s_2 (s_3 - NRQ)$, where $crq = s_1 \times s_2 \times NRQ$ and $ncrq = s_1 \times s_2 (s_3 - NRQ)$.

Searching scheme in candidate subarrays for range key query

The candidate offset values to be searched in a candidate subarray are determined exactly according to a period. The ***candidate range of offsets*** for the period p_k of the range of column values *NRQ* of the known subscript is given by $crq = NRQ \times \prod_{j=1}^{k-1} s_j$ and the ***non candidate range of offsets*** is given by

$ncrq = (s_k - NRQ) \times \prod_{j=1}^{k-1} s_j$ (s_j is the length of dimension of dimension *j*). Note that each of these two kinds of range is consecutive in the sequence set of RDT. Assume that the number of nodes needed for storing $ncrq$ be $ncrq\_nd$ where $ncr\_nd = \left\lceil \frac{ncrq \times \rho}{kn} \right\rceil$ and ρ is density of records in the logical extendible array.

If $h_R < ncrq\_nd$ then (h_R is height of RDT) it will be faster to traverse from the root node of RDT to reach the next candidate node than to continue to search $ncrq\_nd$ nodes in the sequence set sequentially, otherwise all the nodes of the entire candidate offsets of the subarray are searched sequentially starting from the first candidate node. Fig. 5 shows the searching scheme.

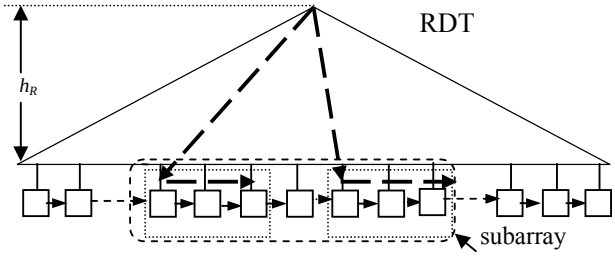


Fig. 5: The search Scheme

Considering User Specified Index for Retrieval: For a user specified indexed column of a relation, index is also constructed in our model by adding one more field (index field) in HT (See Section 3) for the corresponding dimension. For each key of CVT the index field in HT contains the set of candidate <*history, offset*> pairs of the subarrays that do not belong to the corresponding dimension of the CVT. In this case the searching is performed as follows:

For the specified column value the history value h_{max} of the principal subarray is determined by searching the corresponding CVT. Then the RDT is searched with key $< h_{max}, 0 >$. The sequential searching of RDT is continued until encountering any key whose history value is not equal to h_{max}; all of these records in the principal subarray are the candidates. To find the remaining candidate records the <*history, offset*> pairs that are stored in the index field of the corresponding HT are sufficient to be searched. Note that searching RDT is not necessary, because the key <*h,o*> holds the corresponding record itself; each column value can be simply decoded from <*h,o*>.

V. Cost Analysis

In this section the cost analysis of range key query for traditional relational table and the proposed model is developed and compared. The parameters are as follows. Some of these parameters are provided as input, while others are derived from the input parameters. All lengths or sizes are in bytes. Let *R* be a relational table to be implemented.

A. Parameters

NR : Total number of records in *R*
n : Number of columns in *R*
L_i : Number of distinct column values of the *i*-th column ($1 \le i \le n$)
dp_i : Duplicate factor of the *i*-th column,
namely $dp_i = NR / L_i$
ρ : Density of records in the logical extendible array,
namely $\rho = NR / \prod_{i=1}^{n} L_i$
cl : Length of a column value of *R*
P : Disk page size
LR : Length of a record in *R*
d : Order of a B^+ tree node
f : Average fan out from a node. $d \le f \le 2d$, for nodes except root node
kn : Average number of keys in a node, $kn = f - 1$
pl : Length of a pointer
klr : Length of a key (i.e. length of history value and offset value) of RDT
LN_R : Number of leaf nodes for RDT
i.e. $LN_R = \lceil NR / kn \rceil$
NLN_R : Number of non leaf nodes for RDT
The number of non leaf nodes NLN_R is calculated as follows:

$$NLN_R = \lceil LN_R / f \rceil + \lceil \lceil LN_R / f \rceil / f \rceil + \ldots + Y$$

where each term is successively divided by f until the last term Y is less than f. If the last term Y is not 1, 1 is added to the total (due to the root node). The number of terms in the expression for NLN_R represents the number of non leaf nodes that must be accessed while scanning RDT. It is denoted it by h_R. The height of the RDT is therefore, $h_R + 1$.
NRQ : Number of column values present in the specified range.
$SA(hv_1, hv_2, l)$: The set of history values whose dimension is k in the range of history values hv_1 to hv_2 ($hv_1 \le hv_2$), where $hv_1 = h_{min}$ and $hv_2 = h_{min} + (NRQ - 1) \times n$. (Because of assumption Section 5.1(3)(iii)).

(3) Assumptions:
To simplify the cost model, a number of assumptions are made.
(i) The length of column value of *R* is the same for all columns i.e. $LR = n \times cl$.

(ii) The duplicate factor dp_i of the *i*-th column $(1 \le i \le n)$ is the same for all *i*. Hence the number of distinct keys of the *i*-th column is also the same for all *i*.
(iii) The records in the logical extendible array are uniformly distributed.

The values of some parameters that are assumed are shown in Table 1.

Table 1: Assumed parameters

klr	*NR*	*d*	*P*	*pl*
4+8 (history value : 4 offset value : 8)	1000000	102	4096	8

In the following, the Traditional implementation of *R* is named as TI and the new implementation is named as as HI (History offset Implementation. We have developed the detail cost model for both the system here in this paper the analytical cost analysis is presented.

B. Cost Model

All the records in the subarrays of history values in the range $SA(hv_1, hv_2, l)$ are the candidates for the query. Hence all the nodes for each subarray in the range are accessed. See Fig. 3. The number of nodes accessed for a subarray of history values $i \in SA(hv_1, hv_2, l)$ is determined by

$$NA_1 = h_R + N_i$$

Average number of pages accessed for NA_1 is,

$$PA_1 = h_R + \lceil N_i / (P / XR) \rceil, \text{ if } P \ge XR$$
$$= h_R \times \left\lceil \frac{(XR/P)}{2} \right\rceil + \left\lceil N_i \times \frac{(XR/P)}{2} \right\rceil, \text{ if } P < XR$$

The subarrays of $SN(hv_1, l)$ are the remaining candidate subarrays in the extendible array. For each candidate subarray the query is performed by calculating the candidate range (*crq*) and non candidate range (*ncrq*) of the offsets as discussed in Section IV. The subscript of dimension *k* is known in the range *NRQ*.
The number of nodes accessed for a subarray of history value $i \in SN(hv_1, l)$ is determined by

$$K_1 \times (h_R + K)$$

Average number of pages to be accessed is determined by

$$K_1 \times (h_R + \lceil K / (P / XR) \rceil), \text{if } P \ge XR$$

$$K_1 \times (h_R \times \left\lceil \frac{(XR/P)}{2} \right\rceil + \left\lceil K \times \frac{(XR/P)}{2} \right\rceil), \text{ if } XR > P$$

The parameters K_1 and *K* are calculated as follows:

Step 1: $K_1 = \prod_{j=l+1}^{n-1} s_j$, $K_2 = NRQ \times \prod_{j=1}^{l-1} s_j$ and $K = \lceil (K_2 \times \rho) / kn \rceil$

The *K* nodes are organized consecutively in the sequence set of the RDT and they are searched sequentially.Where K_2 is the candidate range of offsets (*crq)* for the range and there are such K_1 candidate ranges exist.

Step2: If the number of nodes for the non candidate range of offsets (*ncrq*) is smaller than h_R then it is preferable to search the non candidate nodes sequentially rather than searching from root of RDT.
Hence if $\lceil (ncrq \times \rho) / kn \rceil < h_R$ then
$K_2 = K_1 \times crq + (K_1 - 1) \times ncrq$ and then $K_1 = 1$.

Cost for user specified index

Number of nodes accessed for RDT to find the records in *principal subarrays* of history values $i \in SA(hv_1, hv_2, l)$ is determined by $h_R + N_i$. This is because the records of the principal subarray are to be searched from RDT.

Number of pages accessed for RDT to find the records of principal subarrays of history values $i \in SA(hv_1, hv_2, l)$ is determined by

$$h_R + \lceil N_i / (P / XR) \rceil, \text{ if if } P \ge XR$$

$$h_R \times \left\lceil \frac{(XR/P)}{2} \right\rceil + \left\lceil N_i \times \frac{(XR/P)}{2} \right\rceil, \text{ if } P < XR$$

For range key query to calculate the number of non leaf pages accessed for CI the parameter of [16] $H(n_1, n_2, n_3)$ as $n_1 = dp \times NRQ$, $n_2 = NR$ and $n_3 = \lceil (NR \times LR) / P \rceil$ is used. This is because $dp \times NRQ$ are the candidate records for the query in CI.

Hence the number of pages accessed for CI if index is considered is determined by

$$= h_I + H(n_1, n_2, n_3), \text{ if } P \ge X$$
$$= h_I \times \left\lceil \frac{(X/P)}{2} \right\rceil + H(n_1, n_2, n_3), \text{ if } P < X$$

where h_I is the height of index;(i.e. a B^+ tree for index) and *X* is the length of a node for index.

B. Comparison between TI and HI

We computed the number of pages accessed to evaluate range key query both for TI and HI. We evaluated the

cost for cl= 16, $n = 6$ and dp between 300 and 9000 with dimension 3 being the known dimension.

Fig. 6 shows the pages accessed for TI and HI with varying duplicate factors.The parameter *NRQ* is assumed to be 25. The retrieval in HI is efficient than TI for large *dp*. The cost of HI for small *dp* but when *dp* increases the cost decreases. This is because when *dp* (i.e. ρ) is very small then HI accesses more nodes. In this case, the records of more than one candidate subarray exist in one node for very small ρ and the same node is accessed more than one time for different candidate subarrays. But when *dp* increases, the performance of HI becomes better and at a certain value of *dp* HI performs very well. For increasing *dp* the cost of HI decreases, this is because if ρ increases (i.e. *dp* increases) then the number of nodes for the non candidate range of offsets increases and when it becomes greater than h_R, HI scans h_R nodes instead of accessing the nodes of non candidate range of offsets. Hence for large *dp* HI has very good performance than TI.

If index is constructed then the cost of HI has improved performance than TI. This is because if *dp* increases then the number of records to be selected randomly from the relational table increases by the factor of *NRQ* which fetches $n \times cl$ bytes from the storage for each record in TI.

However in the range key retrieval, HI has impressive advantage comparing with TI. Throughout the process we used *cl* = 16. For larger *cl*, the impact in HI is negligible because *cl* has no effect in RDT but *cl* has immense effect on TI.

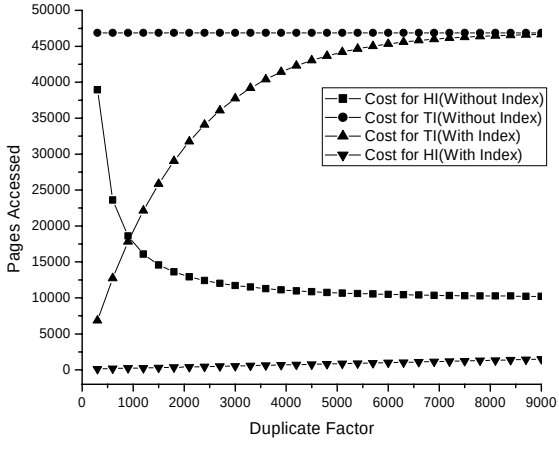


Fig. 6: Comparison between TI and HI for range key query

VI. Conclusion

In this paper, we introduced a new scheme for range key query for multidimensional databases. We developed a cost model and using these models, we computed and compared the range key query with the conventional of relational table. We found better performance than the conventional one. If the duplicate factor is not negligible then the proposed scheme has improved retrieval performance than the conventional relational tables. We believe, our scheme is fully general to apply for the multidimensional Online Analytical Processing or data warehouse systems for multidimensional analysis.

References

[1] K. E. Seamons and M. Winslett, "Physical schemas for large multidimensional arrays in scientific computing applications", Proc. of SSDBM, pp.218-227, 1994.

[2] S. Sarawagi, and M. Stonebraker, "Efficient organization of large multidimensional arrays", Proc. of ICDE, pp.328-336, 1994.

[3] Y. Zhao, P. M. Deshpande, and J. F. Naughton, "An array based algorithm for simultaneous multidimensional aggregates", ACM SIGMOD, pp.159 – 170, 1997.

[4] A. L. Rosenberg, "Allocating storage for extendible arrays", JACM, Vol. 21, pp.652-670, 1974.

[5] E. J. Otoo and T. H. Merrett, "A storage scheme for extendible arrays", Computing, Vol.31, pp.1-9, 1983.

[6] P. Vassiliadis, "Modeling multidimensional databases, Cubes and Cube Operations", Proc. of SSDBM, pp. 53-62, 1998.

[7] T. B. Pedersen and C. S. Jensen, "Multidimensional database technology", IEEE Computer, 34(12): 40-46, 2001.

[8] H. Kang and C. Chung, "Exploiting versions for On-line data warehouse maintenance in MOLAP servers", Proc. of VLDB, pp.742-753, 2002.

[9] J. Marcus, "Index structures for data warehouses", Springer, 2002.

[10]A. L. Rosenberg and L. J. Stockmeyer, "Hashing schemes for extendible arrays", JACM, Vol.24, pp.199-221, 1977.

[11] E. Bertino and W. Kim, "Indexing techniques for queries on nested objects", IEEE Transactions on Knowledge and Data Engineering, Vol 1. No. 2, June 1989.

[12] J. Nievergelt, H. Hinterberger, and K.C. Sevcik, "The grid file: An adaptable, symmetric multikey file structure", ACM Transactions on Database Systems, 9(1):38-71, 1984.

[13] K. Y. Whang and R. Krishnamurthy, "The multilevel grid file: A dynamic hierarchical multidimensional file structure", Proceedings of DASFAA, Tokyo, Japan, 1991.

[14] R. Orlandic and J. Lukaszuk ,"A class of region preserving space transformations for indexing high dimensional data", Journal of Computer Science 1 (1) (2005) 89-97.

[15] J.K. Lawder and P.J.H. King, "Querying multidimensional data indexed using the Hilbert space-filling curve", ACM SIGMOD Record, 30(1) (2001) 19-24.

[16] S. B. Yao, "Approximating block access in database organizations", Communications of ACM, vol.20, Number 4, pp. 260-261, April 1977.

[17] K. M. Azharul Hasan, M. Kuroda, N. Azuma, T. Tsuji, K. Higuchi, "An Extendible Array Based Implementation of Relational Tables for Multidimensional Databases", Proc. of DaWak, pp 233-242, 2005.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Linear Algorithm for Floorplan Compaction

Md. Wasi-ur-Rahman, Nusrat Sharmin Islam and Md. Saidur Rahman
Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology.
Email: saidurrahman@cse.buet.ac.bd

Abstract— A sequence pair is a pair of sequences of n rectangular blocks which represents the relative positions of each of these blocks in a floorplan. In this paper, we present an $O(n)$ time algorithm for constructing a compact floorplan from a given sequence pair representation. The best known previous algorithm takes $O(nloglogn)$ time.

I. Introduction

A *VLSI floorplan* is a topological arrangements of circuit modules called blocks on a two dimensional plane. The *area of a floorplan* is the area of the rectangle which encloses the blocks. To reduce the cost of a VLSI chip, it is desirable to obtain a floorplan with the smallest possible area. The process of taking a floorplan and attempting to move the blocks to make the area of the floorplan as small as possible is known as *floorplan compaction*. The floorplan in Fig. 1(b) is obtained from the floorplan in Fig. 1(a) by compaction. With the increasing number of components in a VLSI circuit, floorplan compaction has become a very crucial and important task in a VLSI chip design [4, 5, 6].

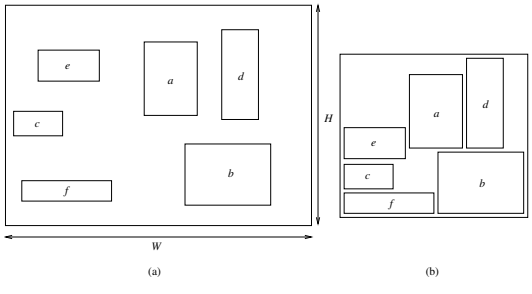


Fig. 1. (a) A floorplan and (b) a compact floorplan.

In this paper, we consider the problem of constructing a compact floorplan from a given floorplan represented by a "sequence pair" [3]. For a floorplan of n modules, a *sequence pair* is a pair of sequences, each of length n, which represents the topological placement of the modules in the floorplan. The sequence pair $(ecadfb, fcbead)$ represents the relative positions of the six blocks a, b, c, d, e and f in the floorplan in Fig. 1(a). The same sequence pair also represents the floorplan in Fig. 1(b), which is a compact floorplan of the floorplan in Fig. 1(a). We thus consider to find a compact floorplan for a given sequence pair.

Murata *et al.* [3] gave an $O(n^2)$ time algorithm for finding a compact floorplan for a given sequence pair. Tang *et al.* sped up the algorithm to $O(nlogn)$ in [7] and later Tang and Wong further improved to $O(nloglogn)$ in [8]. In this paper, we give an $O(n)$ time algorithm for finding a compact floorplan for a given sequence pair. The outline of our algorithm is as follows. From the sequence pair we first construct two graphs, namely a "horizontal constraint graph" and an "upward visibility graph". These two constraint graphs are weighted directed graphs where the vertices represent the blocks in a circuit, the edges represent the left or above relationship among the blocks and the weights of the edges are determined from the widths and heights of the blocks. To determine the position of each block in the compact floorplan, we use weighted topological sorting on the graphs.

The rest of the paper is organized as follows. Section 2 provides background information on the sequence pair representation. Section 3 provides some useful properties of sequence pair. Section 4 deals with the algorithm for constructing floorplan. Section 5 concludes the paper.

II. Preliminaries

In this section, we give some necessary definitions.

A *sequence pair* $\pi = (\Gamma+, \Gamma-)$ is a pair of two sequences each representing a list of *n* blocks in a floorplan. For example, $\pi = (abcd, bacd)$ is a sequence pair of module set { a,b,c,d } in the floorplan. Here, the first sequence $\Gamma+ = abcd$ and the second sequence $\Gamma- = bacd$.

In a sequence pair $\pi = (\Gamma+, \Gamma-)$, $\Gamma+$ denotes the first sequence which is obtained by drawing the "positive step line" for each block in the floorplan as illustrated in Fig. 2(a). Similarly, $\Gamma-$ denotes the second sequence which is obtained by drawing the "negative step line" for each block as illustrated in Fig. 2(b). A *positive (negative) step line* is a sequence of horizontal and vertical line segments drawn from the upper right (left) corner of a block and moving towards right (left)

without crossing any other block's boundary or any previously drawn positive (negative) step lines. The line moves upward only if by moving towards right (left) it would cross another block's boundary or any previously drawn positive (negative) step lines or the boundary of the floorplan. The line will end at the upper right (left) corner of the floorplan. The first sequence $\Gamma+$ is then obtained from the ordering of the positive step lines from left to right. Similarly, $\Gamma-$ is obtained from the ordering of the negative step lines from left to right. In Fig. 2, the first sequence is $\Gamma+ = ecadfb$ and the second sequence is $\Gamma- = fcbead$.

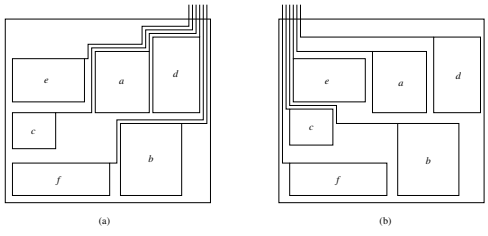


Fig. 2. (a) Positive step lines (b) negative step lines.

In a sequence pair $(\Gamma+, \Gamma-)$ modules x and x' can be related in four ways. These relationships are expressed by the following four disjoint sets [3].

$M^{aa}(x)$={$x'|x'$ is after x in both $\Gamma+$ and $\Gamma-$ },

$M^{bb}(x)$={$x'|x'$ is before x in both $\Gamma+$ and $\Gamma-$ },

$M^{ba}(x)$={$x'|x'$ is before x in $\Gamma+$ and after x in $\Gamma-$ },

$M^{ab}(x)$={$x'|x'$ is after x in $\Gamma+$ and before x in $\Gamma-$ }.

For any two modules x and x',

$x' \in M^{aa}(x) \Leftrightarrow x \in M^{bb}(x')$

$x' \in M^{ba}(x) \Leftrightarrow x \in M^{ab}(x')$

If $x \in M^{bb}(x')$ then x is left of x' in the floorplan. The claim also holds when the pair of words (M^{bb} and left of) is replaced by any of (M^{aa} and right of), (M^{ba} and above) and (M^{ab} and below) [3].

