

CENTRAL LIBRARY



SCHOLARLY ARTICLES

*A CURRENT AWARENESS BULLETIN
OF ARTICLES BY
FACULTY, STUDENTS AND ALUMNI*

~ **OCTOBER 2013** ~

DELHI TECHNOLOGICAL UNIVERSITY

(FORMERLY *DELHI COLLEGE OF ENGINEERING*)

GOVT. OF N.C.T. OF DELHI

SHAHBAD DAULATPUR, MAIN BAWANA ROAD

DELHI 110042

PREFACE

This is the Tenth Issue of Current Awareness Bulletin for the year 2013, started by Delhi Technological University Library. The aim of the bulletin is to compile, preserve and disseminate information published by the Faculty, Students and Alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of DTU as a whole to the academia. It contains information resources available in the internet in the form of articles, reports, presentation published in international journals, websites, etc. by the faculty and students of Delhi Technological University in the field of science and technology. The publication of Faculty and Students, which are not covered in this bulletin, may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose. To make the bulletin more comprehensive, the learned faculty and Students may provide their uncovered publication to the library either through email or in CD, etc.

This issue contains the information published during October 2013. The arrangement of the contents is alphabetical wise starting from A-Z. The Full text of the article, which is either subscribed by the University or available in the web, is provided in this Bulletin.

CONTENTS

1. A modified implementation of tristate inverter based static master-slave flip-flop with improved power-delay-area product ***Kunwar Singh**, Department of Electrical Engineering
2. A Neuro-Fuzzy Classifier for Website Quality Prediction ***Ruchika Malhotra**, *Department of Computer Engineering*
3. *A Variable Length Distributed Source Coding Algorithm for WSNs* K S Shivaprakasha and ***Muralidhar Kulkarni**, Electronics and Communication Engineering (ECE) Department
4. *An FRAM-Based Nonvolatile Logic MCU SoC Exhibiting 100% Digital State Retention at 0 V Achieving Zero Leakage With 400-ns Wakeup Time for ULP Applications* **£Sudhanshu Khanna**, *(Electronics and Communication Engineering)* Steven C. Bartling, Michael Clinton, Member, IEEE, Scott Summerfelt, John A. Rodriguez, Member, IEEE, and Hugh P. McAdams
5. Ant Colony Based Load Flow Optimisation Using Matlab ***Kapil Upamanyu**, ***Keshav Bansal** and ***Miteshwar Singh** Department of Electrical Engineering
6. *Application of Adaptive Neuro-Fuzzy Inference System for Predicting Software Change Proneness* ***Akshit Peer** and ***Ruchika Malhotra** *Department of Computer Engineering*
7. APPLICATION OF SUPPLY CHAIN MANAGEMENT IN A FOOD PROCESSING COMPANY Parvesh Antil and ***Anish Kumar**, ME Department
8. Development of reliability index for combined cycle power plant using graph theoretic approach Nikhil Dev, ***Samsher** (Department of Mechanical Engineering) S.S. Kachhwaha and Rajesh Attri

9. Effect of Indoor Air Pollution on Child Development: A Research Review ***Debapriyo Samanta, *Aditi Singh, *Tarun Joshi** and ***Rajeev Kumar Mishra**, Department of Environmental Engineering
10. Ensemble based Active Annotation for Biomedical Named Entity Recognition ***Mridula Verma** (Department of Computer Science Engineering), Utpal Sikdar, Sriparna Saha and Asif Ekbal
11. Evaluating the Ambiguity of Fuzzy Clustering with Fuzzy Entropy ***Seba Susan, *Saarthak Sankalp, *Prakhar Porwal, *Prasanna Kumar** and ***Simrat Singh**, Department of Information Technology,
12. Grey Relational Effort Analysis Technique using Regression Methods for Software Estimation ***Nagpal Geeta**, Department of Computer Science, **+Uddin Moin**, Faculty of Engineering and Technology and Kaur Arvinder
13. IJSE reviewers September 2013 reviewer acknowledgement ***Naveen Kumar**
14. Magnetotransport Behaviour of Nanocrystalline $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ ($0.40 \leq x \leq 0.60$) Neelam Maikhuri, ***Anurag Gaur**, (Department of Applied Physics) Vasudha Aggarwal, Umesh Gaur, and H. K. Singh
15. Novel Hybrid Spectrum Handoff for Cognitive Radio Networks Nisar A. Lala, **+Moin Uddin** (Faculty of Engineering and Technology) and N A Sheikh
16. Novel natural structure corrector of ApoE4 for checking Alzheimer's disease: Benefits from high throughput screening and molecular dynamics simulations Manisha Goyal, Sonam Grover, ***Jaspreet Kaur Dhanjal** (Department of Biotechnology), Sukriti Goyal, Chetna Tyagi, Sajeev Chacko and Abhinav Grover

- 17.** Prediction of Slot-Size and Inserted Air-Gap for Improving the Performance of Rectangular Microstrip Antennas using Artificial Neural Networks ***Taimoor Khan**, Member, IEEE, ***Asok De**, Senior Member, IEEE, Department of electronics and communication engineering, and **+Moinuddin** (Faculty of Engineering and Technology)
- 18.** STACKED PATCH ANTENNA WITH CROSS SLOT ELECTRONIC BAND GAP STRUCTURE ***N.S. Raghava, Asok De**, Dept. Of Electronics and Communication Engg., Nitish Kataria and Sarthak Chatterjee
- 19.** Visible Luminescence Characteristics of Sm³⁺ doped Zinc Alumino Bismuth Borate Glasses, K. Swapna, Sk. Mahamuda, ***A. Srinivasa Rao**, (Department of Applied Physics) T. Sasikala, L. Rama Moorthy
- 20.** Visible red, NIR and Mid-IR emission studies of Ho³⁺ doped Zinc Alumino Bismuth Borate glasses, Sk. Mahamuda, K. Swapna, P. Packiyaraj, ***A. Srinivasa Rao** (Department of Applied Physics), and G. Vijaya Prakash

+ Vice – Chancellor
*** Faculty**
@ Research Scholar
\$ PG Scholar
£ Undergraduate Student

A modified implementation of tristate inverter based static master-slave flip-flop with improved power-delay-area product

Kunwar Singh ^{a,*}

^a Department of Electrical Engineering,
Delhi Technological University, India

Postal Address: Room No. FW1-SF1, EED, DTU, New Delhi-110042

E-mail: kunwarsingh@dce.ac.in

Mobile: +919711177638

Satish Chandra Tiwari ^b

^b Division of ECE, Netaji Subhas Institute of Technology,
University of Delhi, India

Postal Address: Room No. 16, Division of ECE, NSIT, Sector-3, Dwarka, New Delhi-110078

E-mail: stiwari@cadence.com

Mobile: +919899267644

Maneesha Gupta ^c

^c Division of ECE, Netaji Subhas Institute of Technology,
University of Delhi, India

Postal Address: Room No. 18, Division of ECE, NSIT, Sector-3, Dwarka, New Delhi-110078

E-mail: maneeshapub@gmail.com

Mobile: +919810753962

* Corresponding author

ABSTRACT

The paper introduces novel architectures for implementation of fully static master-slave flip-flops for low power, high performance and high density. Based on the proposed structure, traditional C²MOS latch (tristate inverter/clocked inverter) based flip-flop is implemented with fewer transistors. The modified C²MOS based flip-flop mC²MOSff1 and mC²MOSff2 are realized using only sixteen transistors each while the number of clocked transistors is also reduced in case of mC²MOSff1. Post layout simulations indicate that mC²MOSff1 flip-flop shows 12.4% improvement in PDAP (power-delay-area product) when compared with transmission gate flip-flop (TGFF) at 16X capacitive load which is considered to be the best design alternative among the conventional master-slave flip-flops. To validate the correct behaviour of the proposed design an eight bit asynchronous counter is designed to layout level. LVS and parasitic extraction were carried out on Calibre whereas layouts were implemented using IC station (Mentor Graphics). HSPICE simulations were used to characterize the transient response of the flip-flop designs in a 180nm/1.8V CMOS technology. Simulations were also performed at 130nm, 90nm, and 65nm to reveal the scalability of both the designs at modern process nodes.

KEY WORDS: flip-flops; low power; high speed; power-delay-area product; VLSI; logical effort.