A *horizontal constraint graph*, for a set of n blocks in a floorplan is a directed acyclic graph G_H with n vertices where the vertices represent the blocks and each edge between two blocks represents the immediate left-right relationship between these blocks.

Let G be a digraph with n vertices and m edges. A *topological numbering* of G is an assignment of numbers to the vertices of G, such that, for every edge (u, v) of G, the number assigned to v is greater than the one assigned to u ($i.e.$ $number(v) > number(u)$). A *topological sorting* is a topological numbering of G, such that every vertex is assigned a distinct integer between 1 and n. If the edges of a digraph G have nonnegative weights associated with them, then a weighted topological numbering is a topological numbering of G, such that, for every edge (u, v) of G, the number assigned to v is greater than or equal to the number assigned to u plus the weight of (u, v) ($i.e.$ $number(v) \geq number(u) + weight(u, v)$).

III. Properties of a Sequence Pair

In this section, we will prove some properties of sequence pair which will be helpful for constructing horizontal constraint graph.

Lemma 1 *In a sequence pair, if a block y appears immediately after a block x in at least one of the two sequences and $y \in M^{aa}(x)$, then Block y lies immediately at the right of Block x in the floorplan.*

Proof. For the moment, let us assume that Block y appears immediately after Block x in at least one of the two sequences and $y \in M^{aa}(x)$, but y does not lie immediately at the right of Block x in the floorplan. Let us also assume that only Block z lies immediately at the right of Block x in the floorplan.

According to the position of the Block y and z in the floorplan, we have $three$ cases to consider.

Case 1*: Block y lies immediately at the right of Block z in the floorplan.*

Now, if we consider the positive and negative step lines of these blocks, then we can see that neither of the two sequences contain Block y immediately after Block x, which is a contradiction.

Case 2*: Block y lies after Block z in the floorplan but not immediately after z.*

In this case, we are assuming that Block y lies immediately after Block z_n, which is immediately at the right of Block z_{n-1} and so on. Block z_1 is at immediate right of Block x in the floorplan.

Here we can assume all the blocks between x and y $i.e.$ blocks $z_1, z_2, \cdots, z_n$ as a single block z and then this case reduces to Case 1.

Case 3*: Block y lies elsewhere in the floorplan.*

In this case, let us assume that Block y lies anywhere but at the right of Block z in the floorplan. So, if we consider it at the left of Block z, then it must be at the left or above or below of x, as Block z is the only block which is at the right of x. So, $y \notin M^{aa}(x)$, which is a contradiction. Then, if we consider it above or below of Block z, then it must also be above or below of Block x according to the definitions of the positive and negative step lines. So, $y \notin M^{aa}(x)$, which is again a contradiction.

So, Block y must lie immediately at the right of Block x in the floorplan. $\mathcal{Q.E.D.}$

Lemma 2 *In a sequence pair, if a block y appears immediately after a block x in the first sequence and a block z such that $z \neq y$, appears immediately after x in the second and $y \in M^{aa}(x)$ and $z \in M^{aa}(x)$, then there exists a sequence of blocks $y_1(= y), y_2, \cdots, y_l(= z)$, $l \geq 2$, where $y_1(= y), y_2, \cdots, y_l(= z) \in M^{aa}(x)$, such that all blocks y_i, $1 \leq i \leq l$ lie at the right of Block x and Block $y_1(= y)$ and Block $y_l(= z)$ are immediately at the right of Block x and blocks $y_2, \cdots, y_{l-1}$ maintain below-above relationship with y_1 and y_l in the floorplan.*

Proof. From Lemma 1, we can say that both y and z lie immediately at the right of x in the floorplan. Now, let us assume

that the sequence pair be $(\cdots xy \cdots z \cdots, \cdots xz \cdots y \cdots)$. As $y \in M^{ba}(z)$, y must be above z in the floorplan. Let us also assume that a sequence of blocks $y_2, \cdots, y_{l-1}$ be in between of Block y and z in both the sequences. Then, in the sequence pair, $(y_2, \cdots, y_{l-1}) \in M^{ab}(y)$ and $(y_2, \cdots, y_{l-1}) \in M^{ba}(z)$. So, they must lie below of y and above of z in the floorplan. As y and z both lie immediately at the right of x in the floorplan, $(y_2, \cdots, y_{l-1})$ must also be at the right of x in the floorplan.

$\mathcal{Q.E.D.}$

IV. Algorithm

In this section, we describe our algorithm to obtain the compact floorplan from a given sequence pair. We first construct a horizontal constraint graph from the sequence pair.

A. From Sequence Pair to Horizontal Constraint Graph

First, we place each block in a rectangular grid taking the first sequence position as the x-coordinate and the second sequence position as the y-coordinate. To construct the horizontal constraint graph, we need to iterate through the first sequence in the sequence pair and do the followings.

At each iteration i, we check which block is in position i in the first sequence. We assume it is x. Then we check which block is present in both sequences of the sequence pair just after this Block x. We assume these two blocks are x_1 and x_2. If x_1 and x_2 both maintain the left-right relationship with Block x *i.e.* $x_1 \in M^{aa}(x)$ and $x_2 \in M^{aa}(x)$, then we can say from Lemma 1 that both x_1 and x_2 are immediately at the right of Block x in the floorplan. So, we place an edge originating from x to each of x_1 and x_2. In this case, there may be more blocks present below Block x_1 and above Block x_2(or above Block x_1 and below Block x_2) that also lie at the right of Block x in the placement according to Lemma 2. At the moment, we just mark the Block x as a block for which all the blocks positioned just at the right of Block x in the floorplan may have not been identified.

Now, if one of the two blocks of x_1 and x_2 does not maintain the left-right relationship with Block x, *i.e.* suppose $x_1 \in M^{aa}(x)$ but $x_2 \notin M^{aa}(x)$ then the edge will be given only to the Block x_1 from x. And if both blocks do not maintain the left-right relationship with Block x *i.e.* $x_1 \notin M^{aa}(x)$ and $x_2 \notin M^{aa}(x)$ then no edges are added. And also if there is no block in the second sequence after the Block x then no edge is placed. We then remove Block x from the second sequence of the sequence pair. We will keep another sequence which we name as $revised\ second\ sequence$ and update it in each iteration. Here we keep the marked blocks (*i.e.* the blocks from which we are not sure whther all the necessary edges are given) and the blocks with no incoming edges yet. The blocks with no incoming edges are either the leftmost blocks in the placement which certainly will have no incoming edges or the blocks for which no edge is added till now. Now, by iterating the revised second sequence in the reverse order we can find the remaining edges necessary. The procedure is as follows.

Iterating in the reverse order, whenever we find a Block y with no incoming edges yet, we will check which marked Block x in this sequence just before it exists. If x and y maintains the left-right relationship in the original sequence pair, *i.e.* $y \in M^{aa}(x)$ then an edge is given from x to y. After this, Block y is removed from this sequence. If y does not maintain the left-right relationship with x, *i.e.* $y \notin M^{aa}(x)$ in the original sequence pair then y is not removed and we proceed to next iteration. Now, while iterating if we find a marked Block x, we will check which Blocks $y_1, y_2, \cdots, y_n$ are after it in this sequence. We will start checking from Block y_1 whether it satisfies $y_1 \in M^{aa}(x)$. If so, an edge is placed and y_1 is removed from the sequence. If we find a Block y_i with which x does not maintain the left-right relationship, *i.e.* $y_i \notin M^{aa}(x)$ the checking stops. If x has at least one incoming edge, then x is removed from the sequence. Otherwise, it is left and we proceed to next iteration without removing it. In this way, after iterating the full sequence all the necessary edges are given.

Fig. 3(a) shows all the edges added before iterating revised second sequence and Fig. 3(b) shows all the edges added in the constraint graph after iterating through the revised second sequence for a sequence pair *(abfgchijdklemno,aeondmlkcjibhgf)*.

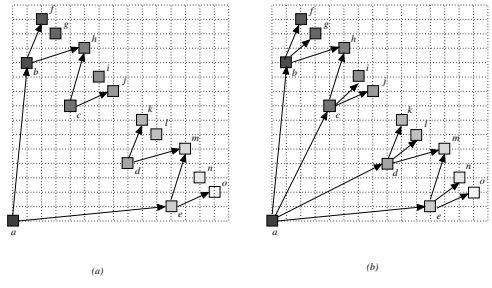


Fig. 3. (a) Partial horizontal constraint graph from sequence pair *(abfgchijdklemno,aeondmlkcjibhgf)* (b) full graph using revised second sequence $aendlcibg$.

Lemma 3 *If there exists a block y with no incoming edges in the revised second sequence and $y \in M^{aa}(x)$ where x is any block in the sequence pair, then there must be at least a marked block before y in the revised second sequence.*

Proof. Let us first assume that there is no marked block in the revised second sequence. Then according to the definition of the marked block, all the necessary edges have already been given. This can't be true as there exists a block y with no incoming edges and its not a leftmost block as $y \in M^{aa}(x)$. So y must have an incoming edge from one of the blocks that are before it in the revised second sequence.

So, there must be a marked block present in the sequence.

$\mathcal{Q.E.D.}$

Lemma 4 *If there exists a block y with no incoming edges in the revised second sequence and there exists a sequence of marked blocks $x_1, x_2, \cdots, x_l$ where $l \geq 1$, before y in the*

same sequence, then y will have an edge from x_i in G_H where $1 \leq i \leq l$ such that x_i is the nearest marked block before y in this sequence and y is an element of both the subsequences formed by placing the two blocks that are immediately after x_i in each sequence of the sequence pair at the beginning and the end.

Proof. Let us first assume that x_j and x_k, such that $j < k$, be two such marked blocks that y is an element of both the subsequences formed by placing the two blocks that are immediately after x_j and x_k in each sequence at the beginning and the end. Let us also assume that x_k is the nearest block to y. But we place an edge from x_j to y. Now, as $y \in M^{aa}(x_j)$ and $y \in M^{aa}(x_k)$, y is at the right of both x_j and x_k. Since, x_k is the nearest block to y, x_k must be after x_j in the second sequence. So, x_k must be at immediate left of y if $x_k \in M^{aa}(x_j)$ and so there should be an edge from x_k to y. Again, if $x_k \in M^{ba}(x_j)$ then x_k is above of x_j and both of them should have an edge to y.

Now, let us assume that x_j and x_k, such that $j < k$, be two such marked blocks that y is an element of the subsequences formed by placing the two blocks that are immediately after x_j, not x_k, in each sequence at the beginning and the end. Let us also assume that x_k is the nearest block to y in the revised second sequence and we place an edge from x_k to y. Now, as y is not an element of the two subsequences formed by the two immediate blocks of x_k, there should be an edge from x_k to y if and only if $y \in M^{aa}(x_k)$ holds. Let, the two immediate blocks of x_k in two sequences be a and b. Then $y \in M^{aa}(a)$ and $y \in M^{aa}(b)$ must also hold. From Lemma 2, we can say that x_k can only have an edge to those blocks that are in between a and b and common in both subsequences formed by placing a and b in the two sequences of the sequence pair at the beginning and the end. So, y should not have any edge from x_k. As the next nearest marked block in the revised second sequence is x_j, which holds y in both the subsequences formed by the two immediate blocks of x_j in each sequence, there should be an edge from x_j to y. $\mathcal{Q.E.D.}$

We are constructing the horizontal constraint graph by considering only the two blocks just after a particular block present in the sequence pair. The following lemma will show the necessary proof of that the blocks existing just at the right of a particular block in the floorplan, can be identified correctly from the sequence pair using this procedure.

Lemma 5 *The horizontal constraint graph G_H shows the correct left-right relationship of the sequence pair for each block.*

Proof. Let x and y be two blocks in the sequence pair and let $y \in M^{aa}(x)$. To prove the lemma, we have to consider two cases.

Case 1*: y is immediately after x in at least one of the two sequences.*
Let the sequence pair be $(\cdots xy \cdots, \cdots x \cdots y \cdots)$. According to Lemma 1 y should be exactly at the right of x in the floorplan and so the corresponding G_H must have an edge from x to y. At each iteration in our algorithm for constructing a horizontal constraint graph from a sequence pair we add two edges from the iterating block to those two blocks that are exactly after that block in each sequence. So, y will have an edge from x in G_H according to our algorithm. This is also true when only the second sequence or both the sequences contain y exactly after x.

Case 2*: y is not immediately after x in any of the two sequences.*
There are $three$ subcases to consider.

Subcase 1*:* Let a and b be two blocks in the sequence pair such that $a \in M^{aa}(x)$ and $b \in M^{aa}(x)$. a is immediately after x in first sequence and b is immediately after x in the second. Let us also assume that $y \in M^{ab}(a)$ and $y \in M^{ba}(b)$. So, the sequence pair is $(\cdots xa \cdots y \cdots b \cdots, \cdots xb \cdots y \cdots a \cdots)$. According to Lemma 4, if x be the nearest marked block in the revised second sequence for y, which is placed in between a and b, the two immediate blocks of x in the sequence pair, then there should be an edge from x to y and by iterating through the revised second sequence, this edge is added to G_H. Again, if x is not the nearest marked block in the revised second sequence for y, then let z be the nearest marked block of y. According to Lemma 4, edge from z to y is then added to G_H.

Subcase 2*:* Let a and b be two blocks in the sequence pair such that $a \in M^{bb}(y)$ and $b \in M^{bb}(y)$. a is immediately before y in first sequence and b is immediately before y in the second. Let us also assume that $x \in M^{ab}(a)$ and $x \in M^{ba}(b)$. So, the sequence pair is $(\cdots a \cdots x \cdots by \cdots, \cdots b \cdots x \cdots ay \cdots)$. According to *Case* 1, y will have edges from both a and b. Now, as Block a exists before x in the first sequence, it will be removed from the second sequence before the iteration for x, according to our algorithm. If any block exists between x and y at the time of iteration of x, then that particular block must be in the set $M^{aa}(x)$. So, it should have an edge from x according to Lemma 1 and y should have an edge from x according to Lemma 2 and the edges are added rightly in our algorithm by iterating through the sequence pair and the revised second sequence respectively. Again, if there is no such block, then y will be exactly after x in the second sequence at that time, and so an edge from x to y is added at that iteration.

Subcase 3*:* Let a and b be two blocks in the sequence pair such that $a \in M^{bb}(y)$ and $b \in M^{bb}(y)$. a is immediately before y in first sequence and b is immediately before y in the second. Let us also assume that $x \in M^{bb}(a)$ and $x \in M^{bb}(b)$. So, the sequence pair is $(\cdots xa \cdots by \cdots, \cdots xb \cdots ay \cdots)$. According to Lemma 2, y should not have any edge from x and the Lemma 4 proves that y will not have any edge from x while iterating the revised second sequence.

$\mathcal{Q.E.D.}$

Lemma 6 *Iterating through the revised second sequence to obtain the necessary edges remaining in the graph G_H takes linear time.*

Proof. Iterating from the right of the revised second se-

quence, we find those edges which are not added yet in the graph G_H. Now, at each iteration, if the block is one of those blocks with no incoming edges, then we just have one checking whether the block maintains the left-right relationship with the first marked block present in this sequence before it. This checking clearly takes constant time as our data structure provides this information.

Now, if a block is marked only during an iteration, then we have to check how many blocks after it in this sequence will have edges from it. Starting from the block just after it we will check each block until we get a block for which there will be no edge. After assigning an edge to a block that block will be removed from the sequence and so no other marked block will check for it. So, each marked block will check for other blocks after it in the revised second sequence. There will be only one extra checking for each marked block for which there will be no edge in the graph. So, the total number of extra checkings is at most n where n is the total number of blocks in the revised second sequence. And the number of checkings for which there will be an edge is also at most n as no blocks can have incoming edges from more than *one* blocks in this revised second sequence. So, the running time of this procedure is linear to the number of blocks n. $\mathcal{Q.E.D.}$

Theorem 7 *The running time of constructing horizontal constraint graph G_H is linear to the number of blocks in the sequence pair.*

Proof. Iterating through the first sequence in the sequence pair takes linear time and from Lemma 6, iterating revised second sequence takes linear time, too. So, the construction process is linear to the number of blocks n present in the sequence pair. $\mathcal{Q.E.D.}$

B. From Horizontal Constraint Graph to Ordered Tree

In this section, we give a linear-time algorithm for constructing an ordered tree from the horizontal constraint graph G_H.

Constructing an ordered tree T from the horizontal constraint graph G_H involves three steps: 1) Each edge of G_H is assigned a weight w, equal to the width of the block from which it originates. 2) Initially, each node of the horizontal constraint graph is assigned a cost equal to $zero$. 3) Weighted topological sorting is applied to G_H to obtain the x- coordinates of the blocks and construct an ordered tree T.

For the sequence pair ($ecadfb$,$fcbead$), the widths and heights of the modules are assumed and G_H with all edge weights assigned is shown in Fig. 4(a).

At each iteration of weighted topological sorting, a node u with $in\text{-}degree = 0$ is selected. The cost of all its neighbours are updated. If node v is a neighbour of node u, then the cost of v is $cost(v) = max(cost(v), cost(u) + w(u, v))$, where $w(u, v) =$ weight of edge (u, v). Then node u and all edges originating from it are deleted.

The final cost of each node is equal to the cost of the longest path from the source to that node and this is the x-coordinate of the corresponding module in the floorplan. Each node v now has unique parent u and each node lies along a unique path originating from the source. In this way, an ordered rooted tree T is constructed as shown in Fig. 4(b).

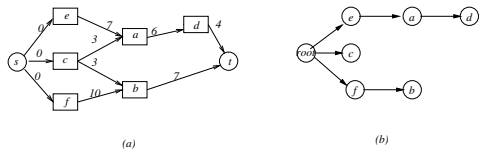


Fig. 4. (a) Horizontal constraint graph with all edge weights assigned (b) Ordered Tree after applying weighted topological sorting.

Lemma 8 *Weighted topological sorting computes the minimum possible x-coordinates for each node in the horizontal constraint graph G_H.*

Proof. Let v be a node in the horizontal constraint graph, which has l parents u_1, u_2,$\cdots$,u_i,$\cdots$,u_j,$\cdots$,u_l. At the end of weighted topological sorting the cost of v is $cost(v) = cost(u_i) + w(u_i, v)$ and u_i is the parent of v. This cost corresponds to the x-coordinate x_v of the module v in the placement. Let $x_v^{'}$ be the minimum possible x-coordinate of the module v such that $x_v^{'} \leq x_v$. v is adjacent to u_i when its x-coordinate is x_v, while it is adjacent to u_j when its x-coordinate is $x_v^{'}$. Let us assume that the starting x-coordinate of u_i is x_{u_i} and that of u_j is x_{u_j}.

As we assumed, $x_v^{'} \leq x_v$ which implies that $x_v^{'} \leq cost(u_i) + w(u_i, v) \Rightarrow x_v^{'} \leq x_{u_i} + w(u_i, v)$ According to Lemma 5, blocks u_1, u_2,$\cdots$,u_i,$\cdots$,u_j,$\cdots$,u_l all must be at immediate left of v in the floorplan. So, if we choose $x_v^{'}$ as the x-coordinate of node v, then it will overlap with the Block u_i in the floorplan, which will violate floorplan constraints. So, $x_v^{'}$ must be at least equal to x_v. So, weighted topological sorting computes the minimum possible x-coordinates for each node. $\mathcal{Q.E.D.}$

Lemma 9 *The ordered tree T can be constructed from the horizontal constraint graph G_H in linear time.*

Proof. The construction of the ordered tree T from the horizontal constraint graph G_H involves iterations through all the n nodes of G_H. At each iteration, a node with in-degree=0 is selected and the cost of all its adjacent nodes are updated. The total time required to update the cost of the neigbouring nodes of the n nodes in G_H is $O(m)$, where $m =$ number of edges in G_H. Thus the running time of the algorithm is $O(n + m)$. Since the graph G_H is planar $m = O(n)$, the running time is $O(n)$. $\mathcal{Q.E.D.}$

C. From Ordered Tree to Upward Visibility Graph

Constructing an upward visibility graph from the ordered tree is done in the same way as in [2]. The running time of this procedure is also proved to be linear ([2]).

D. From Upward Visibility Graph to Floorplan

To obtain the y-coordinates of the rectangular modules, we apply weighted topological sorting to the upward visibility graph, which requires linear time. For this, each edge of the upward visibility graph is assigned a weight equal to the height of the block from which it originates. Weighted topological sorting can now be applied in the same way as to the horizontal constraint graph.

At the end of weighted topological sorting, the cost of each node represents the y-coordinate of the corresponding module in the floorplan. Since we have both the x- and y- coordinates, we can construct the floorplan, which is compact for the given sequence pair representation. The floorplan is shown in Fig. 5.

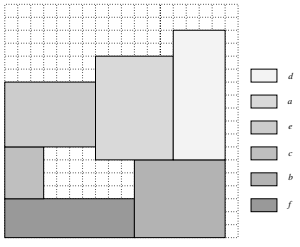


Fig. 5. Compact floorplan for the sequence pair $(ecadfb, fcbead)$

Lemma 10 *Weighted topological sorting computes the minimum possible y-coordinates for each node in the upward visibility graph G_v.*

Proof. This proof is similar to the proof of Lemma 8 replacing x-coordinates by the y-coordinates. $\mathcal{Q.E.D.}$

Theorem 11 *For the given sequence pair, the constructed floorplan is the most compact.*

Proof. Let CF be the constructed floorplan by using our algorithm for a sequence pair SP. Suppose there exists a more compact floorplan CF' for the same sequence pair. Let the width, height and area of CF be W, H and A respectively and those of CF' be W', H' and A' respectively. Now, there are $three$ cases to consider:

Case 1*: $W' < W$ and $H' = H$*

In this case the x-coordinate of at least one block in CF' must be less than the x-coordinate of that block in CF. Let such a block be a. From Lemma 8, weighted topological sorting computes the minimum possible x-coordinates for each block. So, Block a in CF', having an x-coordinate less than that in CF, must overlap with its immediate left block in the floorplan which violates the floorplan constraints. So, W' must not be less than W when $H' = H$.

Case 2*: $H' < H$ and $W' = W$*

This case is similar to *Case* 1 if we interchange heights with widths.

Case 3*: $A' < A$ and $H' \neq H$, $W' \neq W$*

As from Lemma 5, the horizontal constraint graph G_H consists of only those edges that show the immediate left-right relationship for each block. So, for any sequence pair SP, G_H must be unique. Similarly, for any ordered tree, if the widths of the blocks are same, the upward visibility graph G_{uv} must be unique as it consists of only those edges, that show the upward visibility for each block in the floorplan. Now, as $A' < A$, either H' or W' or both must be less than that in CF. According to Lemma 8 and Lemma 10, weighted topological sorting computes the minimum possible x- and y-coordinates for each block. So if the heights and widths of the blocks are fixed, the computed x- and y- coordinates for each block should be unique. Therefore, to obtain a more compact floorplan CF', by changing the height and width of CF, the sequence pair must be violated.

So, for the given sequence pair, CF is the most compact floorplan.

$\mathcal{Q.E.D.}$

V. Conclusion

In this paper, we have presented an algorithm for obtaining a compact floorplan for a given sequence pair representation of rectangular blocks of fixed heights and widths. Our algorithm has two significant improvements: 1) the runtime of our algorithm is linear, and 2) the constructed floorplan is the most compact for the given sequence pair. We have also proved some useful properties of the sequence pair, which are later used in finding the compact floorplan in linear time.

References

[1] Pei. Ning Guo, Toshihiko Takahashi, Chung-Kuan Cheng and Takeshi Yoshimura,*Floorplanning using a tree representation*, IEEE Trans. Computer-Aided Design of integrated circuits and systems, vol. 20, No. 2, 2001.

[2] Masud Hasan, Md. Saidur Rahman and Takao Nishizeki,*A linear algorithm for compact box-drawings of trees*, Proc. of CCCG'02, pp.154-157,2002.

[3] H. Murata, K. Fujiyoshi, S. Nakatake and Y. Kajitani, *VLSI module placement based on rectangle-packing by the sequence pair*, IEEE Trans. Computer-Aided Design, vol. 15, pp. 1518-1524, 1996.

[4] T. Nishizeki, Md. Saidur Rahman, *Planar Graph Drawing*, World Scientific Publishing Co. Pte. Ltd., Singapore, pp. 9-15, 2004.

[5] N. Sherwani, *Algorithms for VLSI Physical Design Automation*, Kluwer Academic Publishers, USA, pp. 423-450, 1995.

[6] S. M. Sait, H. Youssef *VLSI Physical Design Automation*, World Scientific Publishing Co. Pte. Ltd., Singapore, pp. 100-220, 1999.

[7] X. Tang, R. Tian and D. F. Wong, *Fast evaluation of sequence pair in block placement by longest common subsequence computation*, Proc. Design Automation Test Eur., pp. 106-111, 2000.

[8] X. Tang and D. F. Wong, *FAST-SP: A fast algorithm for block placement based sequence pair*, Proc. ASPDAC, pp. 521-526, 2001.

[9] Evangeline F. Y. Young, Chris C. N. Chu and M. L. Ho,*Placement constraint in floorplan design*, IEEE, Transactions on VLSI Systems, Vol.12, No.7, 2004.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A Behavioral Model of Writing

M. Naghibolhosseini and F. Bahrami

Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran.

m.naghib@ece.ut.ac.ir, fbahrami@ut.ac.ir

Abstract - In this paper we propose a new model to generate handwriting based on behavioral patterns we believe is to be found in humans when imitating a written character. The proposed algorithm has a hierarchical structure. It is consisted of two main levels. At the first level the graphical features of the written letter to be imitated are extracted. These features are the directions of movement for each stroke. To extract the strokes that shape a letter, zero crossings of vertical and horizontal spatial velocity profiles are determined. At the second part a given letter is regenerated by using the extracted directions. At this level, the letter is linearly divided into several subdivisions. After that the excursion length (step) and the slope of line-segments producing the subdivisions are estimated. In each trial of learning, trajectory points are chosen with the estimated step and a random slope. The final trajectory is produced by successive arrangement of the strokes. In each trial the slope and step of the target point has a distance less than a specified threshold from the actual path is stored in memory and others are generated again. This process will continue during different trials until the trajectory can be generated using only memory. The resulted trajectories for different letters written by different subjects are qualitatively and quantitatively similar to the actual trajectories generated by human.

I. Introduction

Handwriting stands among the most complex tasks performed by literate human beings [1]. Handwriting is a main point for several motor control problems, such as sequencing of stroke order, decomposition of movements into basic segments, characterization of mental movement coordinate systems, and the role of sensory feedback for motor planning [2]. Examination of these subjects can be done in non-invasive, inexpensive, and easily executed experiments on human subjects via the study of handwriting [2]. This study forms a very broad area that lets researches with various backgrounds to collaborate and interact with different goals [1, 3].

The study of handwriting is based on its modeling. Several modeling techniques are proposed to simulate the handwriting process [4]. There are two general methodologies of handwriting modeling. The first methodology takes into consideration computational models which are aimed to reproduce some features of human handwriting movements such as trajectory. Hollerbach oscillatory model of writing is one of these methodologies [3, 5].

The second methodology of handwriting modeling focuses on psychologically illustrative models. These methodologies consider cognitive aspect of writing such as learning and memory of the movements, which are often omitted from the first methodology, and not an aspect of trajectory formation [3, 6, 7].

Trajectory Formation is one of the basic functions of the neuromotor controller. A group of studies on trajectory formation is directly related to mechanical properties and the geometry of the muscles [6].

In this study we propose a new method that is the combination of both methodologies. We propose an algorithm to plan and generate the trajectory and also learn the trajectory of the handwriting independently of the actual joint and muscle patterns. This model relies on Arabic handwriting data. The model has a target point estimation algorithm that estimates target points and save them according to the distance that they have with the actual trajectory. This model not only generate the trajectory of a letter almost the same as human trajectory but also save the information that is needed to write the letter in working memory. So when the model learns to write a letter it can use only memory to generate the letter trajectories.

The overall approach is based upon the hypothesis that complex human movements are made up of, and can be segmented into, basic and simple parts [1, 2, 6, 8]. In other words, due to the properties of the neuromuscular system involved in a rapid writing task, there is a class of simple movements, called strokes, that are preferentially produced by such a system, once it is well-trained. More complex movements are thus generated by the vectorial addition of various strokes belonging to such a class [1].

In the case of handwriting, it is important that the model to be able to generate a script in such a way that can be recognized by almost everyone, not to try to regenerate an 'ideal' script [9].

This paper aims to show how to study and to understand the basic properties of pen-tip trajectories as produced by human subjects that perform handwriting.

This paper is organized as follows. The database is introduced in Section II. Then in section III the method that is proposed for planning and learning to write is described. The results of the proposed algorithm are then demonstrated in Section IV. The final conclusions are discussed in Section V.

II. Data Set

Handwriting data is collected using a Wacom 9 · 12 Intuos digital writing tablet with sampling frequency of 206 Hz. The data was collected from four subjects. The subjects were asked to write all the Arabic letters ten times. During every trial horizontal (X) and vertical (Y) coordinations of the pen-tip were recorded.

III. Methods

In this paper we use the hypothesis that complex human movements such as handwriting can be segmented in to some basic and simple units, called strokes. First of all we find these basic units. In Arabic letters we have both up-down and right-left strokes. So in order to find the strokes we should find vertical and horizontal zero crossings of spatial velocity. We name these estimated points stopping points. Indeed stopping points are the points in which velocity in each direction changes its sign, and also velocity became zero or near zero.

After finding stopping points we should define the movement directions during each stroke. The direction of the movement changes at each stopping point. We consider four different directions: positive and negative x directions and also positive and negative y directions.

In next stage we find the slopes between every two adjacent points of the letter. Then we divide the letter into some parts in which the changes in slope is greater than a specified value. We define the slope of each part as a mean value of the slopes of every two adjacent points in that part. Then we find the length of each part and we assume these lengths as the model's steps for choosing target points in order to form the trajectory of the movement.

After the mentioned steps we run the main algorithm of learning to write. In order to find the target position (TPx, TPy) we should have the present position (PPx, PPy) and the "step" and angle of the movement toward the target. So we can calculate target position using equations (1) and (2):

$$TPx=PPx+\cos(\text{random angle})*step \quad (1)$$

$$TPy=PPy+\sin(\text{random angle})*step \quad (2)$$

We start from the first point of the letter. We choose the variable "step" the same as the "step" that we calculated in the previous stage. Then we choose a random angle according to the direction of the first stroke.

$$initial\ random\ angle = \begin{cases} 0-90 & if\ direction = +x,+y \\ 90-180 & if\ direction = -x,+y \\ 180-270 & if\ direction = -x,-y \\ 270-360 & if\ direction = +x,-y \end{cases} \quad (3)$$

Now we have calculated target position of the movement. To estimate the next point we compare the slope of the previous stage with the actual slope of the next stage and calculate the error percentage and error sign. Error sign shows the estimated slope is lower or higher than the actual one. We randomly choose the next angle according to the following rule:

$$Random\ angle = previous\ random\ angle \pm \begin{cases} 0-18, & if\ 0 \le error\ percentage < 20 \\ 18-36, & if\ 20 \le error\ percentage < 40 \\ 36-54, & if\ 40 \le error\ percentage < 60 \\ 54-72, & if\ 60 \le error\ percentage < 80 \\ 72-90, & if\ 80 \le error\ percentage \le 100 \end{cases} \quad (4)$$

The sign is chose according to the error sign. In each target selection at the start of each stroke we check the movement direction and preserve it up to the end of the stroke. With the composition of the produced strokes the complete trajectory will be generated.

We use the above rule and equation (1), (2) to estimate target points. After choosing each target position we check the distance between the chosen point and the nearest point in the original trajectory, if the distance is less than the specified value that depends on the accuracy that we want, we save the step and the slope of the chosen point in memory. The first trial will be completed when the comparison with the last slope is done and the last target position is chosen.

During each target selection if the slope and step that is needed to choose a target already exist in memory, these values will be used to select the target position so the movement is memory based. But if the slope and step values don't exist in memory we should choose the random angle again, but this time we choose an angle between the angle that was chosen in previous trial and the actual angle. After every trial the trajectory becomes more similar to the original trajectory and the number of the saved points in memory increase. We repeat this algorithm until all points can be chosen using only memory and so the trajectory formation become completely memory based.

IV. Results

We studied the performance of the proposed model for all the letters that can be written without taking the pen-tip off the paper, and as a sample from our simulation results we show the results for 4 letters..'Lam' that is shown in fig.1 is almost the simplest letter with comparison to other 3 chosen letters. 'Ein', 'Sin' are among the most complicated Arabic letters. Letter 'Mim' has a closed circular shape.

The best correlations between the handwriting generated by the model and the recorded data, over X and Y directions for these 4 letters are reported in table 1.

Table 1 correlations between human and model trajectory in x and y directions for 4 chosen letters shown in figures 1-4.

Letter/Subject	X position	Y position
Lam/1	0.9886	0.9994
Sin/2	0.9830	0.9676
Ein/3	0.9551	0.9623
Mim/4	0.9759	0.9699

The correlations is calculated using the following equation [2, 8, 10]:

$$c(a,b) = \max_{0 \le r \le R} \frac{\sum_{i=0}^{n-r} (a_i - \bar{a})(b_{i+r} - \bar{b})}{(n-r)\sqrt{\frac{1}{n}\sum_{i=0}^{n} (a_i - \bar{a})^2}\cdot\sqrt{\frac{1}{n}\sum_{i=0}^{n} (b_i - \bar{b})^2}} \quad (5)$$

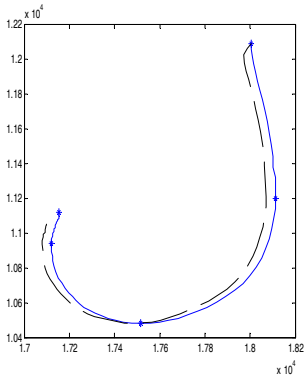


Fig. 1 Letter Lam from subject 1, recorded data (solid) and trajectories generated by the model (dash) are shown. The stars on the trajectory show the stop points which are the end of a stroke and the beginning of a new stroke.

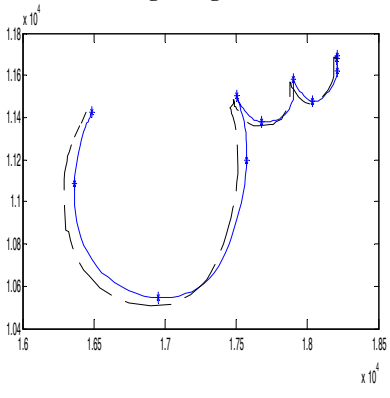


Fig. 2 Letter Sin from subject 2, recorded data (solid) and model (dash) trajectories are shown.

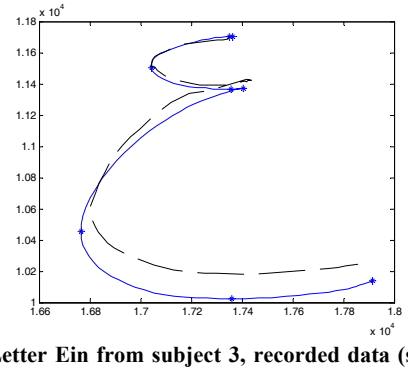


Fig. 3 Letter Ein from subject 3, recorded data (solid) and model (dash) trajectories are shown.

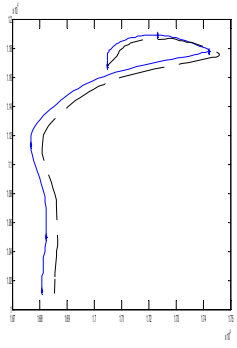


Fig. 4 Letter Mim from subject 4, human (solid) and model (dash) trajectory is shown.

We also examine the performance of the proposed model on other Arabic letters collected from the four subjects. The results show that performance of the proposed model was different among various letters, with a maximum total correlation of 0.99 (mean of x and y direction correlations) for letter 'Lam' and a minimum of 0.78 for letter 'Ein'. The reason for this observation returns to the complexity of the letter, the more complex the letter is the more the quality of performance is decreased. The model shows small variability among different subjects. Also in conditions that the generated trajectory has lower correlation with human handwriting, still the shape of the generated letter is similar to the recorded one.

V. Conclusion

In this paper we proposed a behavioral model of writing. This model can plan and learn human handwriting trajectories. This model breaks the trajectory into strokes and with the composition of these strokes the final trajectory will be generated. To form each stroke we choose some target points. These target points are chosen with a new algorithm based on human writing behavior. The strokes are generated successively. Each target point is checked to see whether it is appropriately chosen or not and according to this some information about that target point will be saved in memory. As learning proceeds over multiple trials, the memorized information becomes more complete. So after the last trial the trajectory will be generated using only memory.

This model can also generate letters with different shapes. In future work we want to examine this model on the generation of letters with different scales.

Different models are proposed for handwriting. One of these models is the Edelman and Flash (1987) minimum snap model. In this work they presented a model of trajectory formation based on dynamic minimization of jerk [8,10]. Their approach was a computational one. Another model that is proposed in modeling handwriting was AVITEWRITE which was developed by Grossberg and Paine (2000). AVITEWRITE was a physiological inspired model. The model that we have proposed is essentially inspired by human behavior during writing.

Our model although is simple, yields good correlation with recorded human handwriting data. The maximum correlation that we got for this model is 0.99 for letter "Lam" which is comparable with the results reported by

Edelman and Flash (1987) which was 1 in the best case and in the worst case the minimum correlation that our model showed is 0.78 for letter "Lam" which is better than AVITEWRITE (0.76).

References

[1] R. Plamondon and W. Guerfali, "The generation of handwriting with delta-lognormal synergies," Biol. Cybern., vol. 78, pp. 119–132, 1998.

[2] S, Grossberg and R.W. Paine, "A neural model of cortico-cerebellar interactions during attentive imitation and predictive learning of sequential handwriting movements," Neural Networks, vol. 13, pp. 999–1046, 2000.

[3] H. Bezine, A. M. Alimi and N. Sherkat, "Generation and Analysis of handwriting script with the beta-eliptic model," I.J. of Simulation, vol.8, 2004.

[4] W. Abend, E. Bizzi, and P. Morasso, "Human arm trajectory formation," Brain, vol. 105, pp. 331-348, 1982.

[5] J.M. Hollerbach, "An oscillation theory of handwriting," Biol. Cybern., vol. 39, pp. 139-156, 1981.

[6] P. Morasso and F. A. Mussa Ivaldi, "Trajectory formation and handwriting: a computational model," Biol. Cybern., vol. 45, pp. 131-142, 1982.

[7] F. Bouslama, M. Benrejeb, "Exploring the human handwriting process," Int. J. Appl. Math. Comput. Sci., vol. 10, n. 4, pp. 877-904, 2000.

[8] R.W. Paine, S. Grossberg, A.W.A. Van Gemmert, "A quantitative evaluation of the AVITEWRITE model of handwriting learning," Human Movement Science, vol. 23, pp. 837–860, 2004.

[9] Y. Wada, M. Kawato, "A via-point time optimization algorithm for complex sequential trajectory formation," Neural Networks, vol. 17, pp. 353–364, 2004.

[10] S. Edelman and T. Flash, "A Model of Handwriting," Biol. Cybern.,Vol. 57, pp. 25-36, 1987.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

A New Approach of Dynamic Encoded Bitmap Indexing Technique based on Query History

Md. Golam Rabilul Alam, Mohammed Yasir Arafat, Mohammed Kamal Uddin Iftekhar

Department of Computer Science & Engineering
International Islamic University Chittagong Bangladesh.
E-mail: gra9710@yahoo.com, arafat_cst@yahoo.com, kamal_iftekhar@yahoo.com .

Abstract-In this paper we have proposed a new approach of Encoded Bitmap Indexing [3] that makes the Encoded Bitmap Indexing Technique well defined for most of the selection queries by using association rule of data mining. Furthermore, this approach makes Encoded Bitmap Indexing (EBI) up to date after specific duration. Actually, this thesis incorporates an idea of using dynamic n-items pattern selection for Encoded Bitmap Indexing technique lookup table to be well defined. We have used 2-items pattern selection of query predicate, 3-items pattern selection of query predicate up to n-items pattern selection of query predicate to select the most frequent pattern for the lookup table dynamically, consisting of database table attribute distinct values of character string and numbers (integer). The main feature of our work is that we have used association rule of data mining to determine the most frequent n-items pattern. It is the first time we select the n-items most frequent pattern dynamically for each attribute of the table which involved in indexing the table data.

Keywords: Encoded Bitmap Indexing, Data Mining, Lookup Table, Query Predicate Pattern, Most Frequent N-items Pattern.

I. Introduction

After the evaluation and development of Database System, indexing [9] is one of the important issues in the computer science for finding the required data from the database system efficiently. As the advancement and increasing Database application areas, the need of advance indexing technique is increasing day by day. Again recently another type of database system evolves i.e. the data warehouse system [4][5], which requires more advances indexing technique. That's why more and more research is going for finding and development of the advance indexing techniques.

Therefore, the Encoded Bitmap Indexing [3] technique has evolved that speed up query processing up to 20% to 40% are typical, and saves the indexing space at a logarithmic scale. Encoded Bitmap Indexing uses lookup table to store the encoded value of the keys and the ***bitmap vectors*** that stores the actual indexes of the database table.