1. Introduction

Flip-flops are the key elements used in sequential digital systems. The appropriate selection of flip-flop topologies is instrumental in the design of VLSI integrated circuits such as microprocessors, microcontrollers and other high complexity chips. However, factors such as high performance, low power, transistor count, clock load, design robustness, power-delay and power-area tradeoffs are generally considered before choosing a particular flip-flop design. The highest operating frequency of clocked digital systems is determined by the flip-flops. Flip-flops and clock distribution network generally account for 30-70% of the total chip power consumption [1, 2]. Clock load is another major concern for digital system designers and several contributions have been reported in the past to reduce clock load and the associated power dissipation in the clocking network [3-5]. A design with elevated transistor count occupies larger area on chip and leads to an increase in the overall manufacturing cost. Hence, design and implementation of low power high performance flip-flops with least possible chip area is the main target of the modern chip manufacturing industry.

Flip-flops are broadly classified into three main categories viz. master-slave [6-11], pulse triggered [12-17] and differential flip-flops [18-21]. Among them, master-slave and pulse-triggered flip-flops are the most efficient in terms of power-delay product. Master-slave flip-flops exhibit positive (negative) set-up time (hold time) requirements and hence not suitable for high speed systems due to extended data to output delays. But they are power efficient and can be used in low power applications. However, their main limitation is less robustness to clock skew. Pulse-triggered flip-flops have negative set-up time and thus lead to smaller data to output delay. They exhibit inherent soft clock edge property which minimizes clock skew related cycle time loss.

A classification of master-slave flip-flops is further elaborated in Figure 1. Clock-gated topologies exhibit internal clock gating to suppress the power consumption at lower data switching activities based on a clock gating logic

and a comparator circuit. However, clock gated flip-flops have extended latency due to enhanced clock to output delays along with increased chip area overhead. Clock gated structures generally consume lesser power at low switching activities [22]. TGFF represents the best choice in the non-clock gated flip-flop category in terms of power-delay product [6] whereas existence of NMOS transistors in the critical path along with partially non-gated keepers leads to less significant power-delay tradeoff characteristics in case of write port master-slave flip-flop (WPMS) [7, 8] and pass transistor logic based flip-flop (PTLFF) [9].

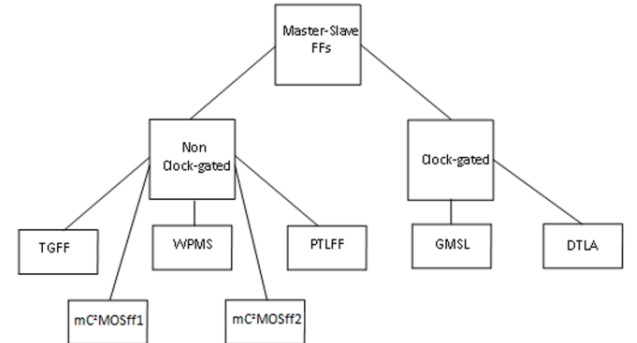


Figure 1. Classification of master-slave flip-flops.

In this paper, we introduce an alternative design approach for designing C²MOS based master-slave flip-flop, based on a new architecture with reduced transistor count and improved power-delay-area product. The proposed configurations mC²MOSff1 and mC²MOSff2 fall under the non-clock gated flip-flop category as shown in Figure 1.

The rest of the paper is organized as follows. Section 2 compares the conventional master-slave flip-flop configurations with proposed designs. Section 3 highlights the simulation parameters and test bench along with techniques used for transistor sizing and methodology adopted for optimization of timing and power-delay product. Section 4 describes the simulation results. Section 5 concludes the paper. An appendix is added to show calibration of parameters for delay calculations using LE

theory and to outline the strategy followed for designing the eight-bit ripple counter.

2. Overview of previous work and proposed designs

Figure 2 shows the conventional master-slave flip-flop architecture whereby two regenerative loops (L1 and L2) are present in the master and slave sections to account for a static functionality. Both the loops operate independent of each other on complementary clock signals. Regenerative loops are composed of cross coupled inverters. It can be observed from Figure 2 that for each loop regenerative action is achieved through one inversion in the forward (critical) path while the other (clocked) inversion takes place in the feedback path. Moreover, there is no common component between both the loops.

Since an inverter followed by transmission gate is equivalent to a clocked inverter, the combination is replaced by a clocked inverter to form a C²MOS based flip-flop architecture as shown in Figure 3 [23]. Two regenerative loops L3 and L4 are used in a similar manner as in the previous case to maintain the static nature of the flip-flop.

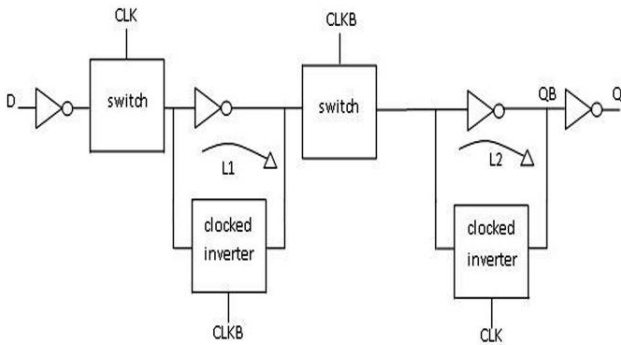


Figure 2. Conventional architecture with inverting latches in the feedback path.

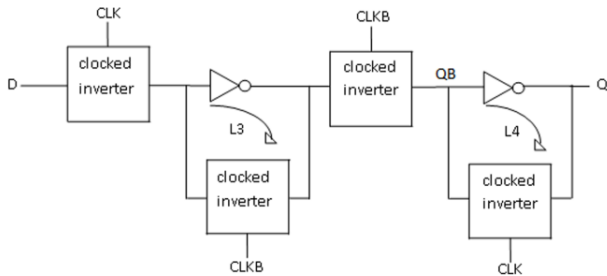
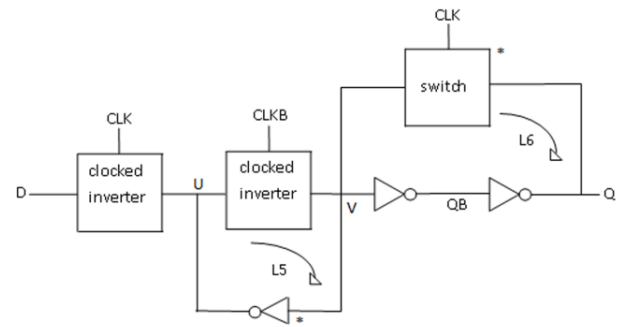


Figure 3. Conventional architecture with clocked switches in the feedback path.

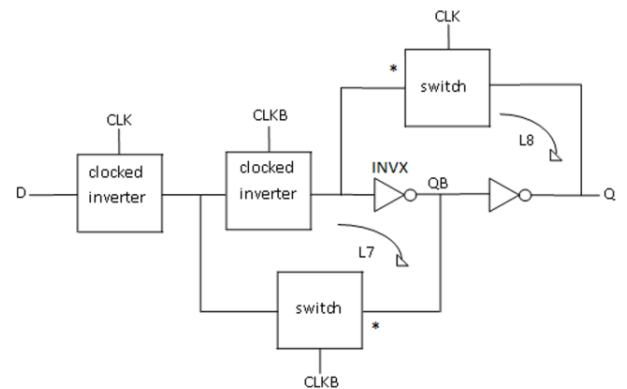
However, in the proposed architecture as reported in Figure 4 (a), both the inversions take place in the forward (critical) path and the loop is completed by a clocked switch for loop L6 while loop L5 is completed by using an inverter in the feedback path. It is clearly noticed from Figure 4(a) always driven and never floating thus ensuring a static flip-

flop operation. The size of transistors in the feedback path marked by asterisks (*) is kept at 360nm (minimum technology width) to eliminate race conditions at nodes U and V. Yet another implementation is shown in Figure 4 (b) which uses inverter INVX in the critical path and a clocked switch to form a regenerative loop L7. It is to be noted that INVX is common to both the regenerative loops L7 and L8 which is contrary to the realization of previous architectures.

Figure 5 represents the actual circuit design based on the proposed architectures in Figure 4, while TGFF is implemented using transmission gates as switches in the conventional architecture as demonstrated in Figure 6.