In our proposed indexing technique, we first scan and read the query history file for the attribute of the *dimension* table of which lookup table to be remapped to select the actual pattern (which is most frequent) dynamically and have used this pattern to remap the lookup table. We have used association rule of data mining for selecting the n-items pattern. We have evaluated the support and confidence for 2-items ruleset, 3-items ruleset up to n-items ruleset to find the actual pattern. When the confidence of any n-item ruleset is equal to or greater than the minimum threshold value then we have taken that n-item ruleset in the candidate ruleset. In this way, we find the most frequent n-item ruleset and thus find the actual pattern for the lookup table key values arrangement. Then this pattern is used to arrange the lookup table.

II. Literature Review

A. Encoded Bitmap Indexing

Given a table $\tau\{t_1,\cdots\cdots\cdots\cdots t_n\}$, where tj is a tuple of $\tau\{j=1\cdots n)$, let A be an attribute of τ , denoted by τ .A, and the domain of A be $\{a_1,\cdots\cdots a_m\}$. Then, a **encoded bitmap index $\mathbf{B^A}$** on τ . A, is a set of bitmap vectors $\{B_{k-1},\cdots\cdots,B_0\}$, a one-to-one mapping ($\mathbf{M^A}$: A -> $\{<b_{k-1}\cdots b_0>|\ \ b_{i\in}\{0,1\},\ \ i=0,\cdots k-1\})$) and a set of *retrieval Boolean functions* ($\{f_{a1},\cdots\cdots,f_{an}\}$), where $k=\lceil\log_2 m\rceil$. The bitmap vectors are defined as follows: $\forall B_i$ (i=0…k-1), tj (j=1…n) => B_i [j] = 1 if $M^A(\tau$ .A) [i] = 1, else $B_i[j] = 0$, where $B_i[j]$ denotes the j-th bit of B_i and $M^A(\tau$ .A) [i] the i-th bit (from LSB to MSB) $M^A(\tau$ .A). In addition, $\forall\alpha\ \in\{a_1,\cdots,a_m\}$, the retrieval function for α , f_α is a k-variable min-term (i.e., a fundamental conjunction) $x_{k-1},\ldots x_0$, where $x_i = B_i$, if $M^A(\tau$ .A) [i] = 1, otherwise $x_i = B_i$ (i=0,…k-1). Let us suppose an attribute STATE (Figure 1), with 3 possible values: {NY, MA, CA}. In the simple bitmap case, we would need 3 vectors. However, in the **encoded bitmap index** case, $\lceil\log_2 3\rceil$ = 2 bitmap vectors would suffice.

Table τ

NY	…
MA	…
CA	
MA	
NY	

B_{NY}	B_{MA}	B_{CA}
1	0	0
0	1	0
0	0	1
0	1	0
1	0	0

B_1	B_0
0	0
0	1
1	0
0	1
0	0

Mapping Table

NY	00
MA	01
CA	10

Figure 1. Encoded Bitmap indexing of typical STATE attribute

A retrieval function must be defined. Say A is a value encoded as $b_{k-1},\cdots\cdots,b_0$. The *retrieval function* f_A is defined as

$$f_A = x_{k-1}\cdots\cdots x_0$$

Where x_i is B_i, if $B_i = 1$, and the negation of B_i, call it B_i'' otherwise. In our example, $f_{CA} = B_1B_0'$. We will represent the OR of two retrieval functions as $f_A + f_B$, and the AND, as $f_A\, f_B$.

B. Data Mining

Data mining [7][8] is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. Data mining consists of finding interesting patterns in large datasets in order to guide decisions about future activities. The first and simplest analytical step in data mining is to **describe** the data — summarize its statistical attributes (such as means and standard deviations), visually review it using charts and graphs, and look for potentially meaningful links among variables (such as values that often occur together).

C. Association Rule

Association rule mining searches for interesting relationship among items in a given data set. Suppose we want to learn about the relationship between item ***pen*** ***and ink*** in a file. This information that ***pen*** tends to stay closer to ***ink*** in a file is represented in association rule below: ***{pen}*** $\Rightarrow$ ***{ink}***

More generally an association rule has the form *LHS* $\Rightarrow$ *RHS*, where both *LHS* and *RHS* are the sets of items. The interpretation of such rule is that if every item in *LHS* is occurred in a string, then it is likely that the items in *RHS* are occurred as well.
There are two important measures of association rule:
Support: The support for a set of items is the percentage of occurrences that contain all desired items together. For example, consider the rule ***{pen}*** => ***{ink}***. The support of this rule is the frequency of the itemset ***{pen, ink}*** *or* ***pen ink***.
Confidence: The confidence for a rule *LHS* => *RHS* is the percentage of such occurrences that also contain all items in *RHS*. More precisely, the confidence of a rule is an indication of the strength of the rule. As an example (Table 1), consider again the rule ***{pen}*** => ***{ink}***. The confidence of this rule is *sup(**penUink**) / sup(**pen**)*.
Rules that satisfy both minimum threshold value of support (*min_sup*) and minimum threshold value of confidence (*min_conf*) are called strong.

Table 1 Transaction Table

Transaction	Item
111	pen, ink, milk, juice
112	pen, ink, milk
113	pen, milk
114	pen, ink, juice

Consider the rule ***{pen}*** => ***{ink}***.
The support of this rule
= No. of Occurrences of ***{pen, ink}*** / Total Transaction = 3 / 4 = 0.75 x 100 = 75 %
The confidence of this rule
= Sup(***pen, ink***) / Sup(***pen***) = 0.75 / (4/4) = 0.75 / 1 = 0.75 x 100 = 75 %.

III. Existing Methods of Bitmap Encoding

Bitmap Indexing [1][2] have many variations and they are evolved for different purpose. The most recent variations on Bitmap Indexing are the Encoded Bitmap Indexing [3]. It has evolved for overcoming the problem of other encoding techniques, such as, Range Encoding, Equality Encoding etc. The main objective of the [**EBI**] is to overcome the problems of low efficiency and space consumption which affect bitmapped indices when the attributes being indexed have high cardinality domains. Since the existing technique [3] does not support most of the complex and iterative queries [6] efficiently of the data warehouse, the new approach needs to overcome this problem. Again this technique is static and thus it is not well suited for the data warehouse environment.

Our proposed dynamic Bitmap Indexing technique [**EBI**] overcomes the problems of the existing Bitmap Indexing technique.

IV. Proposed Encoded Bitmap Indexing

Here we have discussed our proposed method for well-defined encoding scheme of Encoded Bitmap Indexing in detail. The application of association rule in selecting n-items pattern in specific order is also shown.

A. Proposed Procedure for Finding Well Defined Encoding Scheme

1. Generate the 2-items rules for the dimension attribute distinct value (i.e. cardinality of the attribute), which involve in indexing fact table.
2. Read the query history source file for the dimension table attribute which involve in indexing fact table transaction by transaction and count the frequency for each rules and find support and confidence for each rules.
3. Compare the confidence of each rule against minimum confidence threshold value (i.e. 50%) to select 2-items candidate pattern.
4. Again generate 3-items rules for each of the 2-items candidate pattern starting with a highest confidence valued candidate pattern.
5. Rescan the query history source file for the 3-items rules and count the frequency for each rule and find support and confidence for each rule.

6. Compare the confidence of each rule against minimum confidence threshold value (i.e. 50%) to select 3-items candidate pattern.
7. Continue the same process for 4-items candidate pattern, up to n-items (i.e. up to attribute cardinality value) candidate pattern.
8. Select the highest confidence valued n-items pattern.
9. Remapped the mapping table of the Encoded Bitmap Index for this attribute with this n-items pattern in the way that the pattern order exists.

B. Flow Chart/Proposed Model of Encoded Bitmap Indexing

The proposed model of Encoded Bitmap Indexing technique shows the overall procedure for finding well-defined encoding scheme. The flow chart of the proposed technique is in Figure 2.

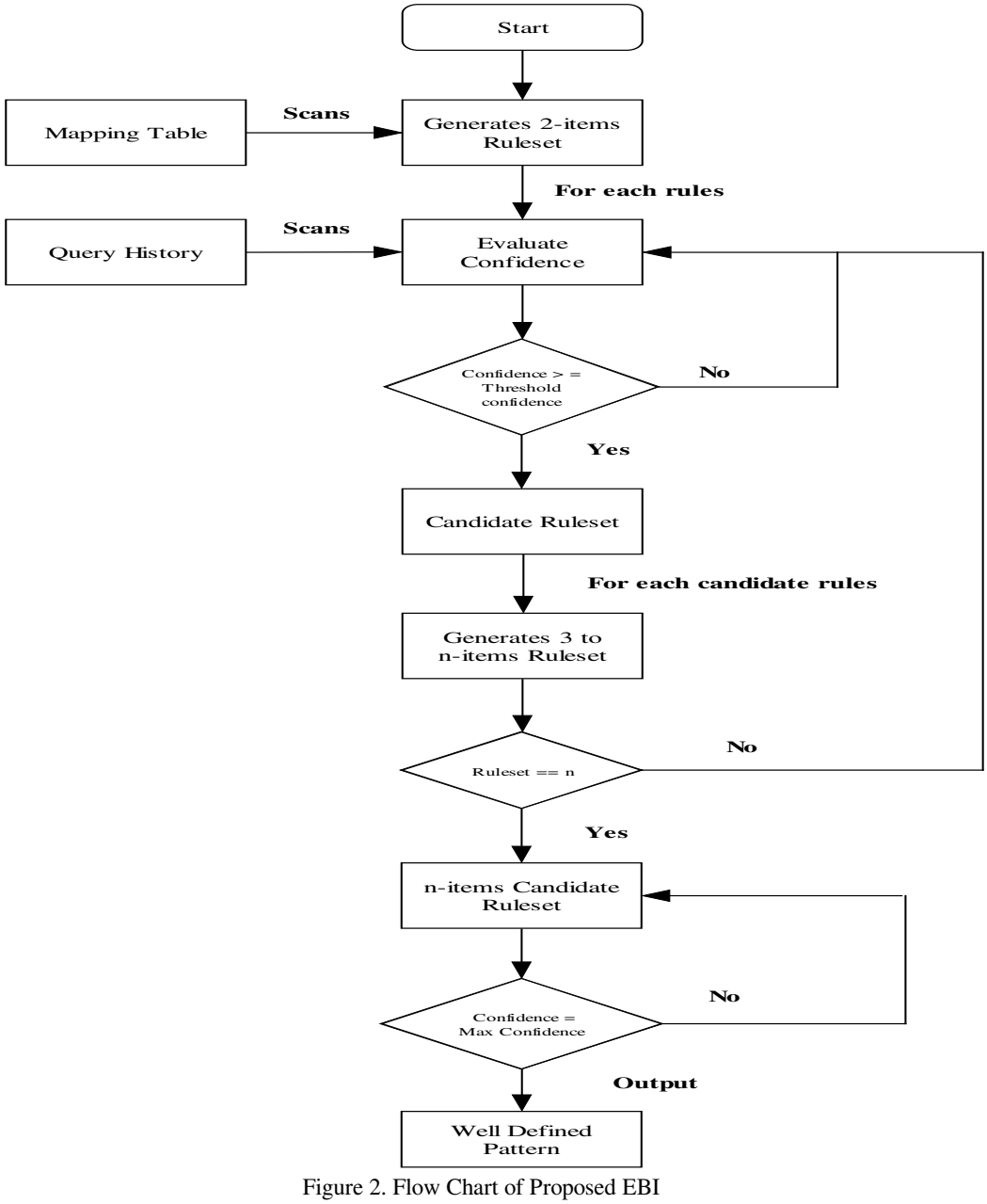


Figure 2. Flow Chart of Proposed EBI

C. Association Rule in Selecting n-items Pattern

Let consider that the cardinality of the *state* attribute of the *customer* table is 5. And *state* attribute contains the distinct values {NY, MA, CA, FA, LA}. At first the support and confidence for the 2-items rules set R_1 are evaluated (figure 3). We have selected the minimum confidence threshold for 2-items candidate pattern set C_1 is 50%. The frequent 2-items pattern set C_1 is found by comparing the confidence of the candidates with the minimum confidence threshold value. Then the 3-

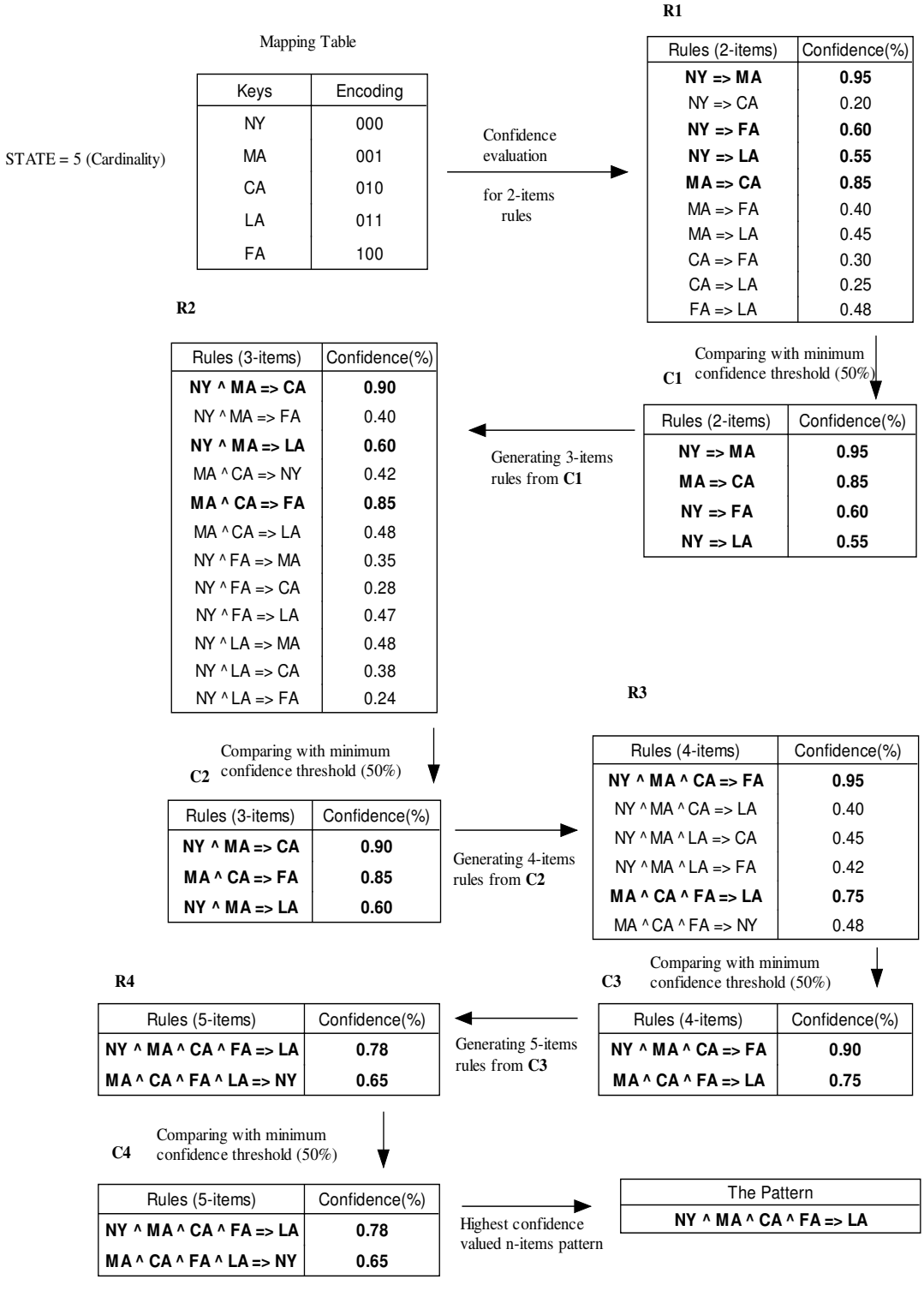


Figure 3. Selection of n-items pattern (up to 5-items) by using proposed algorithm

items rules are generated from the 2-items candidate set C_1 and the support and confidence for the 3-items rules set R_2 are evaluated. Again the frequent 3-items candidate patterns set C_2 is found form R_2 by evaluating their confidence against minimum confidence threshold value.

Then the 4-items rules are generated from the 3-items candidate set C_2 and the support and confidence for the 4-items rules set R_3 are evaluated. Again the frequent 4-Items candidate pattern set C_3 is found form R_3 by

evaluating their confidence against minimum confidence threshold value. We have decided to continue this process to find n-items pattern. And then we will select the highest confidence valued n-items pattern. For 2-items pattern we have selected the minimum threshold for confidence as 50(%) and 50(%) as well for the minimum threshold value for 3-items pattern. The 50(%) minimum threshold confidence value is selected for each n-items pattern. In each level of items pattern the pattern which exceeds this minimum confidence threshold value is selected as the candidate this level pattern from which next level patterns are generated and again compared and generated up to n-level pattern. Thus the highest confidence valued n-items pattern is selected from the n-level patterns.

For the example we have the following 5-items pattern that is very frequent which is obtain through association rule of mining of items pattern. The pattern –

NY ^ MA ^ CA ^ FA => LA

NY → MA → CA → FA → LA

This pattern will then mapped into the lookup table in the order that is extracted in the order through mining of items pattern.

IV. Comparative Study

Here we have compared between the existing Encoded Bitmap Indexing and our proposed Encoded Bitmap Indexing technique. The comparison is done through the experiment result of both the technique. In the comparison process we have used some useful criteria such as – query processing time for the sample selection predicate, scan time and found set percentage. Hence, we introduce the cost model to measure the performance of these two indexing techniques –

The response time of a query = Access time of the table level + Access time of the index level + Complexity of the algorithm.

A. Evaluation of Both Techniques with Example

For the evaluation of both the techniques we use the sample selection predicate and mapping table as an example. The cost model analysis for this sample selection predicate is given below –
If the following query were posed,

```
SELECT *
    FROM Customer c
    WHERE c.state IN {NY, MA, LA, VA} OR
          c.state IN {MA,VA,FA,NC};
```

As an example, let us consider the following encodings:

Mapping Table1 (P.EBI) Mapping Table2 (E.EBI)

NY	000
MA	001
CA	010
FA	011
LA	100
VA	101
NA	110
NC	111

NY	000
MA	001
CA	010
LA	011
FA	100
VA	101
NA	110
NC	111

Figure 4. Two different mapping for the same attribute

Here, we would need only one bitmap vectors for answering it, if we have mapping table number1 i.e. B_1'.On the other hand, six bitmap vectors are required according to the second mapping table i.e. $B_2'B_1' + B_2'B_0 + B_1'B_0$. For the first mapping table we need no logical operation also but the second one needs 3 AND and 2 OR logical operation.

Now, evaluate the response time of both the technique by using the following cost model formula:
CostModel(sc) = 1/8 * ((|sc| * att) + (|sc| * ati) + |sc| * itime(ebm))
Where sc = selective conditions
|sc| = the total number of found sets in selective conditions
m = total number of tuples in the relation
att = access time per tuple
itime = instruction time of an algorithm
ati = access time per index
ebm = Encoded Bimap Index algorithm
For the, above selection predicate
Let |sc| = found set * total number of tuples = 10% * 20000 = 2000.
att = 0.1
ati = 0.01
itime(ebm) = 1
The response time of a query using mapping table1
= 1/8 * (|sc| * (att + ati + itime(ebm)))
= 1/8 * (2000 * (0.1 + 0.01 + 1))
= 277.5 sec.
The response time of a query using mapping table2
= 1/8 * (|sc| * (att + ati + itime(ebm)))
= 1/8 * (2000 * (0.1 + 0.01*6 + 1.4))
= 390 sec.

V. Conclusion

In our proposed method towards dynamic n-items pattern selection for the remapping of lookup table of the specific attribute we have tried our best to reduce scan time complexity of query history file using normal approach. Though scan time raises significantly it may be noted that still the query processing time remains reasonable for most of the ad hoc queries.

References

[1] Alejandro A. Vaiseman, *Bitmap Indexing: FROM MODEL 204 to DATA WAREHOUSES*, University of Buenos Aires, Argentina, 1998.

[2] Chee-Yong Chan, *Bitmap Index Design and Evaluation*, Department of Computer Science, University of Wisconsin-Madison, 1998.

[3] Ming Chan Wu, Alejandro P. Buchmann, *Encoded Bitmap Indexing for Data Warehouse*, DVSI, Computer Science Department, Technical University of Darmstadt, Germany, 1998.

[4] Ming Chan Wu, Alejandro P. Buchmann, *Research Issues in Data Warehousing*, DVSI, Computer Science Department, Technical University of Darmstadt, Germany, 1998.

[5] Maria Lupetin, *A Data Warehouse Implementation using the Star Schema*, InfoMaker Inc., Glenview, Illinois.

[6] Ming Chan Wu, *Query Optimization for Selections using Bitmaps*, DVSI, Computer Science Department, Technical University of Darmstadt, Germany, 1998.

[7] Amir Michail, *Data Mining Library Reuse Patterns using Generalized Association Rules*, Dept of Computer Science and Engineering, University of Washington.

[8] www.twocrows.com, *Introduction to Data Mining and knowledge discovery*, Third Edition by Two Crows Corporation, pp, 1-10, 1999.

[9] Silberchatz, Korth, Sudarshan, *Database System Concepts*, Fourth Edition by McGraw-Hill, pp. 1-156, 445-500, 817-845, 2006.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Design of Smart Card for Automated Toll Collection at Jamuna Multipurpose Bridge in Bangladesh

Md. Arafatur Rahman, Md. Saiful Azad, Farhat Anwar and Md. Rafiqul Islam

ECE, Faculty of Engineering, IIUM
Malaysia
E-mail: arafatiiuc@yahoo.com, G0623131@stud.iiu.edu.my, farhat@iiu.edu.my, rafiq@iiu.edu.my

Abstract - This paper presents the design and development of a 16-bit smart card for automated toll collection at Jamuna Multipurpose Bridge in Bangladesh. Since it is simpler and faster than the traditional token based ticket system, it has all the potential to replace the existing system. Moreover, it saves users' valuable time by reducing the queue length in front of the toll counter. The proposed smart card has been designed using Very (High-Speed Integrated Circuit) Hardware Description Language (VHDL) and simulated in Quartus II. Finally, it is downloaded in a Field Programmable Gate Array (FPGA) chip and tested on some given scenarios.

I. Introduction

Smart cards are secure, compact and intelligent data carriers. Smart cards should be regarded as specialized computers capable of processing, storing and safeguarding thousands of bytes of data. Smart cards have electrical contacts and a thin metallic plate just above center line on one side of the card. Beneath this dime-sized plate is an integrated circuit (IC) chip containing a central processing unit (CPU), random access memory (RAM) and non-volatile data storage. Data stored in the smart card's microchip can be accessed only through the chip operating system (COS), providing a high level of data security. This security takes the form of passwords allowing a user to access parts of the IC chip's memory, or encryption/decryption measures which translate the bytes stored in memory into useful information [1].

It is economical to integrate complex systems on a single silicon die. Designing such a system on a chip is a complex process. Most chips have one or more embedded processors. Platform-based chips designs provide integrated solutions to challenging design problems in the multimedia, telecommunications, and consumer electronics domains [2]. Success will rely on using appropriate design methods as well as on the ability to test and integrate existing components including processors, controllers, and memories [3].

The automated chip card was invented by Helmut Grottrup and his colleague Jurgen Dethloff as early as 1968. In 1970's, a similar kind of patent application is submitted by a Japanese scientist Kunitaka Arimura. However, the first real progress in the development of smart card came when Roland Moreno registered his smart card for patents. He patented his first concept of the memory card in France, 1974. In 1977, Michel Ugon from Honeywell Bull invented the first microprocessor smart card. In 1978, Bull patented the SPOM (Self Programmable One-chip Microcomputer) that defines the necessary architecture to auto-program the chip. Three years later, the very first "CP8" based on this patent was produced by Motorola. Today, Bull has 1200 patents related to smart cards [4].

Jamuna Multipurpose Bridge opened in June 1998, is the longest bridge in Bangladesh as well as in South Asia, and the 11th longest bridge in the world [5, 6]. It was constructed over the river jamuna, mightiest of the three major rivers of Bangladesh, and fifth largest in the world in terms of traffic flow. Table 1 and Fig. 1 presents the average daily traffic flow of Jamuna Multipurpose Bridge in 1998-2005 respectively.

Table 1 Traffic flow of Jamuna Multipurpose Bridge in 1998-2005

Year	Light Vehicle	Bus	Truck	Total
1998*	773	660	645	2078
1999**	580	836	903	2319
2000**	602	1059	1130	2791
2001**	563	1194	1404	3161
2005***	706	1649	2358	4713

Sources: *JMBA materials [7]; **Jamuna Bridge Impact Study, ADB, 2003 [8]; ***World Bank [9]

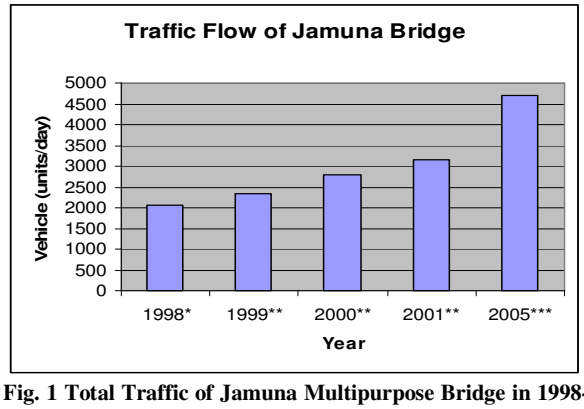


Fig. 1 Total Traffic of Jamuna Multipurpose Bridge in 1998-2005

According to the World Bank, traffic over the bridge has increased by 11.5 percent per year since its opening in 1998. The flow of trade and services is expected to more than triple by 2010 than that of 1999 [9]. Hence, approximate traffic flow by the year 2010 will be on average 7000 vehicles per day. If four booths collect toll from the customer manually at that time, each vehicle needs to be served within one and half minutes or less. However, traffic flow is not constant and it varies all the time. Number of vehicles will be lot higher in peak hours than that of in off peak hours. A huge queue will build up in front of each booth in peak hours resulting in longer waiting time for the customers.

This problem can be solved by implementing modern technology, like smart card. It will save valuable time of the customer as well as give security and it is more accurate than other available toll collecting techniques, which is shown in Table 2.

Table 2 Accuracy of several toll collection techniques

Toll Options	Toll Volumes (VPH*)	Accuracy (%)
Manual	250 - 350	98.00%
Automatic Coin Machine w/ Barrier (five coins)	450 - 550	98.50%
Automatic Coin Machine w/o Barrier (one coin/token)	500 - 700	95.00%
Vouchers/Script	500 - 900	98.50%
Smart Card w/Barrier	700 - 900	99.50%

*VPH = Vehicles per Hour; Source: World Bank, Washington DC, 3rd April, 2006 [10]

The rest of the paper is organized as follows. Design overview is described in Section II. Section III discourses VHDL modeling. Section IV and V shows and discusses simulation results and hardware implementation of the proposed design respectively. Result discussion is given in section VI. Finally, conclusion is drawn in section VII.

II. Design Overview

The proposed smart card design has three different phases and vehicles are classified into four general types illustrated in Table 3 and Table 4 respectively.

Table 3 Different phases of the system

Serial No.	Module Name	Purpose	Bit representation
0	Validation	Entry	00
1	Transition	Exit	01
2	Update	Add Balance	10

Each vehicle has a 3-bit vehicle identification number and toll is varied according to their types. The information on the smart card is stored in 16 bits, depicted in Fig. 2.

Table 4 Classification of vehicles and Fare information

Serial No	Vehicle Type	Bits Repr.	Toll (Tk.)
1	Motor Cycle	001	30.00
2	Car/Light Vehicle	010	400.00
3	Small Bus (29 or less seats)	011	550.00
4	Large Bus (30 or more seats)	100	800.00
5	Light Good Vehicle (less than 5 tonnes)	101	750.00
6	Medium Track (5 to 8 tonnes)	110	1000.00
7	Heavy Goods Vehicle (more than 8 tonnes)	111	1250.00

Sources: Jamuna Multipurpose Bridge Authority [5]; Bangladesh Bridges Authority (BBA) [11]

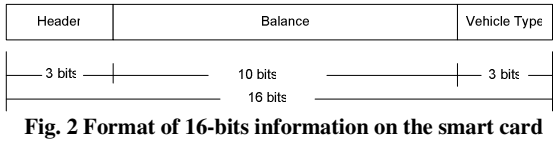


Fig. 2 Format of 16-bits information on the smart card

The purpose of "101" header is to indicate that up to 13 more bits of information need to be read by the controller. The balance indicates the remaining balance in the card where balance is represented in a special binary format e.g. equivalence of 10 taka is "00001". The last 3-bits indicate the vehicle types.

When a passenger enters the toll plaza, the system will read the information from the smart card (only header and balance) and determine if the card has sufficient funds for a minimum one-way trip according to their vehicle types. If the amount is insufficient, the card controller will generate an insufficient fund signal (gate open = 0). If the amount on the smart card is sufficient, the system will return the card to the passenger, and open the gate (gate open=1). When a passenger exits from the other end of the bridge, the system will read the 16-bit information from the card. The information of the card will be erased first (by writing 16 zeroes) and overwrite it with the updated balance and other information. It will then generate a gate open signal. When a passenger needs to add new funds to the card, the add fund function will update the balance based on added funds. It will do so by first reading the value, next resetting the smart card to zero and then writing the new 16-bit field (which includes the updated balance).

III. VHDL Modeling

The combined Smart Card module consists of 3 sub-modules as shown in Fig. 3.

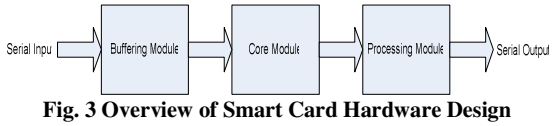


Fig. 3 Overview of Smart Card Hardware Design

All these inner modules including combined module are explained briefly in the following sections:

A. Buffering Module

Buffering module temporarily stores the incoming message block and pass this information to the Core module for further computation.

B. Core module

Core module is the most important module in our design which accomplishes the core tasks. This module comprises of 6 input and 5 output signals. The layout of the core module is shown in Fig. 3. This module performs the following tasks:

- Check the validity of the card
- Balance transition
- Balance update

This module comprises of three inner modules to perform the tasks described above. They are, Validation Module, Transition Module and Update Module. The VHDL [12] coding of these modules are written based on description given below. Validation module performs two tasks. It verifies the smart card using the header field (describe in section 2) and also checks the availability of sufficient balance in the card for a trip. This module takes 3 input (Data Input, Vehicle Type and Task Controller) and 2 output (V_Data Out and V_Gate Open) signals. Transition module performs the task of subtraction from the existing balance. This module also takes 3 inputs (Data Input, Vehicle Type and Task Controller) and 2 outputs (T_Data Out and T_Gate Open). Update module performs addition when a user needs to update his/her balance in the card. This module takes 3 inputs (Data Input, Add Fund and Task Controller) and 1 output (U_Data Out) as well. The layout of the core module is shown in Fig. 4.

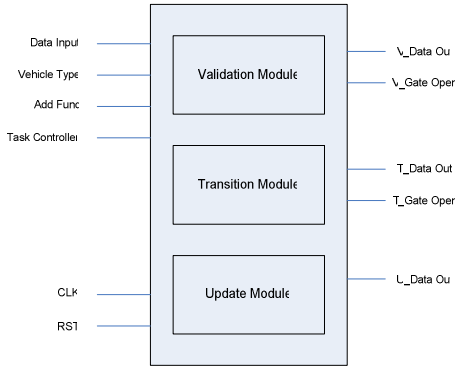


Fig. 4 Core module hardware layout

C. Processing Module

Processing module acquires the core module's output as its input. It also takes task controller signal as one of its input. Finally, it produces the output of the system according to task control bit. The hardware model layout of the processing module is shown in Fig. 5.

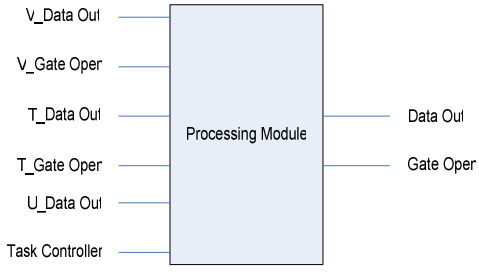


Fig. 5 Processing module hardware layout

D. Combined Module

Smart Card COMBINE Module is the combination of 3 different modules: Buffering, Core and Processing. Three modules are linked together as illustrated in the Fig. 6. The VHDL coding of each and every module is coded as a separate entity and finally combined together with a top level design.

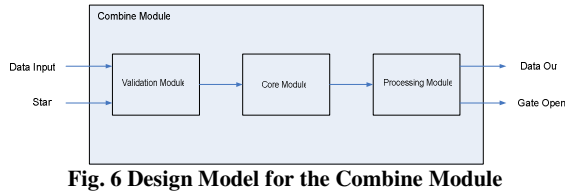


Fig. 6 Design Model for the Combine Module

The flow chart in Fig. 7 illustrates the algorithm of the proposed system.

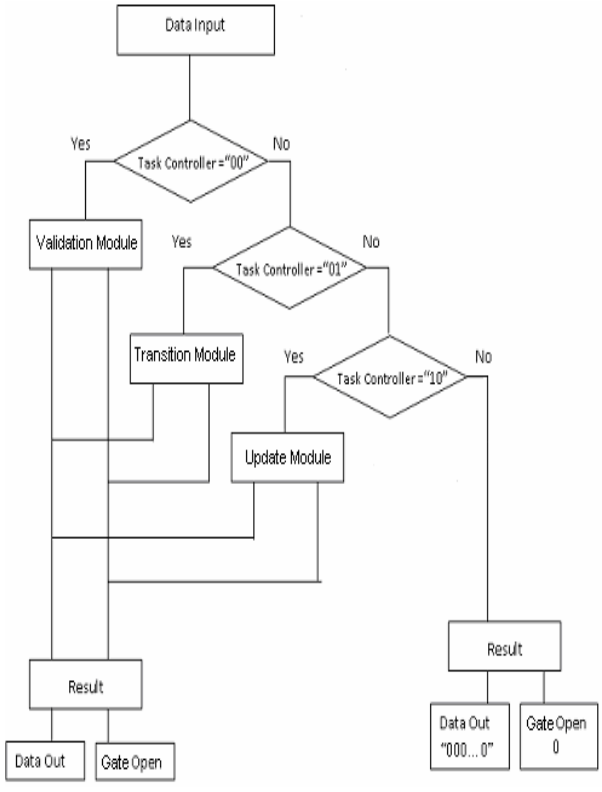


Fig. 7 Flow chart of proposed system

IV. Simulation, Synthesis and Discussion

A. Timing Simulation Module

The timing simulation is performed by using some arbitrary values. In this part the timing simulation of the core module is performed for its 3 inner modules. For the simulation following input parameters are used, which is shown in Table 5.

Table 5 Input parameters used in simulation

Serial No	Module Name	Parameter Name	Value (binary)
1.	Validation	Data_in	1010000100000000
		vehicle_type	001
		Task_controller	00
2.	Transition	data_in	1010000100000000
		vehicle_type	001
		task_controller	01
3.	Update	data_in	1010000100000000
		Add_fund	100000
		task_controller	10

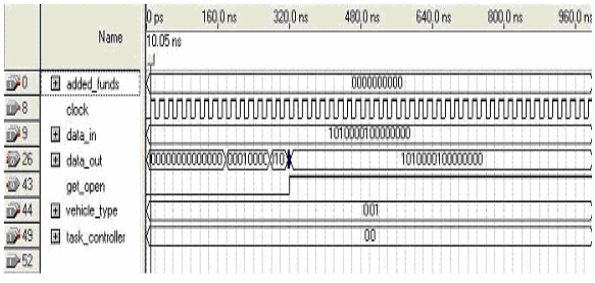


Fig. 8 (a) Timing simulation of Validation module

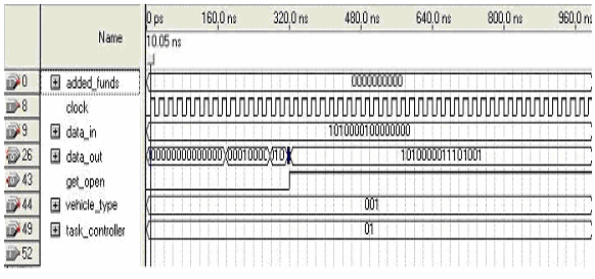


Fig. 8 (b) Timing simulation of Transition module

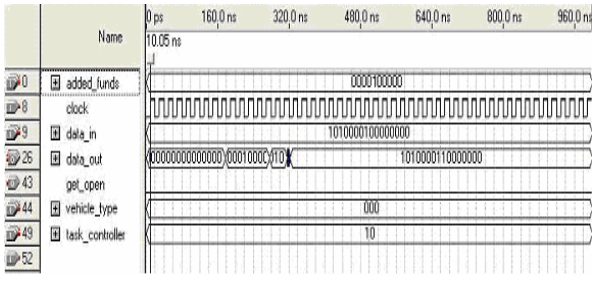


Fig. 8 (c) Timing simulation of Update module

The simulation result of the validation is shown in Fig. 8 (a). Fig. 8 (b) depicts the simulation of the transition module, while Fig. 8 (c) shows the simulation result of update module. The timing simulation is performed with 20ns simulation clock period. The parameter values of core module are given in Table 5. It can be noticed that all the inputs are in binary format. In Fig. 8 (a), 16 bits data_in, 3 bits vehicle_type and 2 bits task_controller are given as input. Since this is verification phase and balance in the card is adequate to perform at least one trip, the system sets the gate open bit high and generates 16 bits output. However, in Fig. 8 (b), similar input values are applied except task_controller bits ("01"), which indicate transition phase. As balance in the card is ample to perform the trip, the system subtracts the toll from the balance and sets the gate open bit high. Finally, 16 bits data_in, 10 bits add_fund and 2 bits task_controller are given as input in Fig. 8 (c). Here, the balance will be updated according to the add_fund value and updated balance will be stored in the card.

B. Synthesis

The VHDL code is synthesized onto Altera STRATIX II device family. For synthesis EP2S30F484C3 has been chosen to implement the smart card. From the synthesis result, it is found that the device is 31% utilized. All different modules are successfully tested and verified with a clock period 20 ns. The Register Transfer Language (RTL) view for the combine module is shown in Fig. 9.

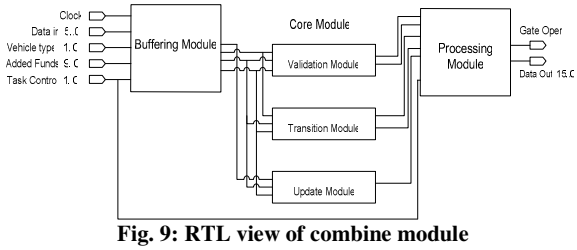


Fig. 9: RTL view of combine module

V. Hardware Implementation

The project is successfully configured and downloaded to the STRATIX II EP2S30F484C3, tested and validated. The following steps are performed:

- Prepare project for hardware design
- Set-up
- Apply power
- Configure the STRATIX II device
- Configure the Pattern Generator (PG) and Logic Analyzer (LA) devices
- Test and verify results

In the set-up, the Altera DSP Development Board is connected to a laptop using an Altera ByteBlaster download cable. On the other hand, both the PG and LA mainframes are connected to the laptop through a USB cable. All the three devices are then connected together using signal connectors, ground lines, probes and jumper wires. Power is applied to the board by connecting the 5.0V DC power supply adapter provided in the DSP kit. Once

power is applied to the board, the POWER ON LED turns on. The STRATIX II device is directly configured using Quartus II software version 5.0 Programmer. Next, the Joint Test Action Group (JTAG), a type of protocol, configures the STRATIX II device through the download cable via the ByteBlasterMV cable. Among the 16 available channels in PG, 14 are utilized as inputs to the Altera FPGA. After downloading the project into FPGA, it is not sufficient by only performing software simulations. Thus, PG is used to generate FPGA input label signals and feed them into the chip and capture signals by LA from these chip's output. This verification model reflects the real responses not only from virtual simulation by software, but it is also a real chip working result. Combining PG and LA provides an auto testing system or auto verification system. The ground lines of PG and LA are connected to the tested-circuit ground while the LA and PG grippers are connected according to the order of channel field number. The input wave patterns are then sent from the PG to the tested-circuit and the LA captures the outputs from the tested-circuit [13].

VI. Result and Discussion

Due to the shortage of pin number in the FPGA Board, 4 bits for balance, 4 bits for balance update, 2 bits for vehicle types (V(1), V(0)), 2 bits for task control (T(1), T(0)), one bit for get open and 16 bits for seven segment display are used. Thus the necessary assumptions are made as given in Table 6.

Table 6 Input parameters used in hardware implementation

Serial No.	Type	Bit	Fare(in unit)
1	Bus	00	0100 (4)
2	Truck	01	0011 (3)
3	Car	10	0010 (2)
4	Other	11	0001 (1)

In Fig. 10 (a), we assume that the initial balance of the smart card is 1000 (8) and task control bits are 01, i.e. module is in transition phase. Since vehicle type is bus ("00"), the fare of the vehicle is 0100 (4). Therefore, final balance is 8 - 4 = 4 (0100), which is displayed by the seven segment display in Fig. 10 (a).