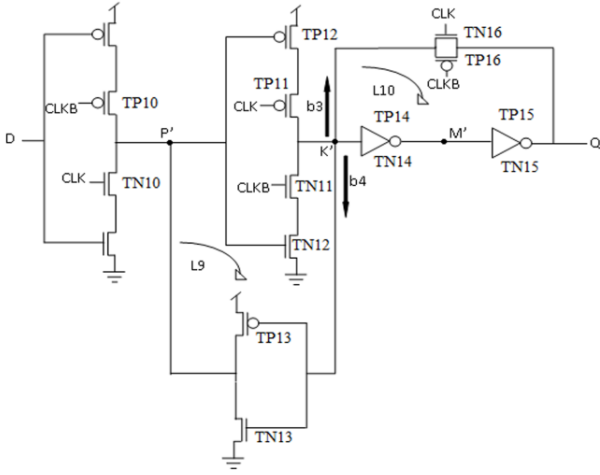


(a)

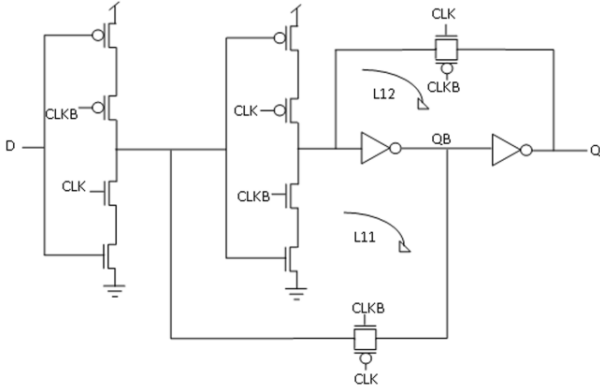


(b)

Figure 4. Proposed architectures.



(a) mC²MOSff1



(b) mC²MOSff2

Figure 5. Schematic diagrams of proposed designs.

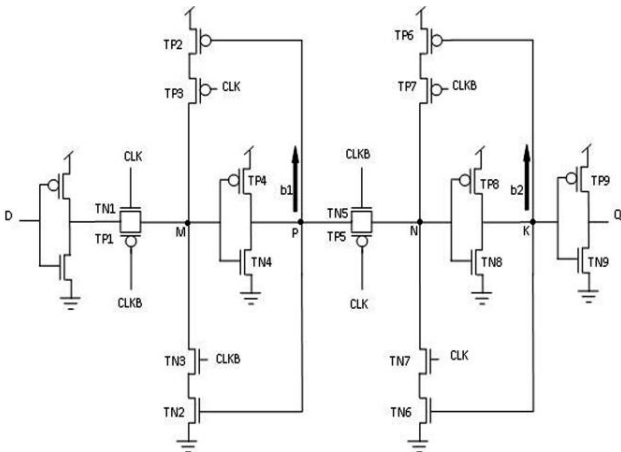


Figure 6. TGFF based on conventional architecture.

It can be clearly observed that mC²MOSff1 and mC²MOSff2 both are realized using sixteen transistors each. As a result, the area occupied by the proposed designs is significantly lesser than the conventional designs. Moreover, the number of clocked transistors in mC²MOSff1 is six as compared to eight in case of TGFF or conventional clocked inverter based flip-flop C²MOSff [23].

To illustrate the superior performance of the proposed flip-flop configurations, other flip-flop topologies viz. TGFF, WPMS, PTLFF, gated master-slave latch (GMSL) [10] and data transition look ahead flip-flop (DTLA) [11] belonging to the master-slave class have been used for comparisons. Out of the above mentioned topologies GMSL and DTLA represent flip-flops with internal clock gating. Schematic diagrams of WPMS, PTLFF, GMSL and DTLA are shown in Figures 7-10 respectively.

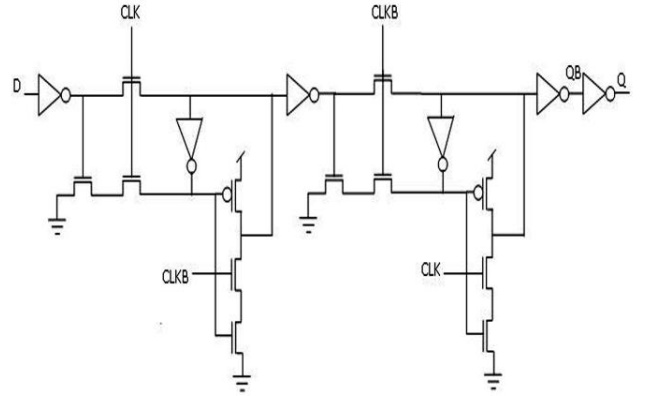


Figure 7. Schematic diagram for Write port master slave flip-flop (WPMS).

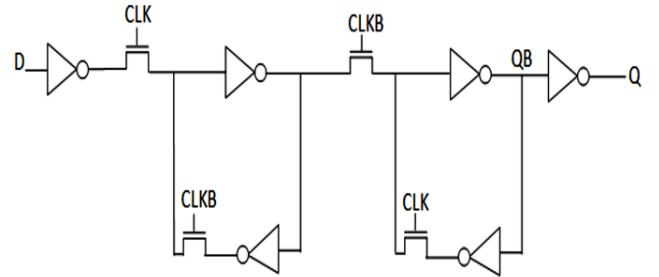


Figure 8. Schematic diagram for Pass transistor logic style flip-flop (PTLFF).

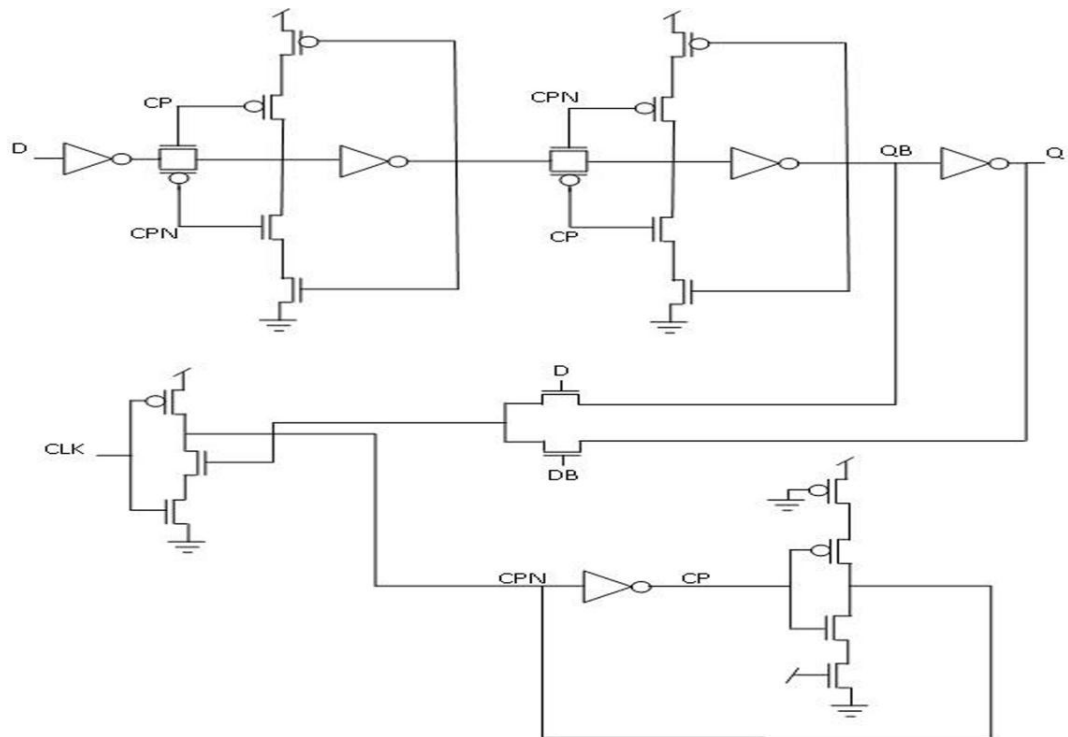


Figure 9. Schematic diagram for Gated master slave latch (GMSL).