The second test is depicted in Fig. 10 (b) where task control is 10 i.e. module is in update phase. Assuming, the user's initial balance is 1000 (8) and wants to add 0011 (3) units. As a result, the updated balance of the card becomes 8 + 3 = 11, which is shown by the seven segment display in Fig. 10 (b).

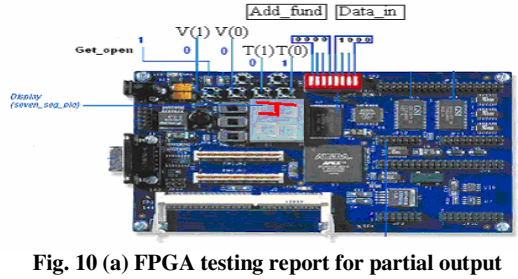


Fig. 10 (a) FPGA testing report for partial output

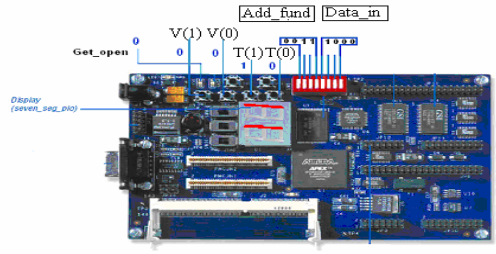


Fig. 10 (b) FPGA testing report at peak hour

From the results comparison between the two approaches software and hardware implementation, it can be concluded that the results satisfies our proposed design.

VII. Conclusion

The primary goal of this paper was to develop smart card for automated toll collection at Jamuna Multipurpose Bridge in Bangladesh. Since the traffic over the bridge increases each year by 11.5 percent from its opening in 1998, it is necessary to take some preventive measure to reduce queuing delay in front of toll plaza. An attempt is made to keep the design as simple as possible to ease the implementation. Smart card will save users' valuable time by reducing the queue length in front of the toll counter and it also ensure security. It is designed using VHDL, simulated in Quartus II, downloaded in a FPGA chip and it's functionality has been verified.

References

[1] W. Kim, S. Kim, Y. Bae, S. Jun, Y. Park and H. Cho, "A Platform-Based SoC Design of a 32-Bit Smart Card," ETRI Journal, Vol. 25, Num. 6, pp. 510-516, December, 2003.

[2] H. Chang, L. Cooke, M. Hunt, G. Martin, A. McNelly, and L.Todd, "Surviving the SOC Design evolution," A Guide to Platform-Based Design, Kluwer Academic Pub., 1999.

[3] Kyung-Hoi Huh, Yong-Seok Kang, and Sungho Kang, "Efficient Path Delay Testing Using Scan Justification," ETRI J., vol. 25, no.3, pp. 187-194, 2003.

[4] W. Rankl & W. Effing, Smart Card Handbook, John Wiley & Sons, ISBN 0-471-96720-3, 1997.

[5] Jamuna Multipurpose Bridge Authority, Available: http://www.citechco.net/jmba/

[6] http://banglapedia.search.com.bd/HT/B_0119.htm

[7] Jamuna Multipurpose Bridge Project, Report date: March 2001, Field survey: July 2000, Available: www.jbic.go.jp/english/oec/post/2001/pdf/e_project_65_all.pdf

[8] Shaikh Moniruzzaman, "Effect of Infrastructure Development on Income Generation: A Study on Jamuna Bridge Project in Bangladesh," Journal of Social Sciences, vol. 4, no. 2, pp. 80-87, 2008.

[9] World Bank, "Jamuna Bridge Connects Bangladesh's Two Halves," Available: http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21271555~menuPK:3266877~pagePK:51236175~piPK:437394~theSitePK:73154,00.html.

[10] Booz, Allen and Hamilton, Presentation "Toll Collection Systems Technology Trend Impact on PPP's & Highways' Transport," Jack Opiola, World Bank Washington D.C., April, 2006.

[11] Bangladesh Bridge Authority (BBA) (erst. Jamuna Multi-purpose Bridge Authority), Available: http://www.bba.gov.bd/index.php?option=com_tollandtariff&Itemid=30.

[12] J. Bhasker, "A VHDL Primer," Prentice Hall, Third Edition, ISBN 0-13-096575-8, 1999.

[13] M. I. Ibrahimy, M. B. I. Reaz, K. Asaduzzaman and S. Hussain, "FPGA Implementation of RSA Encryption Engine with Flexible Key Size," International Journal of Communications, vol. 1, no. 3, pp. 107-113, 2007.

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

SET Based Logic Realization of a Robust Spatial Domain Image Watermarking

D. Samanta*, A. Basu#, T. S. Das, V. H. Mankar, Ankush Ghosh, Manish Das and Subir K Sarkar

*Centre for Educational Technology, Indian Institute of Technology, Kharagpur, India
#Dept. of Radiophysics and Electroncs, University of Calcutta, Kolkata-700 092, India
Dept. of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata-700 032

E-mail: sksarkar@etce.jdvu.ac.in

Abstract - The proliferation of the digitized media introduces an exigent problem of security in multimedia data transmission in the network surroundings. Consequently, digital watermark technique needs to be incorporated in a digital rights management framework to address different aspects of the content management. The objective is to develop low power, real time, reliable and secure data hiding system that can be achieved all the way through hardware implementation. As an attempt towards the power efficient system, here we present an oblivious, spatial domain watermarking based authentication algorithm and its VLSI/ULSI architecture. It is realized with the help of Single electron tunnelling (SET) devices where classical binary bit '1' and '0' are encoded. The advantage of the SET devices is that it requires much less power than conventional technologies. The proposed scheme employs the lower bit plane modulation for hiding a binary image like information within a gray scale image. All the modules of the watermarking algorithm are implemented in SET device based Logic along with its computer simulation. It is shown that the low cost data embedding algorithm can conceal watermark into original cover image coming from a sensor much faster than software implementation and the embedded image is easily transmitted to PC by using proper interface. The quality of the transmitted image is also comparable to that of the implemented by software algorithm.

I. Introduction

rapid development of internet and the wide availability of digital consumer devices such as digital cameras, scanners etc. improved the ease of access to digital information. This leads to the problem of illegal copying and redistribution of digital media. The digital watermarking came as an efficient solution of problems related to protection of copyright and the intellectual property of the media. The watermarking embeds a signal into the host data in some invisible way that is supposed to identify the owner [1-3].

Watermarking schemes can be classified into various ways. There are basically two types of watermarking scheme, spatial domain and frequency domain according to working field. The frequency domain schemes are generally considered more robust than the spatial domain schemes and are based on DCT (discrete cosine transform) and DWT (discrete wavelet transform) in general. Various techniques are introduced and applied to watermarking schemes such as Lower Bit Plane Modulation (Least Significant Bit), Human Visual System (HVS), Spread Spectrum (SS) and Quantization Index Modulation (QIM) [4].

Over the past decade, numerous watermarking algorithms have been introduced and their software is available, however recently some hardware implementations are being presented in the literature. Software implementations have been developed due to the ease of use, upgrading and flexibility but at the cost of limited speed problem and vulnerability to the offline attack. On the other hand hardware realizations offer advantage over the former in terms of less area, low execution time and low power. Therefore, it may be desirable to mark the image inside digital devices to authenticate the original data content right at the origin since offline watermarking process can not guarantee the tamper proof [5-6]. Hence spatial domain is always preferred for hardware implementation compared to the transformed domain owing to a smaller amount of computational complexity.

SET is one of the new technology where device size is very small and have ultra low power dissipation capabilities. SET devices are indispensable elements in nano-electronic field and used the circuits such as memory [7], logic circuits [8], digital to analog converter [9] etc. In previous research, different logic gates, flip-flops, addition, multiplication, division, ALU etc. using single electron devices have been implemented [10-14].

A tunnel junction can be thought of as a leaky capacitor. Electrons are considered to tunnel through a tunnel junction one after another. Even only one electron tunneling may produce a charge e/C across the tunnel junction (where C is total capacitance and e = 1.602 x 10^{-19} C). The critical voltage Vc across the tunnel junction can be calculated with the equation

$$V_c = \frac{e}{2(C_e + C_j)} \quad (1)$$

Where C_j is the junction capacitance and C_e is the equivalent capacitance for remainder circuit [14].

In view of the above discussion it appears that the spatial domain data hiding scheme has lot of significant applications and single electron devices are potential better performer than their traditional electronic counterparts. These in essence motivate us to realize single electron logic based hardware implementation of a spatial watermarking technique for the applications in intellectual property protection. In this paper a spatial domain data hiding scheme and its hardware implementation using SET device based logic circuits are proposed. The algorithm can be used for real time data transmission through digital media. The reliability in the transmission of message is further improved by using private key encoding scheme incorporated to the specific bit plane of the watermark signal based on their relative significance.

II. Watermark Embedding and Detection

A. Review Stage

The algorithm considered here is a digital modulation scheme that employs synchronous detection for decoding of the information. The binary image is used as an information bearing signal. It conveys the unique information with a good degree of resiliency after various forms of image impairments. The cover data is considered as a carrier, a gray scale watermarked image called as modulated signal is generated using spatial domain LSB modulation scheme. The modulated signal is transmitted either through a secure channel or noisy one subject to the distortion constraint. Finally, the message is extracted at the receiver from this modulated signal using synchronous detection.

In the watermark encoder, the 2-D pixel values of the gray cover as well as binary message are quantized into their stored integer (SI) data type with respective word lengths. Then all these SI type 2-D pixel values (cover and message) are converted into 1-D signal individually. The buffer block redistributes these signals to a new frame size. The binary message string is then redundantly embedded using a pseudorandom (PN) sequence corresponding to a specific state as a private key. Each bit of the binary string replaces suitable LSB plane of the specified cover pixel values. The choice of bit plane for embedding the watermark is made by a proper trade-off between the estimated tendency of possible changes in the gray values and imperceptibility. Here the spatial pixel values as well as binary string both are undergone through fixed point quantization. As the data hiding operation is over, the unbuffer block unbuffers frame-based input into a sample-based output and further converted into 2-D watermarked image. The block diagram of the watermark encoder is shown in Fig. 11.

Similarly, the stego image with or without external attack is transformed to 1-D sequence having SI type pixel values with fixed point quantization in the watermark decoder. The embedded binary data is picked from the bit plane specified by the same private key being used during data embedding and the final estimated bit has been selected using a majority rule. These values represent the pixel values of the decoded message signal. Finally, the same buffer and unbuffer blocks are required here also for the speed of response compatibility. The watermark decoder as a model is shown in fig. 12.

III. Hardware Realization

Low dimensional structures, also called nanostructures, provide opportunities to realize high speed, low power consuming devices. As a consequence, the search for the new principle of the small size devices is becoming more and more important. At present there are two main branches of the proposal on the suitable operation principles, so called “quantum electron devices” and “single electron devices”. Here a spatial domain watermarking encoder and decoder are designed using single electron logic devices.

IV. Encoder

The watermarking encoder embeds the binary string into the grayscale host image. An encoder in block diagram level (so designed) is depicted in Fig. 11. At first, 8 bit serial data is inserted into the shift register. A combinational circuit determines the perceptually significant pixel values for watermarking. This is basically a HVS scheme, where a pixel seems to be greater than threshold is selected for the data concealment. The combinational function is given as follows:

f (A, B, C, D, E, F, G, H) = ABC’DE’ + ABCDE’ + ABCD’E’ + ABEF + ABEF’ (2)

The output of this combinational block acts as a selection input to a multiplexer (2:1) so that this multiplexer output results in a clock (R=1). This is further forwarded to the preloaded shift register for giving a watermark bit to the input of second multiplexer. Now, depending on the logical result of the combinational function, either multiplexer results in binary bit to replace the third bit plane position of the selected spatial pixel value or the pixel is left unchanged. The 8:1 multiplexer performs this embedding or bit replacement operation. Finally the changed/unchanged pixel as a watermarked is loaded into the output shift register. The mod 8 counter along with two combinational circuits count the total number of bit replacement operation and reset the shift register. The block diagram of the encoder is given in Fig. 11.

V. Decoder

The decoding operation is just the reverse to that of the embedding or image noising. The watermarked data in serial binary form is loaded into the 8 bit shift register. The same combinational circuit as being used in the encoder is engaged to compute the threshold limit and to select the perceptually significant, appropriate pixels. Then through the use of multiplexer, the hidden watermark bit in the third bit plane is selected and loaded

into the output shift register. Here the mod 8 counter helps to count the total number of decoding operation and when the condition is equal to the length of the watermark, it finishes the decoding operation. The decoding block is given in Fig. 12.

VI. Operation Principle

The proposed circuit constructed here using single electron devices. An electron (say, messenger electron) travels along the path and reaches leaf (0 or 1). Figure 1 and figure 2 shows the block diagram and circuit diagram of node represented in BDD graph using single electron device [15]. It consist of four tunnel junctions J_1, J2, J3 and J_4 driven by a clock Φ. If variable is 1, X will be positive voltage (X' negative) and if variable is 0, X will be negative voltage (X' positive). A messenger electron received by device element from preceding device through entry branch or root and sends the electron through exit branch. Multiple clock phases (Φ_0, Φ_1, Φ_2 and Φ_3) are required for this purpose.

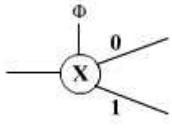


Fig. 1: Block Diagram of a BDD Graph

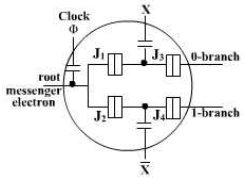


Fig. 2: Basic operation of Tunneling Junction

VII. Explanation of SET based Circuit Operation

As evident from the picture delineated in Fig 11 the spatial domain watermarking based authentication algorithm comprises of several individual units. It is comprehensive to explain the operation of each unit based on SET devices and then amalgamate the units to interpret the overall functionality of the circuit.

A. Unit 1 (2x1 Multiplexer)

This unit is employed for performing the multiplexing operation in order to generate the variable clocking/signaling action prescribed for the realization of the system. The SET device based diagram of this unit, depicted in the figure, incorporates using two input AND gate, two input OR gate and one inverter. Now if S be the select line and I_1, I_2 be the inputs to the MUX then the output is given by

$$Y = \overline{S}I_1 + SI_2 \tag{3}$$

This logical operation is efficiently performed by the SET device based architecture depicted in the Fig.3. I_1, I_2 and S are applied to the inputs of circuit couple with their corresponding select signals and are propelled towards AND and next OR operation in their wake respectively, so that the logical signals can be generated at ultimate output hat emerges reflects the normal multiplexing functionality.

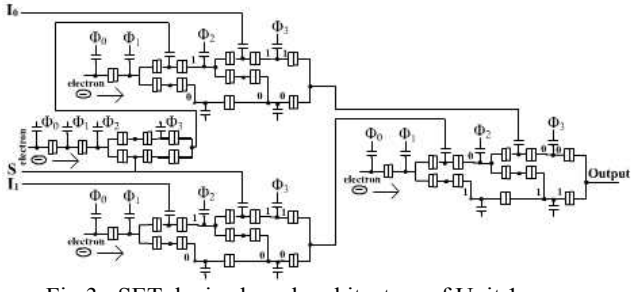


Fig.3 : SET device based architecture of Unit 1

B. Unit 2 (8x1 Multiplexer)

This unit is employed for performing the watermark encoding operation by selecting the proper bit plane in order to generate the watermarked signal prescribed for the realization of the system. 8x1 multiplexer is derived here using 3nos. NOT gates, 8nos. four inputs AND gate and one eight inputs OR gate. If S_0 S_1 and S_2 be the select lines and I_0, I_1, I_2, I_3 I_4, I_5, I_6, I_7 be the inputs to the multiplexer then the output is given by

$$Y = \overline{S_0}\overline{S_1}\overline{S_2}I_0 + \overline{S_0}\overline{S_1}S_2I_1 + \overline{S_0}S_1\overline{S_2}I_2 + \overline{S_0}S_1S_2I_3 + S_0\overline{S_1}\overline{S_2}I_4 + S_0\overline{S_1}S_2I_5 + S_0S_1\overline{S_2}I_6 + S_0S_1S_2I_7 \tag{4}$$

This logical operation is efficiently performed by the single electron device based architecture depicted in the fig. 4. The outputs of OR gate is any one of the inputs depending on select lines and the ultimate output reflects the normal multiplexing functionality.

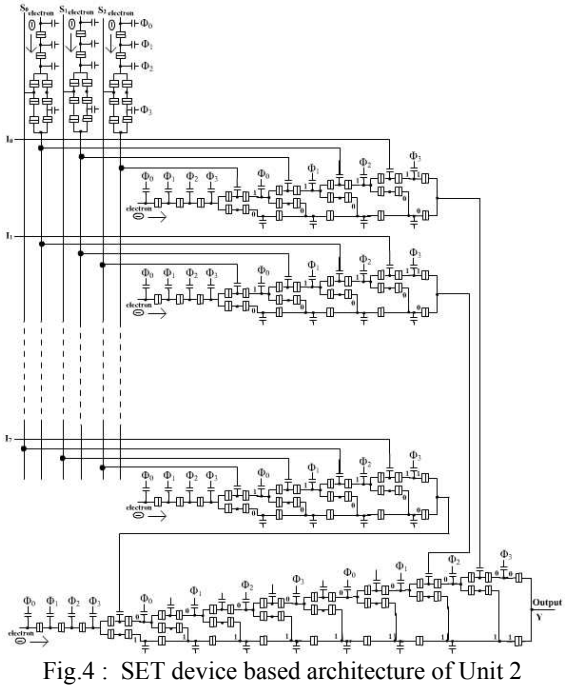


Fig.4 : SET device based architecture of Unit 2

C. Unit 3 (8-bit Shift Register)

This unit imports the spatial pixel values into the subsequent encoder/decoder circuit as well as exports or loads the output watermarked pixels or decoded data. Logic design of this unit is composed of eight D flip flops that are cascaded and connected to a synchronized clock. The operation of the single electron device based architecture of a shift register using D flip flops is depicted in the fig. 5.

The SET device based architecture is needless to be explained as the readers will certainly recognize the algorithm behind such a design.

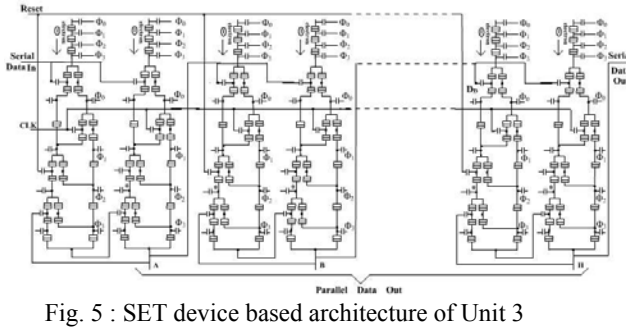


Fig. 5 : SET device based architecture of Unit 3

D. Unit 4 (Mod 8 Counter)

This unit simply counts the number of bit wise watermark encoding and decoding operations. Units 6 and 7 help this block in this regards. Basically when the unit 6 reaches to a logical state with all the selection lines high, unit 7 momentarily reset the shift register and complete the encoding operation. SET device based Mod 8 counter constructed using three D- flip flops and three XOR gates.

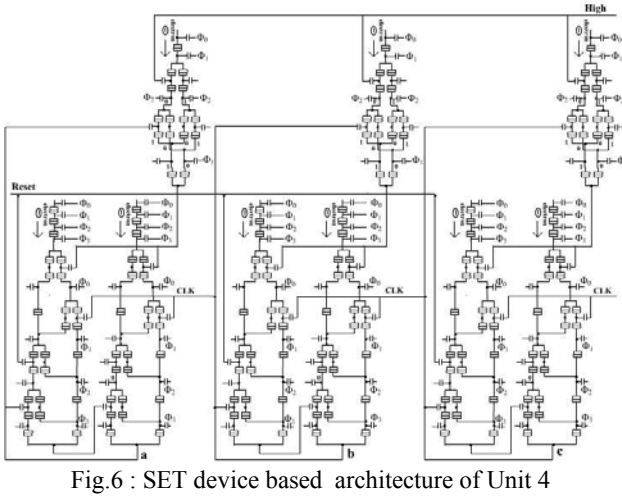


Fig.6 : SET device based architecture of Unit 4

E. Unit 5 (Combination Circuit-1)

This unit is used to select the specific pixel values for data hiding. It provides a private key based adaptive threshold scheme so that perceptually significant spatial pixels can be obtained. The pixels so received will be used for the data concealment using the lower bit plane modulation methodology. The combinational circuit for unit 5 is implemented using SET device based AND, OR, NOT gate as shown in figure 7.