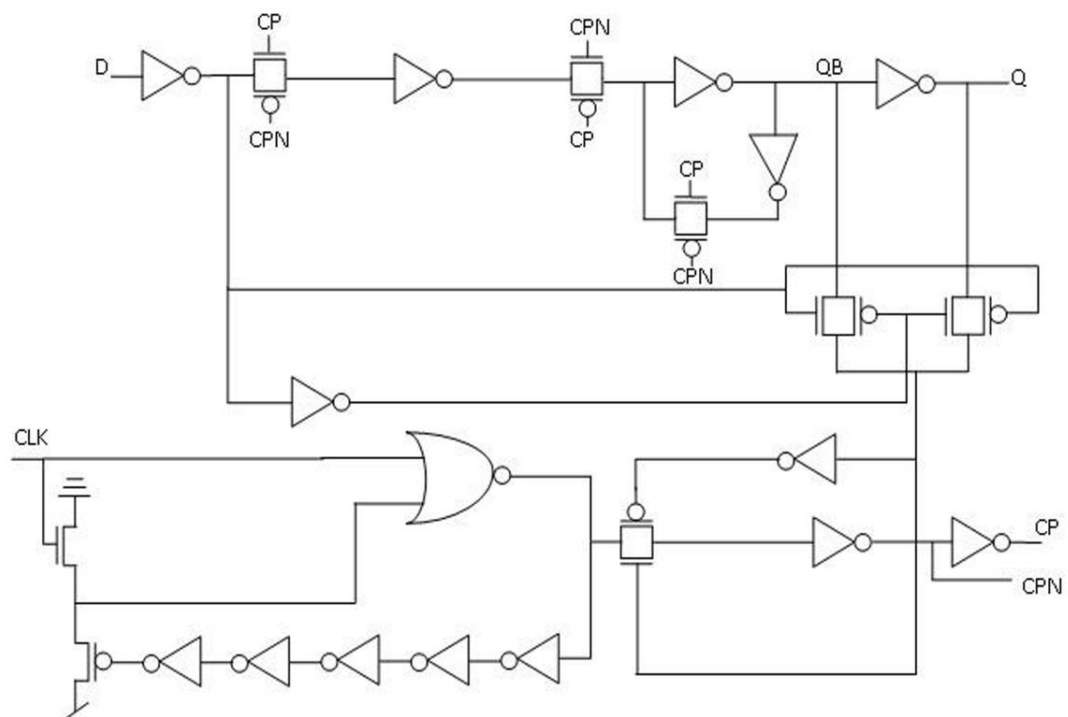


Figure 10. Schematic diagram for Data transition look ahead flip-flop (DTLA).

3. Simulation parameters, test bench and optimization methodology

Table 1 lists the CMOS parameters used for creating the simulation environment. The flip-flops were designed to layout level in 180nm/1.8V CMOS process at 250 MHz clock frequency. The width of transistors in the feedback structures were invariably fixed at the minimum value 360nm while the slope of the data and clock signals was kept

at 100ps. Performances of the various flip-flop configurations are evaluated through SPICE simulation of the circuits extracted from the layout with the inclusion of parasitics.

Table 1: Simulation parameters.

Parameter	Value
W_{\min}	360nm
L_{\min}	140nm
C_{\min}	1.24fF
V_{DD}	1.8V
Frequency	250MHz
Signal Slope	100ps

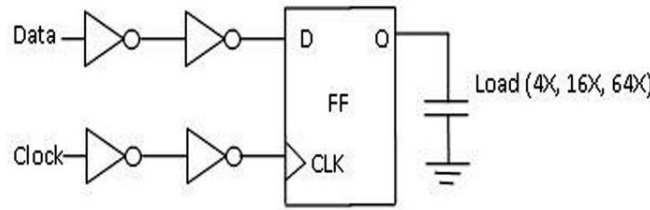


Figure 11. Simulation setup.

Figure 11 shows the simulation test bench for characterization and comparison of the FF designs [3]. The clock and data signals are fed to the flip-flop through a two stage buffer. Data-to-output delay ($T_{DQ,\min}$) is used for performance comparisons. Logical effort theory is extensively used for designing fast CMOS circuits based on pencil and paper calculations and is widely adopted in the literature [24]. Hence, the delay sensitivity factor introduced by Alioto et. al. [25] based on logical effort theory has been used for performance optimization.

A 16 cycle long pseudorandom sequence with a switching factor $\alpha=0.5$ is supplied at the data input for measurement of average power [26]. Since the delay and power characterization is strongly dependent on the capacitive load offered to FFs [27], varying capacitive loads $\{4, 16, 64\}C_{\min}$, where C_{\min} is the input capacitance of a symmetrical minimum inverter ($W_p = 2W_n = 2W_{\min}$) have been used to test the FF behaviour. Transistor sizing methodology adopted is the same as that in [28, 29] whereas power-delay product (PDP) and power-delay-area product (PDAP) are the chosen figures of merit (FOM).

The expression relating the absolute gate capacitance (C_{GATE}) in terms of fF (femtofarads) and absolute transistor width (W) in terms of nanometers (nm) obtained at 180nm process node by fitting simulation data [30] is given as

$$C_{GATE} = (1.15 \cdot 10^{-3}) \cdot W \quad (1)$$

LE method states that the optimized delay D of a path of N cascaded stages is

$$D = N\sqrt[N]{GBH} + P \quad (2)$$

$$D = N\sqrt[N]{F} + P \quad (3)$$

where G, B, H ($= C_L/C_{in}$), are the logical effort, branching effort, electrical effort while P, F($= GBH$) and C_L are parasitic delay, path effort and final load capacitance respectively.

$$D = P(1 + t) \quad (4)$$

From (2) and (4)

$$t = \frac{N\sqrt[N]{GBH}\sqrt[N]{C_L}}{P\sqrt[N]{C_{in}}} \quad (5)$$

where t represents the relative delay increment with respect to parasitic delay. Equation (4) and (5) indicate that larger values of C_{in} lead to a saturation in the optimized delay and

based on the above analysis, the delay sensitivity factor introduced by Alioto et. al. [25] is utilized to obtain the upper bound on the transistor widths for exploration of the power-delay design space with least computational effort.

$$S_D^{C_{in}} = \frac{\partial D}{\partial C_{in}} \frac{C_{in}}{D} = -\frac{1}{N} \frac{t}{t+1} \quad (6)$$

where $S_D^{C_{in}}$ is the delay sensitivity factor and is obtained from equations (3)-(5). The upper bounds on the normalized transistor widths w_i (normalized with respect to W_{min}) have been obtained such that the delay sensitivity remains under a minimum value S_{min} which is chosen as -5% for our analysis. The input capacitance C_{in} of the flip-flop is expressed in terms of normalized width $w1$ as shown below

$$C_{in} = (w1 \cdot 360 + 2 \cdot w1 \cdot 360) (1.15 \cdot 10^{-3}) \quad (7)$$

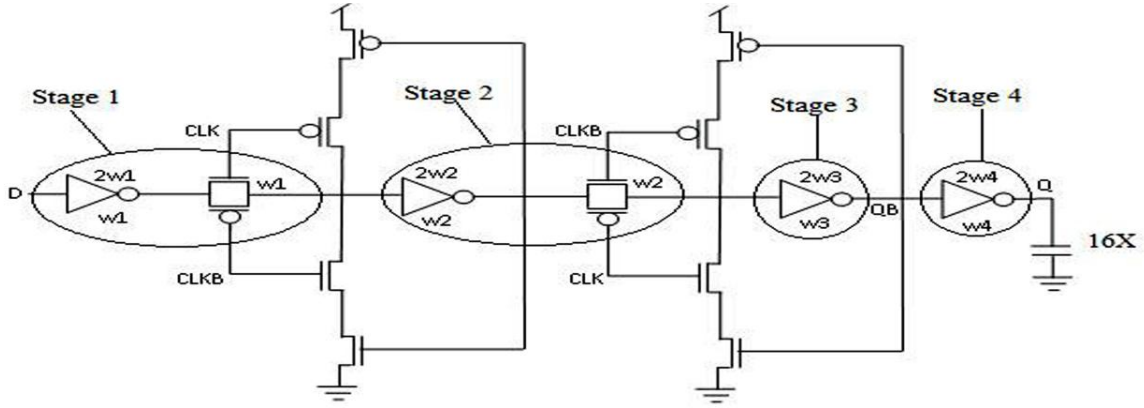


Figure 12. LE theory based transistor sizing methodology for transmission gate flip-flop.

Table 2: Traditional transmission gate flip-flop at 19.92fF load (16X).

C_{in} (fF)	w1	w2	w3	w4	$T_{DQ,min}$ (ps)	Power (uW)	PDP (fJ)
2.48	2	2.35	2.79	6.65	226	554	125.2
4.96	4	3.95	3.95	7.91	191	585	111.7
7.44	6	5.35	4.84	8.76	173	599	103.6
9.92	8	6.65	5.59	9.41	166	615	102
12.4	10	7.86	6.25	9.95	162	632	102.3
14.8	12	9.01	6.85	10.4	159	648	103
17.3	14	10.1	7.40	10.8	157	665	104.4
19.8	16	11.1	7.91	11.2	155	675	104.6
22.3	18	12.2	8.39	11.5	154	682	105
24.8	20	13.2	8.84	11.8	153	689	105.4

Table 3: Technology parameters used for estimation of capacitances.

Parameter	C_{gdo} (F/m)	C_{gso} (F/m)	C_{isw} (F/m)	C_i (F/m ²)	L_D (m)	L_S (m)
NMOS	2.78E-10	2.78E-10	7.9E-10	0.00365	31.6E-09	31.6E-09
PMOS	2.78E-10	2.78E-10	1.44E-9	0.00138	31.6E-09	31.6E-09

Figure 12 shows the conventional TGFF design. The sizing is done by assuming the transistors in the critical path to be independent design variables (IDVs) and optimizing for maximum performance using LE theory. The inverter before transmission gate in the first stage protects the input terminal from noise variations [31]. Table 2 exhibits delay variation for increasing C_{in} values. It is noteworthy that the delay saturates at 153ps for $C_{in} = 24.8$ fF. As a result, the upper bounds on transistor widths are exposed and the limits of power (energy)-delay design space are defined early in the design cycle [32]. The table also includes the corresponding power dissipation along with the power-delay product and it is observed that minimum power-delay product is obtained at $C_{in} = 9.92$ fF. The technology parameters used for capacitance calculations throughout this paper are listed in Table 3.

4. Results and discussion

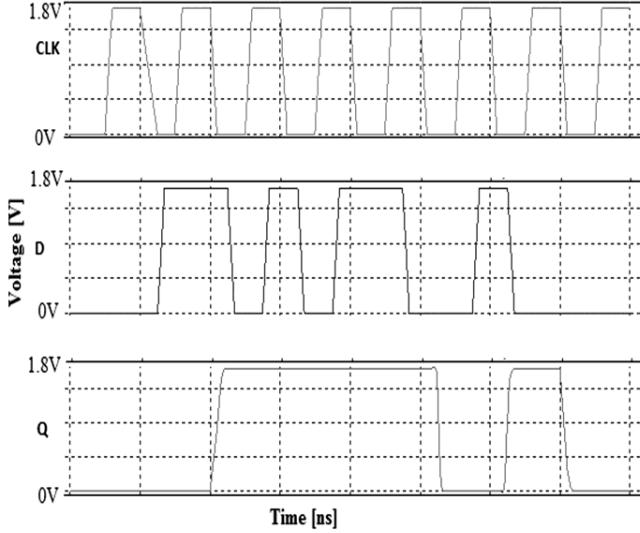


Figure 13. HSPICE simulation waveforms at 1GHz clock frequency for mC²MOSff1.

It is a well established fact that the conventional C²MOS although slower, is skew tolerant and occupies lesser area than TGFF [23, 33]. Moreover, mC²MOSff1 and mC²MOSff2 show nearly identical characteristics in terms of power, delay and area and hence only mC²MOSff1 is considered for comparisons.

The waveforms in Figure 13 represent the transient analysis of mC²MOSFF1 carried out over a period of 8 clock cycles. The SPICE simulation results verify the correct flip-flop operation at 1GHz clock frequency (all the flip-flops reported in the paper are designed for negative edge triggered operation). The variation of absolute data-to-output delays $T_{DQ,min}$ with FF input capacitance (C_{in}) for 16X (19.92fF) capacitive load is illustrated in Figure 14.

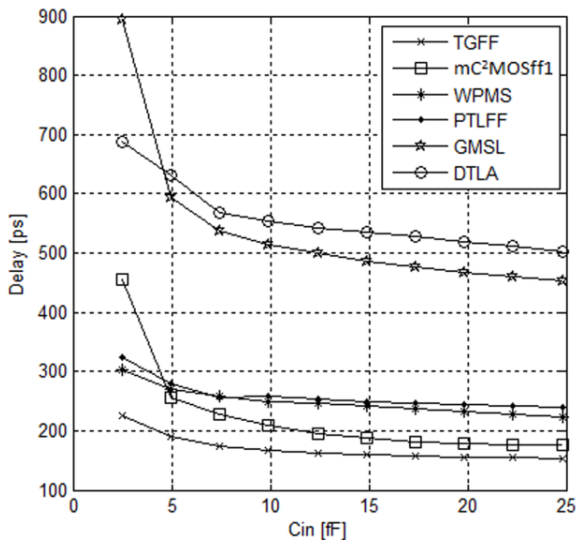


Figure 14. Variation in data-to-output delay with respect to FF input capacitance.

TGFF utilizes transmission gates in the critical path and hence faster than the rival designs. There are exactly the same number of stages in the critical path of TGFF and mC²MOSff1 the only difference being that the latching circuit in case of TGFF is an inverter followed by a clocked transmission gate (inverting latch) whereas a clocked/tristate inverter is present in mC²MOSff1. Logical effort of both the latches is considered to be two, however it is apparent that an inverter followed by a transmission gate is faster because the output node is driven by both the transistors of the transmission gate in parallel and this behaviour is reflected in Figure 14. From the above discussion, it is obvious that the value of logical effort for an inverting latch can be assumed to be two for most theoretical purposes, but for comparison with a C²MOS latch it must be slightly less than two if delays are to be modelled precisely.

Equation (2) clearly indicates that lesser branching effort leads to a faster circuit operation. The branching effort for a path with internal fan-out is expressed as [24]

$$b = \frac{C_{on-path} + C_{off-path}}{C_{on-path}} \quad (8)$$

where

$C_{on-path}$ represents the load capacitance along the path under analysis;

$C_{off-path}$ represents the capacitance of the connections that lead off the path.

The branching effort along the critical path is given as

$$B = \prod b_i$$

There are two branches each in TGFF and mC²MOSff1 represented as b1, b2 and b3, b4 in Figure 6 and Figure 4(a) respectively. The branching effort corresponding to branches b1, b2, b3 and b4 is calculated as follows.

4.1. Branching Effort in case of TGFF

b1 calculation:

$$C_{on-path} = C_{gd}(TN5) + C_{db}(TN5) + C_{gd}(TP5) + C_{db}(TP5) = 8.43fF \quad (9)$$

$$C_{off-path} = C_g(TN2) + C_g(TP2) = 1.12fF \quad (10)$$

$$b1 = 1.13$$

b2 calculation:

$$C_{on-path} = C_g(TN9) + C_g(TP9) = 12.33fF \quad (11)$$

$$C_{off-path} = C_g(TN6) + C_g(TP6) = 1.12fF \quad (12)$$

$$b2 = 1.09$$

$$B = b1 * b2 = 1.23$$

4.2. Branching Effort in case of $mC^2MOSff1$

b3 calculation:

$$C_{on-path} = C_g(TN14) + C_g(TP14) = 7.76fF \quad (13)$$

$$C_{off-path} = C_{gd}(TN16) + C_{db}(TN16) + C_{gd}(TP16) + C_{db}(TP16) = 1.47fF \quad (14)$$

$$b3 = 1.18$$

b4 calculation:

$$C_{on-path} = C_g(TN14) + C_g(TP14) = 7.76fF \quad (15)$$

$$C_{off-path} = C_g(TN13) + C_g(TP13) = 0.828fF \quad (16)$$

$$b4 = 1.10$$

$$B = b3 * b4 = 1.30$$

where

C_{gd} is gate to drain capacitance,
 C_{db} is drain to body capacitance,
and C_g is the gate capacitance of respective transistors.