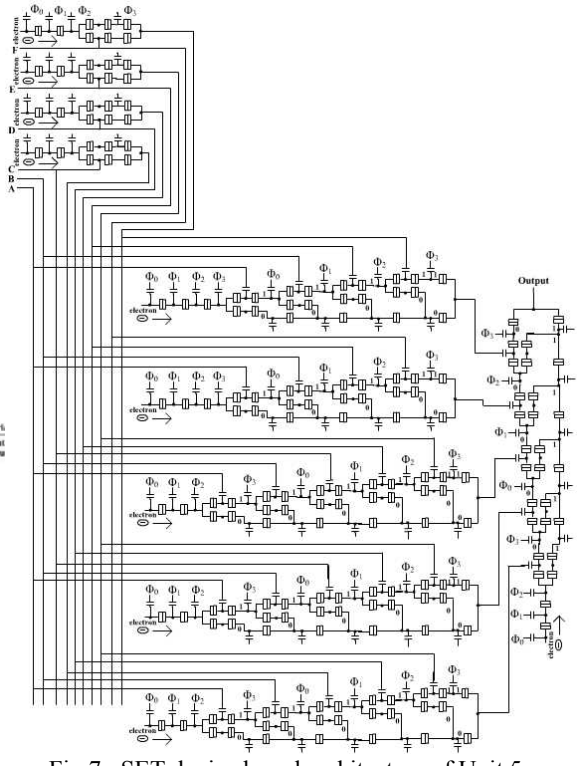


Fig.7 : SET device based architecture of Unit 5

F. Unit 6 (Combination Circuit-2)

Unit 6 is an accessory block to give support to the mod-8 counter (unit 4) in order to calculate the total number of bit replacing operations. The output of this unit is connected to the I_3 input of the multiplexer. The output of AND gate is connected with the input of a NOT gate and produce the desired value.

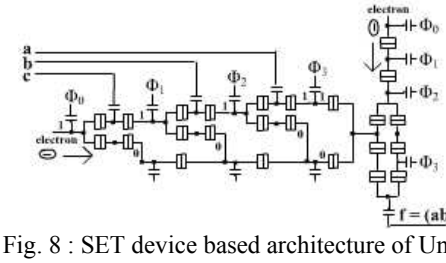


Fig. 8 : SET device based architecture of Unit 6

G. Unit 7 (Combination Circuit-3)

Unit 7 is an inversion block to that of the unit 6 and helps to format the input shift register after hiding the entire message bit string into the selected pixel values.

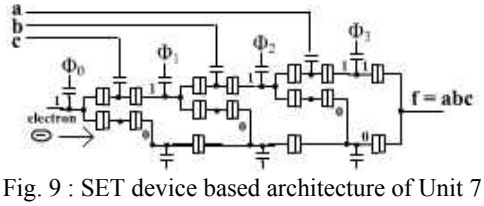


Fig. 9 : SET device based architecture of Unit 7

VIII. Results and Discussion

A SET device based circuit for a spatial domain watermarking based authentication algorithm is designed and implemented in the present work. Single electron devices are the basic building blocks because of their low power consumption, high speed operation and tiny size. In order to substantiate our designed SET device spatial domain watermarking based authentication algorithm we have made computer based simulation. Simulated signals for different situations are fed to the watermarking system so designed. It is observed that activities of the designed circuits are as per expectation thereby establishing the fact that the design is proper.

Table 2 Results for PSNR, SSIM, Security Values and Mutual Information

Image	PSNR (dB)	SSIM	Mutual Information
Lena	39.74	0.9927	0.287313
Fishing Boat	39.68	0.9976	0.287313
Peppers	42.11	0.9951	0.287313
Baboon	40.81	0.9943	0.287313
USAir Force	37.61	0.9809	0.287313

Table 3: Robustness Efficiency

Attacks	Bit Error Rate
Median Filtering	0.1117
LPF	0.1635
Histogram	0.3559
Cropping	0.2194
Inverting	0.1178
Edge Encoder	0.2107
Range [up-215 low-25]	0.2114
Gaussian Filtering	0.1313
Add Noise [pix-10% amnt-20%]	0.1252
Scaling [256-128-256]	0.2763
Erode	0.2133
Dilate	0.2133
Edge	0.3695

The cover image I is a gray scale image having a size M X M where $M=2^i$. The auxiliary message or watermark is a binary image of size N X N where $N=2^j$, where $i>j$. Results are computed here considering one bit per pixel binary image of size 16X16 as watermark and 256X256, 8 bits per pixel gray as cover data. The buffer block frame size being used in data concealment is 1 by16. The data is embedded in third bit plane in the host signal. The resiliency performance of the proposed algorithm has been evaluated in terms of various image distortions as well as noisy modulated signal similar to the concept of drift in oscillator carrier frequency in synchronous detection. It has been observed that the detected watermarks are quite subjectively recognizable even after a higher depth of degradations occurred in the watermarked image. The various transparency metrics are given in table 1. The experimental results of the imperceptibility and robustness performance are given in table 2 & 3 and Fig. 13.

During the hardware realization, in order to obtain the simulation output block level models is run under the fix point quantization with stored integer (SI) data type format. Afterward the VHDL codes (subsystem, subsystem_pkg, embedded function and timing controller) are generated from the subsystem-watermark encoder having embedded functions. The VHDL codes are simulated in ModelSim SE/PE with appropriate input signals (i.e. cover data and watermark signals in standard logic form). All the VHDL codes are then simulated in Xilinx ISE 9.1i for obtaining the RTL representations as well as technology views of the proposed watermark encoder required for the proposed hardware design. The similar process is applied for the subsystem watermark decoder. The RTL representations and technology views for the proposed watermarking technique along with modelsim output are given in Fig. 10.

IX. Conclusion

An algorithm for spatial domain oblivious watermarking and its VLSI/ULSI realization using Single Electron Devices is proposed in this paper. Subjective recognition of the decoded message is possible without the original cover image i.e. the scheme is blind. The simulated results show better statistical invisibility and resiliency against several image degradations. The proposed scheme requires only a few simple computations and VLSI/ULSI implementation allows the scheme for real time multimedia data transmission. This anticipated work has immense application for commercial applications in the area of medical imagery, low power verification system etc. The present work is step forward to develop watermarking technique as a digital communication system i.e. incorporates proper type of interleaving, source and channel coding, modulation and appropriate synchronous or asynchronous detection method using the single electron device chip in the future.

Acknowledgment

Subir Kumar Sarkar thankfully acknowledges the financial support obtained from DRDO, Govt. of India vide order no. (ERIP/ER/0503561/M/01/905 dated 01/08/2006)

References

[1] I. Cox, M. Miller, J. Bloom, "Digital Watermarking", Morgan Kaufmann Publishers, 2002.

[2] I. J. Cox, J. Kilian, F. T. Leighton, T. Shamoon, "Secure Spread spectrum watermarking for multimedia," *IEEE Transaction on Image Processing*, volume 6, pages 1673-1687, 1997.

[3] M. Barni, F. Bartolini, A. Piva, "Improved Wavelet based Watermarking through Pixelwise Masking," IEEE Trans. On Image Processing, Vol. 10, No. 5, pp. 783-791, 2001.

[4] T. S. Das, V. H. Mankar, S. K. Sarkar, **"**Spread Spectrum based *M*-ary Modulated Robust Image Watermarking" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.10, pp 154-160, October 2007.

[5] H. Lim, S. Park, S. Kang, W. Chao, "FPGA Imaplementation of Image Watermarking Algorithm for a Digital Camera," IEEE Conf. pp. 1000-1003, 2003.

[6] S. P. Mohanty, R. K. C, S. Nayak, "FPGA based Implementation of an Invisible-Robust Image Watermarking Encoder," LNCS3356, pp. 344-353, 2004.

[7] Kozo Katayama et al "Design and analysis of high-speed random access memory with coulomb blockade charge confinement" - IEEE trans. Electron Devices, vol. 46, no.11, November 1999

[8] Hiroki Iwamura et al "Single-electron majority logic circuits" - IEICE trans. Electron vol. E81-C No. 1 January 1998, pp. 42-48

[9] Casper Lageweg et al " Digital to analog conversion performed in single electron technology" – IEEE S3.1 Nanocircuits and architectures, oct 2001, pp. 105-110

[10] Mawahib Sulieman et al "On single electron technology full adders" - 4th IEEE conference on nanotechnology, 2004 pp. 317-320

[11] Casper Lageweg et al "Binary multiplication based on single electron tunneling" – Proceedings of 15th IEEE international conference on application-specific systems, architecture and processors (ASAP '04)

[12] A.K.Biswas, S.K.Sarkar "An arithmetic logic unit of a computer based on single electron transport system" – Semiconductor physics, Quantum electronics and optoelectronics 2003, V 6, pp 91-96

[13] Casper Lageweg et al "Static buffered SET based logic gates" – IEEE Nano, Aug 2002 pp. 491-494

[14] Casper Lageweg et al "Single-electron encoded latches and flip-flops" – IEEE trans. On nanotechnology, vol.3, no.2, june 2004.

[15] Noboru Asahi et al "Single-electron logic systems based on binary decision diagram"- IEEE trans. Electron Devices, vol. 44, no.7, july 1997 pp.1109-1116.

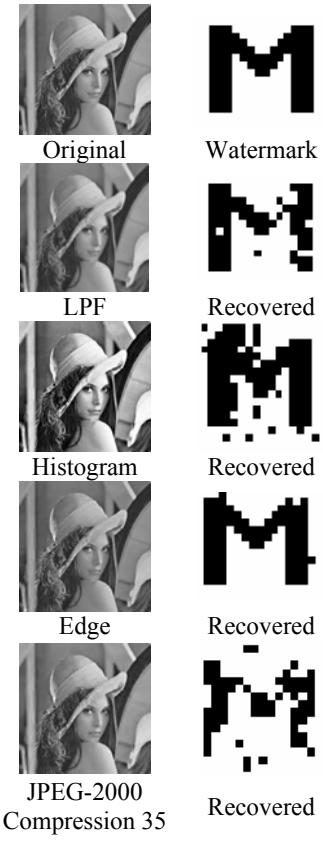


Fig. 10

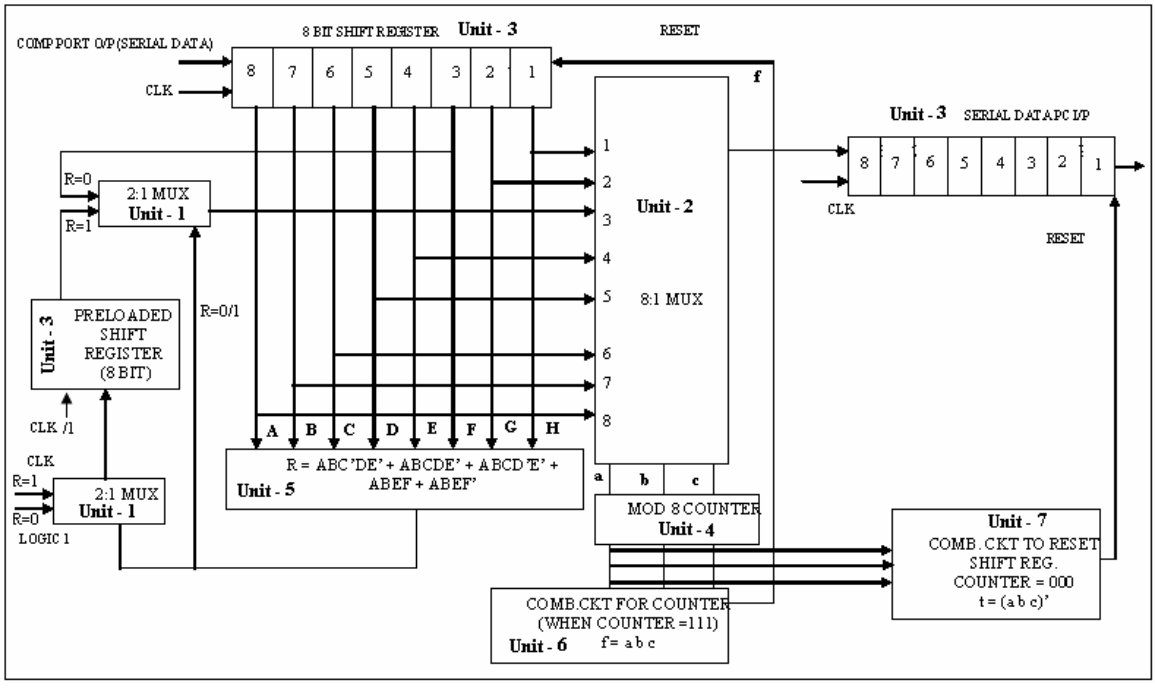


Fig. 11 : Watermark encoder

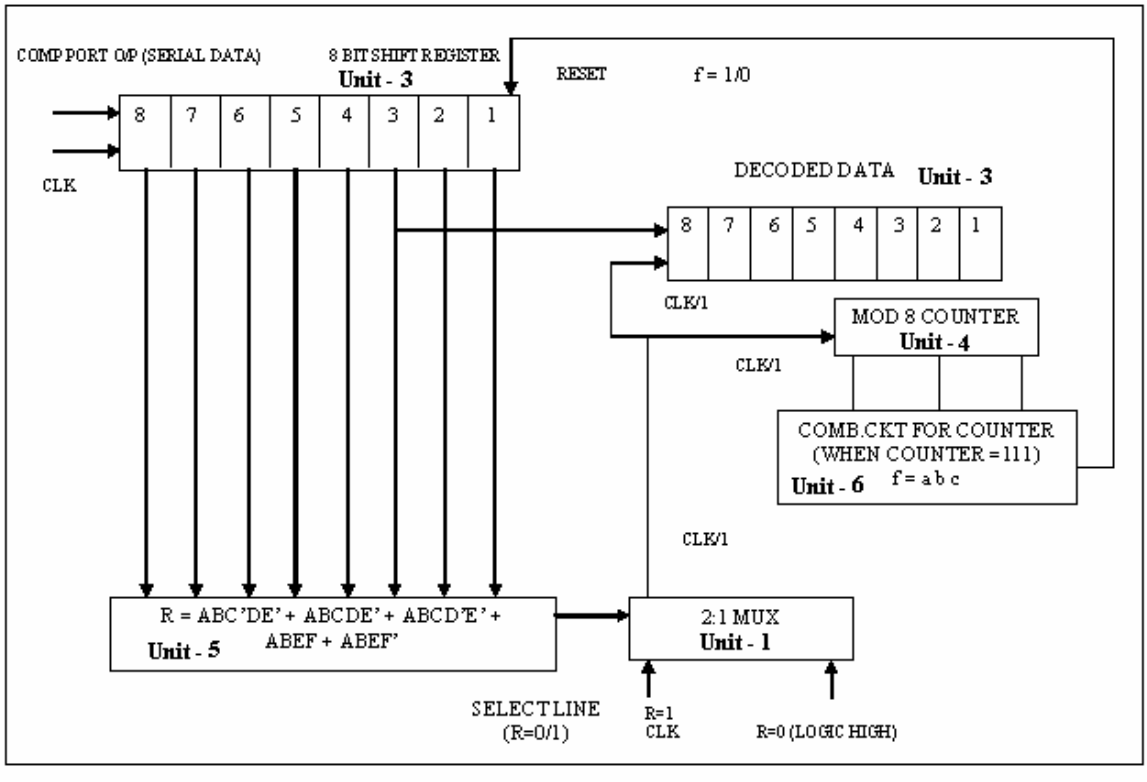
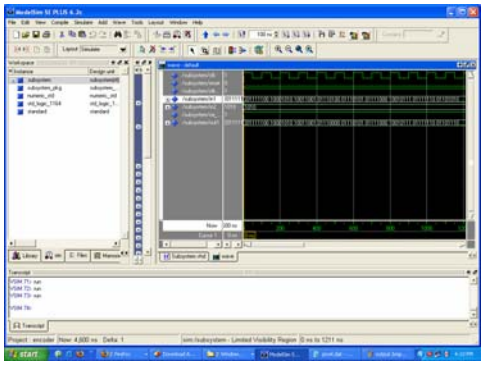
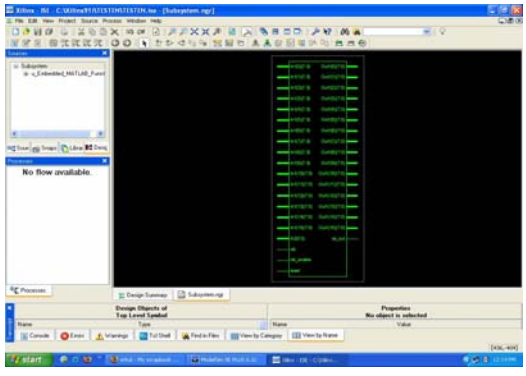


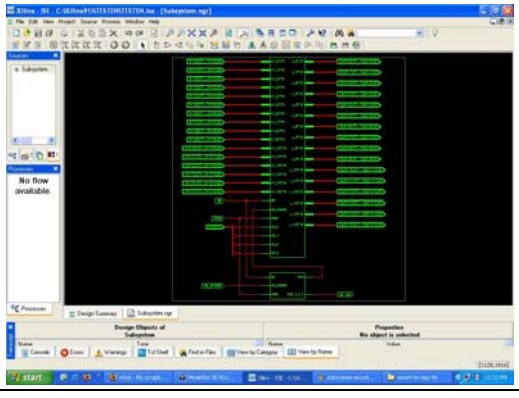
Fig. 12 **:** Watermark decoder



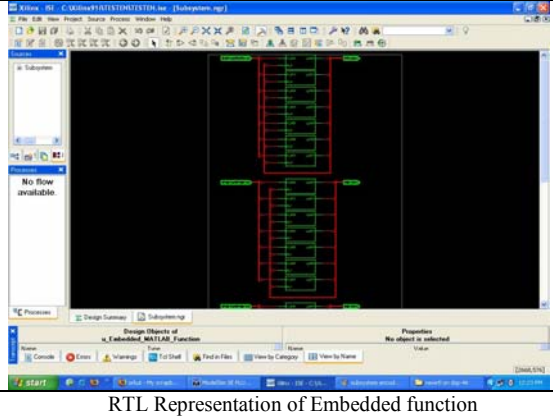
Watermark Encoder output in ModelSim



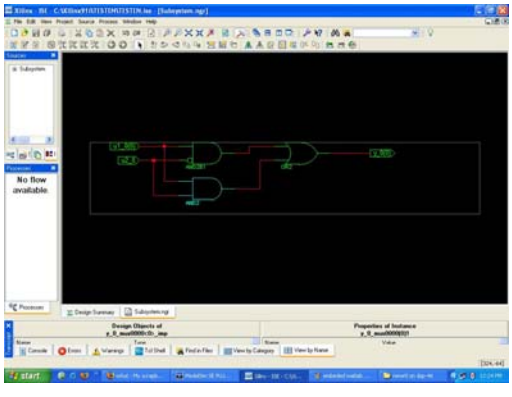
RTL Representation of Top level -Watermark Encoder in Xilinx



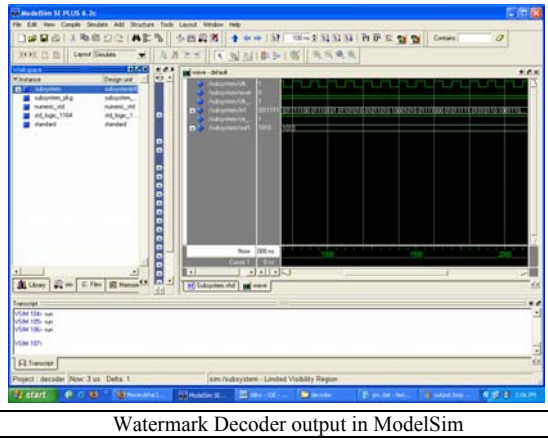
RTL Representation of Subsystem -Watermark Encoder



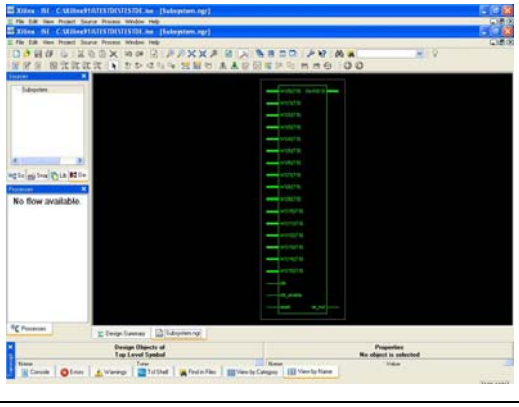
RTL Representation of Embedded function



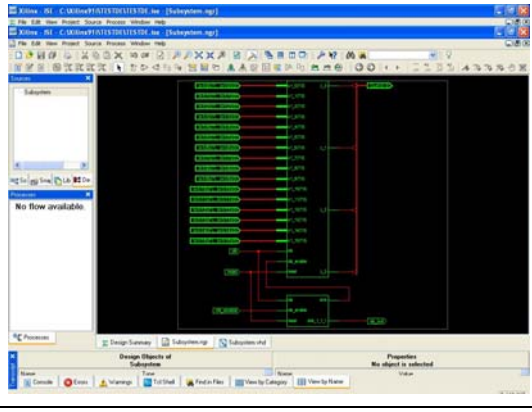
Basic RTL Representation of Watermark Encoder



Watermark Decoder output in ModelSim



RTL Representation of Top level -Watermark Decoder



RTL Representation of Subsystem -Watermark Decoder

Fig. 13

5th International Conference on Electrical and Computer Engineering
ICECE 2008, 20-22 December 2008, Dhaka, Bangladesh

Agent's Decision Making Based on Cultural Values

Zeinab Mazadi, Nasser Ghasem-Aghaee, Mohammad Ali Nematbakhsh

Department of Computer Engineering, University of Isfahan
Isfahan, Iran
z.mazadi@gmail.com, nasser_ga@yahoo.ca, mnematbakhsh@yahoo.com

Abstract- This paper presents a model for cultural intelligent agent decision making. The proposed model is based on Schwartz 10 value type. We follow a fuzzy approach for identification of agent's values and cultural dimension. Each cultural value causes a set of behavior that according to its importance is performed by agent. In each situation agent selects nearest situation in comparison with its criteria. These criteria are explored by agent according to cultural values. We use of fuzzy J and aglet for implementation.