Accordingly, using equation (2) and putting $G=4$, $B=1.23$, $H=19.92/12.4=1.60$, $N=4$ and $P=6$ we have $D=12.7$ (absolute delay 165.1ps) for TGFF whereas putting $G=4$, $B=1.30$, $H=19.92/12.4=1.60$, $N=4$ and $P=6$, $D=12.79$ (absolute delay 166.27ps) for $mC^2MOSff1$. Absolute delays D_{abs} are obtained by multiplying parameter D with parameter τ as shown in the equation below.

$$D_{abs} = D\tau \quad (17)$$

It is clearly observed that the delay of $mC^2MOSff1$ is marginally higher than the delay of TGFF. Now keeping other parameters to be same and assuming the logical effort of inverting latch to be 1.8, the updated value of TGFF is evaluated as $D=12.35$ (absolute delay 160.55ps).

The value of process dependent parameter τ is determined as approximately 13ps using the calibration technique as mentioned by Sutherland et. al. [24]. The detailed procedure is discussed in the Appendix. The absolute delay measurements obtained through simulation are 162ps for TGFF and 196ps for $mC^2MOSff1$ which is in close agreement with the theoretical values 160.55ps and 166.27ps respectively (typically within 15% error).

WPMS and PTLFF topologies show degraded performance due to the presence of pass transistors in the critical path while the speed of clock-gated structures is worst mainly because gating circuit is inserted between the clock and the flip-flop terminals which deteriorates the timing characteristics. The characterizations are done assuming $C_{in}=12.4fF$ and $C_L=19.92fF$ (16X) where C_L represents the flip-flop load capacitance.

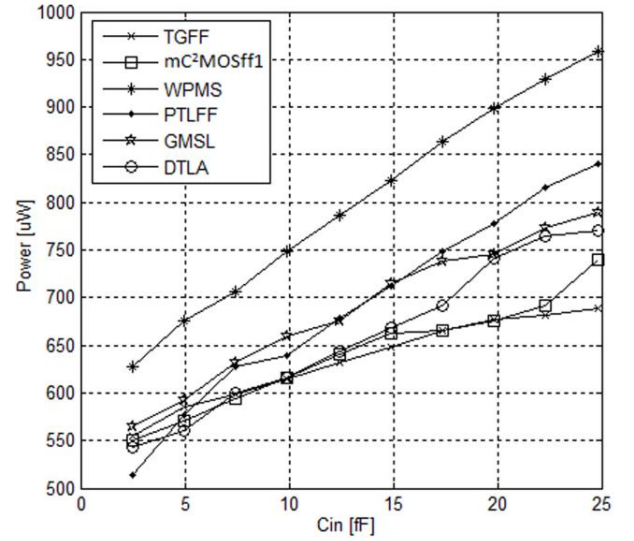


Figure 15. Variation in power dissipation as a function of FF input capacitance.

The variation of average power with C_{in} for 16X loading condition is depicted in Figure 15. Due to threshold voltage drop at internal nodes WPMS and PTLFF display worst power dissipation characteristics because of short circuit power dissipation. GMSL and DTLA exhibit greater power dissipation than non-gated counterparts because pseudorandom sequence has an activity factor of 0.5. The reason being the presence of additional comparator and clock gating circuit which is beneficial only at sufficiently low switching activities or otherwise leads to both increased area and power overhead.

4.3. Clock load calculations

TGFF:

$$\{C_g(TN1) + C_g(TP1) + C_g(TN5) + C_g(TP5)\} + \{C_g(TN3) + C_g(TP3) + C_g(TN7) + C_g(TP7)\} \quad (18)$$

{Transistors contributing towards clock load in the critical path} + {Transistors contributing towards clock load in the feedback structure}

$$= 14.78fF + 1.66fF$$

$$= 16.44fF$$

mC²MOSff1:

$$\{C_g(TN10) + C_g(TP10) + C_g(TN11) + C_g(TP11)\} + \{C_g(TN16) + C_g(TP16)\} \quad (19)$$

{Transistors contributing towards clock load in the critical path} + {Transistors contributing towards clock load in the feedback structure}

$$= 22.18\text{fF} + 0.84\text{fF}$$

$$= 23.02\text{fF}$$

Apart from the clock load, the capacitance value at internal nodes of mC²MOSff1 is reduced as compared to TGFF by eliminating transistors TN6 and TP6 from the feedback structure.

4.4. Capacitance calculations at internal nodes of TGFF

Internal Capacitance at nodes P and K:

$$\text{Node P: } C_g(TN2) + C_g(TP2) + C_{gd}(TN5) + C_{db}(TN5) + C_{gd}(TP5) + C_{db}(TP5) = 9.28\text{fF}$$

$$\text{Node K: } C_g(TN6) + C_g(TP6) + C_g(TN9) + C_g(TP9) = 9.02\text{fF}$$

Internal Capacitance at nodes M and N:

$$\text{Node M: } C_{gd}(TN1) + C_{db}(TN1) + C_{gd}(TP1) + C_{db}(TP1) + C_{gd}(TN3) + C_{db}(TN3) + C_{gd}(TP3) + C_{db}(TP3) + C_g(TN4) + C_g(TP4) = 18.41\text{fF}$$

$$\text{Node N: } C_{gd}(TN5) + C_{db}(TN5) + C_{gd}(TP5) + C_{db}(TP5) + C_{gd}(TN7) + C_{db}(TN7) + C_{gd}(TP7) + C_{db}(TP7) + C_g(TN8) + C_g(TP8) = 14.80\text{fF}$$

4.5. Capacitance calculations at internal nodes of mC²MOSff1

Internal Capacitance at nodes P' and K':

$$\text{Node P': } C_g(TN12) + C_g(TP12) = 9.76\text{fF}$$

$$\text{Node K': } C_g(TN13) + C_g(TP13) + C_g(TN14) + C_g(TP14) + C_{gd}(TN16) + C_{db}(TP16) + C_{gd}(TN16) + C_{db}(TP16) = 10.06\text{fF}$$

Internal Capacitance at node M':

$$\text{Node M': } C_g(TN15) + C_g(TP15) = 12.35\text{fF}$$

It can be easily concluded from calculations above that a total of 19.34fF capacitance has been reduced from the internal nodes in the critical path of mC²MOSff1 in comparison to TGFF. This leads to reduced internal power dissipation at these nodes as lesser capacitance has to be charged or discharged per clock cycle. However, reduction

in the clock load of mC²MOSff1 due to transistors eliminated from the feedback structure is nullified due to PMOS transistors TP10 and TP11 whose size is twice that of transistors TP1 and TP5 in case of TGFF and as a result the total power dissipation of both the flip-flops is nearly the same as can be clearly observed from Figure 16. Following a similar procedure, the clock load of various flip-flops is obtained and listed in Table 4 along with number of clocked transistors and power consumption values. It is seen that TGFF and mC²MOSff1 represent the most efficient designs in terms of reduced power consumption having power dissipation comparable to DTLA at C_{in}=12.4fF and C_L=19.92fF.

Table 4: Comparison of flip-flop parameters at $C_{in}=12.4fF$ and 16X capacitive loading.

Design	TGFF	mC ² MOSff1	WPMS	PTLFF	GMSL	DTLA
Transistor count	20	16	24	16	31	46
# of clocked transistors	8	6	6	4	2	3
Clock-to-output delay (ps)	92	116	206	204	419	683
Optimum setup time (ps)	70	80	40	50	80	-140
Hold time (ps)	-19	-21	-33	-32	-23	25
$T_{DQ,min}$ (ps)	162	196	246	254	499	543
Clock load (fF)	16.44	23.02	9.05	8.22	7.76	7.31
Power dissipation (uW)*	632	640	786	679	676	643
Leakage Power (uW)	59.38	57.51	72.64	69.83	74.91	76.73

* Pseudorandom sequence with $\alpha=0.5$ is used for power calculations.