1. Introduction

Some aspects of human cognitive and affective characteristics are taken into account in advanced agent simulation, i.e., simulation of systems represented as software agents. It is well known that culture also affect human decision making. Our research team aims to glean appropriate theories and pragmatic findings on human culture and explore ways in which they can be embedded on the decision making of advanced software agents.

A group of people's way of life is culture that they accept, generally without thinking about them, and that are passed along by communication and imitation from one generation to the next [1]. Different cultures have usually different standards for perceiving, believing, evaluating, communicating, and acting. Hence culture affects decision making and norm of behavior in different situations [2]. In negotiation, decision making and team-working simulations, using cultural intelligent agents with different norms of behavior and ways of communication can be useful for solving some problems such as conflict [3].

2. What is culture?

A group of people's way of life is culture that they accept, generally without thinking about them, and that are passed along by communication and imitation from one generation to the next [1]. Most of the people who study culture agree that culture is learned. It is not encoded in our DNA. Each culture has standards for acting and communicating that are passed down from one generation to the next. Culture is created by people and is embodied in both physical and social artifacts. [2]

"Studying differences in culture among groups and societies presupposes a neutral vantage point, a position of cultural relativism. A great French anthropologist, Claude Lévi-Strauss (born 1908), has expressed it as follows: 'Cultural relativism affirms that one culture has no absolute criteria for judging the activities of another culture as 'low' or 'noble.' However, every culture can and should apply such judgments to its own activities, because its members are actors as well as observers.'" [1].

To study culture, we first describe manifestations of culture. Cultural differences manifest themselves in different ways and differing levels of depth. Cultures consist of values, rituals, heroes, and symbols [1]. Symbols represent the most superficial and value the deepest manifestations of culture, with heroes and rituals in between.

Symbols are words, gestures, pictures, or objects that carry a particular meaning which is only recognized by those who share a particular culture.

Heroes are persons, past or present, real or fictitious, who possess characteristics that are highly prized in a culture.

Rituals are collective activities, sometimes superfluous in reaching desired objectives, but are considered as socially essential.

Values The core of a culture is formed by values. They are broad tendencies for preferences of certain state of affairs to others (good-evil, right-wrong, natural-unnatural) [1].

Values form the core of a culture. That is why value transfer between cultures is a rare phenomenon. Thus, discussing the cultural differences is in fact highlighting the differences between cultural values. Values also form cultural dimensions of society. For example, American decision making style is more individual than that of Japanese people. Main features of values, as formulated by Schwartz in value theory are the following [4, 5]:

Values are beliefs linked inextricably to affect. When they are activated, they become infused with feelings.

Values refer to desirable goals that motivate action.

Values transcend specific actions and situations (e.g., obedience, honesty, independence). This distinguishes values from narrower concepts like norms and attitudes that usually refer to specific actions, objects, or situations.

Values serve as standards or criteria. Values guide the selection or evaluation of actions, policies, people, and events. People decide what is good or bad, justified or illegitimate, worth doing or avoiding, based on possible consequences for their cherished values. The impact of values in everyday decisions is rarely conscious. Values enter awareness when the actions or judgments one is considering have conflicting implications for different values one cherishes.

Values are ordered by importance, relative to one another. People's values form an ordered system of value

priorities that characterize them as individuals. This hierarchical feature also distinguishes values from norms and attitudes.

The relative importance of multiple values guides action. Any attitude, opinion or behavior typically has implications for more than one value. … The tradeoff among relevant, competing values is what guides attitudes and behaviors [6]. Values contribute to action to the extent that they are relevant in the context (hence likely to be activated) and important to the actor". [7]

3. Schwartz's dimensions of culture

Psychologists have proposed several models of culture dimensions. The present paper will describe Schwartz's structure of core human values. "The value concept … [Is] able to unify the apparently diverse interests of all the sciences concerned with human behavior" [8]. Schwartz represents a model based on 10 value types and four cultural dimensions. Each dimension consists of a subset of value types:

***Power* (PO):** Social status and prestige, control or dominance over people and resources.

Achievement (AC): Personal success through demonstrating competence according to social standards.

Hedonism (HE): Pleasure and sensuous gratification for oneself.

Stimulation (ST): Excitement, novelty, and challenge in life.

Self-Direction (SD): Independent thought and action-choosing, creating, exploring.

Universalism (UN): Understanding, appreciation, tolerance and protection for the welfare of all people and for nature.

Benevolence (BE): Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.

Tradition (TR): Respect, commitment and acceptance of the customs and ideas that traditional culture or religion provide the self.

Conformity (CO): Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms.

Security (SE): Safety, harmony and stability of society, of relationships, and of self [9].

The conflicts and congruities among all ten basic values yield an integrated structure of values. This structure can be summarized with two orthogonal dimensions. Self-enhancement vs. self-transcendence: On this dimension, both value types of self-enhancement emphasize pursuit of self-interests, whereas both value types of self-transcendence concern for the welfare and interests of others. Openness to change vs. conservation: On this dimension, self-direction and stimulation values emphasize independent action, thought and feeling as well as readiness for new experience, on the other hand, security, conformity and tradition values emphasize self-restriction, order and resistance to change.

Hedonism shares elements of both openness and self-enhancement [10]. In order to measure each value type, Schwartz used a special questionnaire which was based on measuring similarities of an individual's goals to his/her value types.

4. Portrait value questionnaire (PVQ)

The PVQ includes short verbal portraits of different people. Each portrait describes a person's goal or wishes which is important to him/her. For example "It is important to him to be rich. He wants to have a lot of money and expensive things" describes a person who cherishes power values [9]. For each portrait, respondents answer, "How much like you is this person?" They select one of these options: very much like me, like me, somewhat like me, a little like me, not like me, and not like me at all. Responses show value types of respondent that is important to him/her but he/she maybe does not necessarily exhibit the corresponding trait, for example, people may value creativity as a goal in life but may not be creative. And, some who are creative may attribute little importance to creativity as a value that guides them.

5. Values and behavior

One way to follow important values is to behave in approach that shows them or promotes them from other values. For example a person who concerned about personal safety has security value. Most behaviors can express more than one value. For instance, people who like adventure (stimulation value), love nature (universalism value) or want to obey his/her friend's opinion might go hiking. [11]

The question why people act according to their values is posed here. One answer for this question is Rokeach's idea. He believes that there is a need for consistency between one's beliefs (values) and actions. [8] Another is that values evaluate people's actions and help them to recognize what action is relevant to their values. Studies emphasize that people act according to their values. [12, 13]

In computer science and artificial intelligent in order to simulate human behavior agents should be able to act similarly to human. In our previous work we studied cultural values in intelligent agents and we have simulated cultural values in agents based on Schwartz 10 value type. [14] Now we plan to simulate human decision making according to these cultural values. Our intelligent cultural agents can make their decisions in each situation because of our proposed model is general. We are implementing this model in e-commerce field.

6. Proposed model (Generic Cultural Intelligent Agent: GCIA)

This study is based on Prelude to Cultural Software Agents: Cultural Backgrounds in Agent Simulation [14]. As we mentioned Agent's cultural values effect on its behavior. Proportional to importance of a value corresponding behavior with it is performed by Agent frequently. As well, values cause to form Agent's cultural dimension. Important behavior and cultural dimension determine Agent's criteria. Thus, attention to value's criteria make process of Agent's decision making similar to human. In our proposed model, we attempt to perform Agent's decision making based on cultural values.

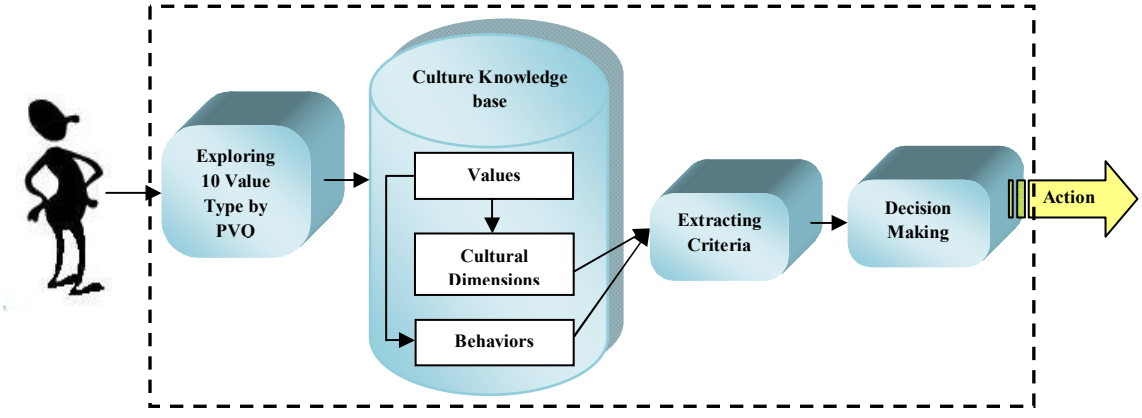


Fig. 1 Proposed Model (Generic Cultural Intelligent Agent: GCIA)

In our previous work, we explain that how 10 value type explore. Now, Agent performs its decision making based on values and its criteria. (Figure 1)

6.1. Extracting Agent's 10 value type based on Fuzzy logic

This Module is initialization Module. At first user answer to PVQ questionnaire and then its cultural values are calculated by fuzzy logic. For each question Respondent enter a number in [0 16] interval that shows how much this goal is important for him/her in life.(Figure 2) We use of average of answers of corresponding questions. [14]

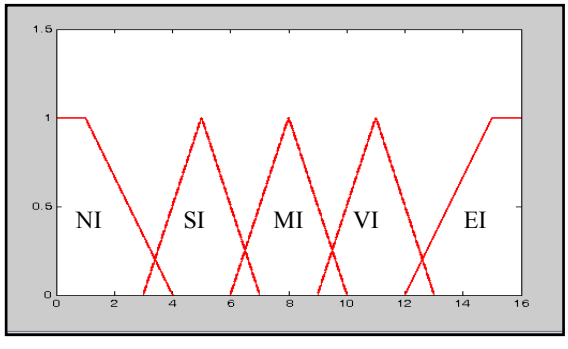


Fig. 2 Fuzzy Set for cultural values (NI: not important at all, SI: Slightly Important, MI: Moderately Important, VI: Very Important, EI: Extremely Important)

6.2. Calculating Agent's cultural dimension based on Agent's cultural values

Each dimension includes several values. We use maximum value of its corresponding values for calculating dimension. This number converts to a fuzzy value by a fuzzy set. We will use cultural dimension for extracting agent's criteria. In order to determine fuzzy set we use the first value for fuzzy interval of Low, the middle three values for Medium and the last one value for High. If we represent jth dimension by D_j, ith value type of jth dimension is represented by V_{ij}, and n is number of value types of each dimension, then the minimum amount of dimension is given by equation 1:

$$\text{Min}(D_j) = \text{Min}\{V_{ij}\} \quad (1)$$

And for maximum amount of dimension we will have equation 2:

$$\text{Max}(D_j) = \text{Max}\{V_{ij}\} \quad (2)$$

We suppose that there are three intervals for each dimension of culture denoted by Low, Medium and High. [14]

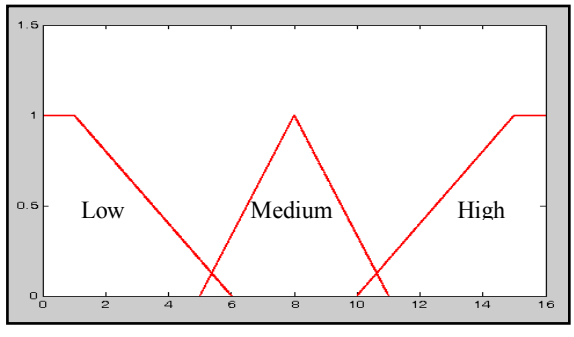


Fig. 3 Fuzzy Set for Importance Degree of Agent Behavior

6.3. Determining agent's degree of importance of behaviors according to values

Based on agent's values some behavior is performed by agent frequently. In order to determine the agent's behavior importance we use of mapping fuzzy sets. This fuzzy set has three numbers such as Low that is provided by NI, SI, Medium and High that is provided by VI, EI. This fuzzy set expresses importance degree of each behavior. (Figure 3)

6.4. Extracting criteria

Each agent according to cultural dimensions and importance degree of its behaviors selects some criteria for decision making. For instance, we are using this model in E-commerce. Suppose that seller agent has high self-enhancement dimension. As we showed, this dimension include of AC and PO values. We can refer to some values shopping behavior such as buying fashionable goods, buying expensive and high quality goods. According to, amount of these values agent determine importance degree of these behaviors. In addition, we should consider cultural dimension for extracting agent's criteria, because of we want to avoid that agent doesn't shopping based on single value.

$$Criteria = [C_1 \quad C_2 \quad \dots \quad C_n]$$

6.5. Decision making

Whereas, values are used as criteria for evaluating actions [11], thus agent compares each action with them. Cultural intelligent agent uses calculating Euclidean distance for its comparison.

$$Action\_param = [P1 \quad P2 \quad .. \quad Pn]$$

$$Dis = \sqrt{(C_1 - P_1)^2 + (C_2 - P_2)^2 + .. + (C_n - P_n)^2} \quad (3)$$

$$Action = Argmin\{ Dis \} \quad (4)$$

For example in shopping an agent with EI achievement value buy fashionable thing, so the fashion is a criteria or an agent with high Self-Enhancement buy high quality and expensive things and also fashion is important for it. So agent selects goods with high quality and high fashion parameters.

7. Implementation

Due to, being generic our model we are using it in e-commerce that we will present its results sooner. We simulate human shopping behavior by seller cultural agents. In our system there are some buyer and seller agents. Seller agents have cultural values according to users. Each buyer has goods with specific features. For example several buyer have goods with high quality and price, some of those have fashionable goods. We will see that each agent shop according to its cultural values. We use fuzzy J for performing sections 4.1, 4.2, 4.3. After extracting each agent behavior importance we use aglet for creating agent. These agents attend in a market place that there is some buyer there and they will shop according to their cultural behaviors. The following represents a few sample rules of cultural behavior in shopping:

Rule 1:

IF		Self-Enhancement	Is	H
	And	Buy_high_Quality_goods	Is	M
THEN		Quality_criteria=1		

Rule 2:

IF		Self-Enhancement	Is	H
	And	Buy_fashionable_goods	Is	M
THEN		Fashionable_criteria= 1		

Rule 3:

IF		conservation	Is	H
	And	Buy_low price_goods	Is	M
THEN		Quality_criteria=0 Low_price_criteria=1		

Rule 4:

IF		PO	Is	EI
	Or	PO	Is	VI
THEN	Quality_criteria=1			

Rule 5:

IF		TR	Is	EI
	Or	TR	Is	VI
	And	AC	Is	NI
	Or	AC	Is	SI
THEN	Fashionable_criteria=0			

Rule 6:

IF		AC	Is	EI
	Or	AC	Is	VI
THEN	Fashionable_criteria=1			

Rule 7:

IF		SE	Is	EI
	Or	SE	Is	VI
THEN	Buy_Internal_goods_criteria=1			

H: High, M: Medium

8. Conclusion and future work

In this paper we presented a model for agent's decision making based on cultural values. In some real world cases as well as in respective simulations, differences between values of individuals can have strong implications on their relationships. For example, in e-commerce each value can be effective in agent shopping. Thus, we are simulating a marketplace for considering these behaviors. Also, this model can apply for agent's judging. For example, each agent for judging about another agent it can compare its values with others. Use of learning can be countered as its future work. If agents can learn others values that it seems good values may be use for social modeling and showing agent's evaluation.

References

[1] G. Hofstede, G.J. Hofstede, *Cultures and Organizations: Software of the Mind*, Revised and expanded 2nd Edition, 436 pages. New York: McGraw- Hill USA, ISBN 0-07-143959-5, 2005

[2] K. Smith, I. Lindgren, and R. Granlund, "Bridging Cultural Barriers to Collaborative Decision Making in On-Site Operations Coordination Centers. Final Report to Räddningsverket. Report LiU - IEI - R-- 07/0002. Linköping University. Sweden, 2007.

[3] I. Samarah, S. Paul, P. Mykytyn, and P. Seetharaman, "The Collaborative Conflict Management Style and Cultural Diversity in DGSS Supported Fuzzy Tasks: An Experimental Investigation", In R.H. Sprague Jr. (ed.), Proceedings of the Thirty-Sixth Annual Hawaii Internaional Conference on Systems Sciences. Los Alamitos, CA: IEEE Computer Society Press, 2003.

[4] ANES-doc (2006). American National Election Studies Proposal on Basic Values. Posted on June 15,2006.ftp://ftp.electionstudies.org/ftp/anes/OC/2006pilot/msshasch.pdf (last visit: 2007-11-18).

[5] S.H. Schwartz, *Universals in the Content and Structure of Values: Theory and Empirical Tests in 20 Countries*. In M. Zanna (ed.), Advances in Experimental Social Psychology, 25:1-65. New York: Academic press, 1992.

[6] S.H. Schwartz, "Value Priorities and Behavior:Applying a Theory of Integrated Value Systems", In C. Seligman, J.M. Olson, and M.P. Zanna (eds.), The Psychology of values: the Ontario Symposium. 8:1-24. Hillsdale, NJ: Erlbaum, 1996.

[7] S.H. Schwartz, "Basic Personal Values", Report to the National Election Studies Board Based on the 2006 NES Pilot Study, 2007.

[8] M. Rokeach, *The Nature of Human Values*. New York: Free Press, 1973.

[9] S.H. Schwartz, G. Melech, S. Burgess, M. Harris, and V. Owens, "Extending the Cross-Cultural Validity of the Theory of Basic Human Values with a Different Method of Measurement, Journal of Cross- Cultural Psychology, Vol. 32 No.5, pp.519-42, 2001.

[10] S.H. Schwartz, "Basic Human Values: Theory, Methods, and Applications", The Hebrew University of Jerusalem, 2006.

[11] A. Bardi, S.H. Schwartz, "Values and behavior: Strength and structure of relations", Personality And Social Psychology Bulletin, *29,* 1207-1220. 2003

[12] N.T. Feather, "Values, valences, and choice: The influence of values on the perceived attractiveness and choice of alternatives", Journal of Personality and Social Psychology, 68, 1135-1151, 1995.

[13] S.H. Schwartz, L. Sagiv, "Identifying culture specifics in the content and structure of values", Journal of Cross-Cultural psychology, 26, 92-116, 1995.

[6] Z. Mazadi, N. Ghasem-Aghaee and T.I. Ören, "Prelude to Cultural Software Agents: Cultural Backgrounds in Agent Simulation", Agent Directed Simulation Conference (ADS'08), Ottawa, Canada, April 14- 17,2008.

[15] Hoefstede, G. *Hoefstede's Cultural Dimensions: http://www.geert-hofstede.com/* (last visit: 2007 November 01).

Appendix 1:

Table 1: Examples of Each Value Behavior Items [Bardi, Schwartz. 2003]

Values	Behavior Items
PO	Choose friends and relationships based on how much money they have.
AC	Study late into the night before exams even if I have studied consistently throughout the semester.
HE	Take it easy and relax.
ST	Do unconventional things.
SD	Examine the ideas behind rules and regulations before obeying them.
UN	Make sure everyone I know receives equal treatment, even if I don't personally like him/her.
BE	Lend things to people I know (e.g., class notes, books, milk).
TR	Show modesty with regard to my achievements and talents.
CO	Refrain from questioning an exam or project's grade even if I think it is unfair.
SE	Vote for candidates who are willing to legislate funds to keep nuclear weapons out of terrorists' hands.