It can be observed that mC²MOSff1 has the least transistor count along with PTLFF while GMSL and DTLA consist of maximum number of transistors. Since only sixteen transistors are used for circuit realization of mC²MOSff1, power dissipation is comparable to TGFF. It is worth noting that GMSL and DTLA offer minimum clock load, as a result these topologies exhibit least power dissipation at lower switching activities. The reason for extended clock-to-output delays of GMSL and DTLA is the insertion of clock gating circuitry while DTLA has a pulsed operation and hence shows negative set-up time requirements. Based on the power and delay measurements, power-delay product characteristics are derived for all the flip-flops as shown in Figure 16. The optimum power-delay product of gated structures, GMSL and DTLA is respectively 3.30x and 3.34x times greater than optimum PDP of TGFF. Among the non-clock gated structures, pass transistors based designs WPMS and PTLFF exhibit 1.77x and 1.57x enhancement in the power-delay product with respect to the benchmark flip-flop TGFF. TGFF also shows 20% improvement over mC²MOSff1 in terms of minimum power-delay product. However, despite the fact that TGFF represents a better alternative in terms of performance and optimum power-delay product, the area requirements also remain a major concern. It has been observed in the literature that conventional C²MOS based flip-flop is upto 20-25% more efficient in terms of occupied chip area. This stems mainly

from the fact that at layout level (i) in comparison to TGFF, diffusion areas of most of the transistors can be shared in C²MOS flip-flop [33] (ii) the number of contact holes can be reduced in the layout pattern [23] (iii) less complicated feedback structure leads to fewer interconnections.

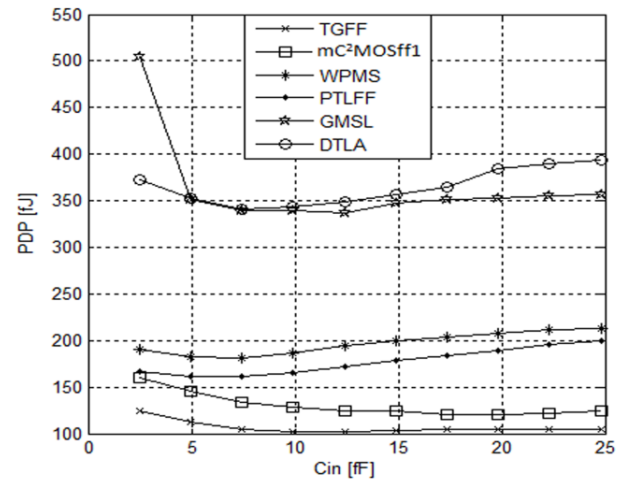


Figure 16. Power-delay product characteristics with varying FF input capacitance at 16X load.

Table 5: PDAP comparison of TGFF and mC²MOSff1.

Design	Transistor count	Transistor widths (um)	Delay (ps)	Power (uW)	Layout area (um ²)	PDP (fJ)	PDAP (fJ.um ²)
TGFF	20	52.52	162	632	175	102.3	17902
mC ² MOSff1	16	58.95	196	640	125	125.4	15675

The layouts were implemented using $C_{in} = 12.4\text{fF}$, indicating almost similar transistor sizes throughout the critical path with the exception of TP10 and TP11 belonging to mC²MOSff1 which are twice in size compared to TP1 and TP5 in accordance with the LE theory. The layouts for TGFF and mC²MOSff1 are shown in Figure 17 and Figure 18 respectively. Table 5 clearly shows that while TGFF is better in terms of PDP by 18.4%, mC²MOSff1 shows 12.4% improvement in the PDAP making it suitable for high density applications where performance can be compromised.

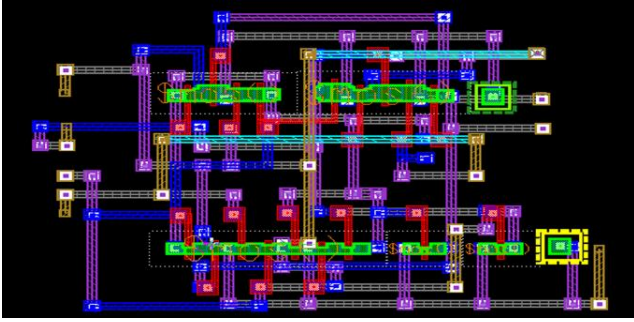


Figure 17. Layout implementation of TGFF.

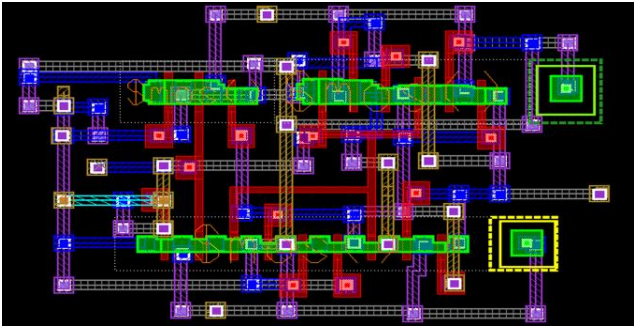


Figure 18. Layout implementation of mC²MOSff1.

The power dissipation results as illustrated in Figure 19 are obtained using $C_{in}=12.4\text{fF}$ which ensures that all the transistors in the critical path have similar widths. At zero switching activity clock-gated topologies are the most power efficient. GMSL and DTLA show GMSL 32.5% and 46.3% reduction in power in case of logic high at the input whereas for logic low, the power consumption is reduced by 19.2% and 35.4% respectively. Again, it can be clearly observed

that there is only a slight difference in the power dissipation of TGFF and mC²MOSff1 at different switching activities.

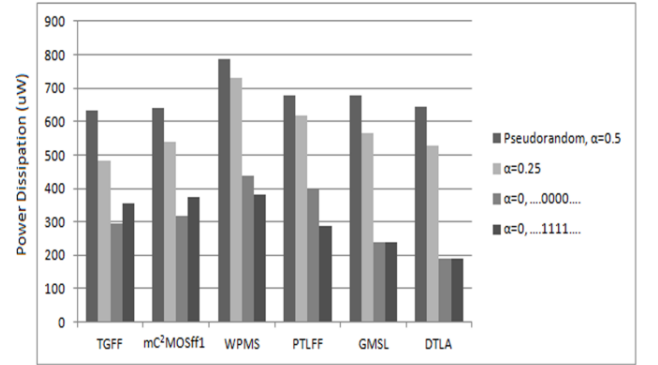


Figure 19. Comparison chart of power dissipation at different switching frequencies.

The correct functionality of the proposed flip-flop mC²MOSff1 is validated by designing an 8-bit ripple counter at 16X capacitive load and the average power measurements were carried out over 256 clock cycles. It was noticed that the power consumption of the mC²MOSff1 based counter is comparable to the TGFF at varying frequencies. Again, LE theory has been adopted for sizing individual flip-flops in each counter for optimum performance which is expressed in detail in the Appendix.

The flip-flops were also designed and simulated to layout level with inclusion of parasitics at 130nm, 90nm and 65nm CMOS processes to address scalability issues at more advanced process nodes. The simulation test bench and optimization methodology are similar as mentioned in Section 3. PVT variations are emphasized to evaluate the performance of flip-flops at all process corners viz. FF, SS, FS and SF with voltages scaled from 0.9-1.1V while the temperatures varied from 0-125 degrees as shown in Table 6. The simulation and technology parameters are also listed in Table 6 where C_G represents the capacitance per unit gate oxide and was evaluated to be $1.3\text{fF}/\mu\text{m}$ by fitting simulation data. In addition, the capacitances per unit length of poly, metal1 and metal2 interconnects are also mentioned.

Table 6: Flip-flop simulation parameters at 65nm CMOS technology.

Process corner	Temperature (°C)	V _{DD}	Simulation/Technology Parameters				
TT	70	1	L _{min}	W _{min}	C _{min}	Frequency	Signal slope
FF	0	1.1	60nm	120nm	507aF	2GHz	20ps
SS	125	0.9					
FS	70	1	C _{Poly} = 0.268C _G				
SF	70	1	C _{metal1} = 0.215C _G C _{metal2} = 0.175C _G				

For illustration purposes, the delay and power variations with the flip-flop input capacitance with respect to different process corners at 65nm CMOS technology for mC²MOSff1 are demonstrated in Figure 20 and Figure 21 respectively at 16X capacitive loading. Both mC²MOSff1 and mC²MOSff2 showed correct circuital behaviour at the aforementioned

process nodes which indicates that no internal noise violations exist especially due to the fact that logic levels are retained even at FF process corner. However, it is to be pointed out that mC²MOSff1 in a manner similar to TGFF starts to fail at SS corner for lower values of C_{in} [34].

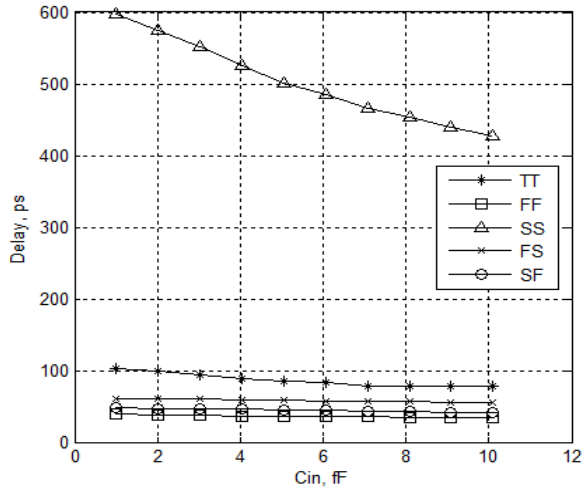


Figure 20. Delay variations for mC²MOSff1 at 16X loading for different process corners.

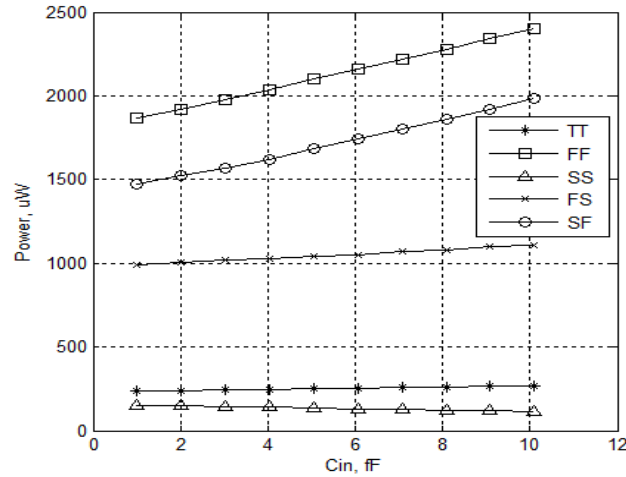


Figure 21. Delay variations for mC²MOSff1 at 16X loading for different process corners.

5. Conclusion

In this paper an alternative architecture for designing C²MOS based flip-flops is presented with a modified feedback strategy while preserving the fully static operation. Using the new feedback approach, a modified topology mC²MOSff1 is proposed with decreased parasitic capacitances at internal nodes in comparison to the TGFF which is the finest design in terms of PDP. However, post layout simulations and analyses indicate that the modified configuration mC²MOSff1 presents the best alternative in terms of PDAP among all the conventional designs. Therefore, for high performance applications TGFF still remains the best choice but it can be replaced by mC²MOSff1 for high density applications. Comparisons were carried out with state-of-the-art flip-flops in the master-slave class. The simulation results are well supported with mathematical analysis based on logical effort theory within acceptable error (typically less than 15%).

Appendix

A.1. Delay calibration using LE theory

For modelling delays using LE theory initially all the delays are expressed in terms of a basic delay unit τ which is process dependent such that the absolute delay is represented as the product of a unit less delay of the gate as shown in equation (2), and the delay unit τ . Accordingly,

$$D_{abs} = D\tau$$

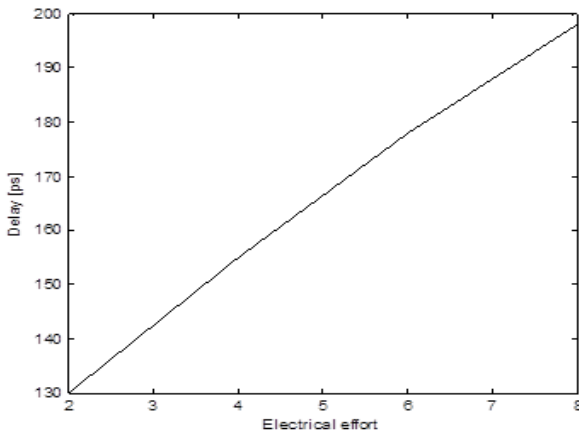


Figure 22. Delay vs. fanout curve for an inverter at 180nm/1.8V CMOS process.

While D represents the delay for a multistage path, d corresponds to the delay of a single stage logic gate. Parameter τ needs to be estimated in order to obtain absolute delays and accordingly a delay vs fanout curve is determined for an inverter as shown in Figure 22 by fitting simulation data. The curve is approximated as a straight line and the slope of the line represents τ since $d = (gh + p) \tau$ and logical effort of an inverter is 1. In our case τ is estimated as 13ps.

A.2. Implementation of 8-bit ripple counter

An 8-bit asynchronous counter was implemented by converting the D flip-flop configuration to a T flip-flop configuration using an EXOR gate as illustrated in the Figure 23 below.

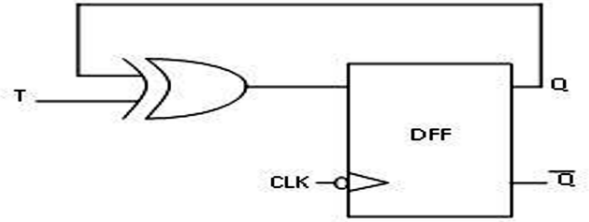


Figure 23. Conversion of D flip-flop to T flip-flop.

The T flip-flop designed using TGFF is shown in Figure 24. It is considered to be a five stage design and optimized for highest speed using LE theory. The EXOR gate was realized using transmission gates as revealed in Stage 1 of Figure 24. A similar procedure was followed for designing mC²MOSff1 based T flip-flop.

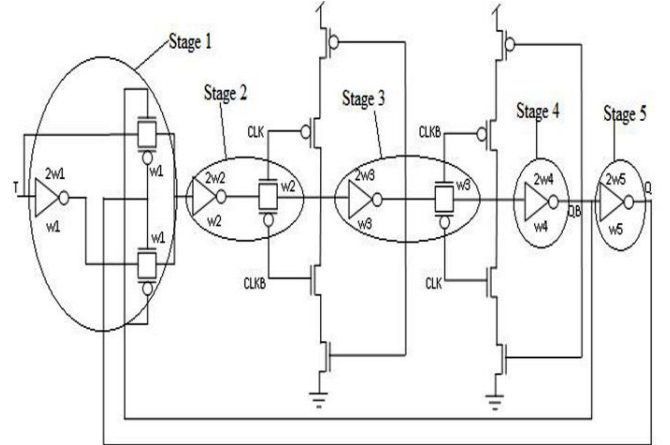


Figure 24. TGFF based T flip-flop.

For designing the modulo 256 counter, the output Q of each stage is connected to the clock terminal of the next stage through two intermediate inverters (acting as a buffer) sized ($W_p=11.52u$, $W_n=5.76u$) such that the input capacitance of the first inverter acts as the load capacitance for the flip-flop configuration of the previous stage as depicted in Figure 25. As a result, the load at the output terminal of each flip-flop is uniformly fixed at 19.92fF.

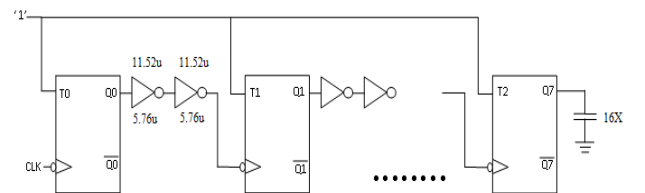


Figure 25. Schematic diagram of a modulo 256 ripple counter with intermediate buffers.

References

- [1] H. Kawaguchi, T. Sakurai, "A reduced clock swing flip-flop (RCSFF) for 63% power reduction," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 5, pp. 807-811, 1998.
- [2] G. Yeap, Practical Low power Digital VLSI Design, Kluwer Academic Publishers, 1998.
- [3] V. Oklobdzija, V. Stojanovic, D. Markovic, N. Nedovic, Digital System Clocking: High-Performance and Low-Power Aspects, Wiley-IEEE Press, 2003.
- [4] B. Mesgarzadeh, M. Hansson, A. Alvandpour, "Jitter Characteristic in Charge Recovery Resonant Clock Distribution," *IEEE Journal of Solid-state Circuits*, vol. 42, no. 7, pp. 1618-1625, 2007.
- [5] C. Giacomotto, N. Nedovic, V. Oklobdzija, "The effect of the system specification on the optimal selection of clocked storage elements," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 6, pp. 1392-1404, 2007.
- [6] G. Gerosa, S. Gary, C. Dietz, P. Dac, K. Hoover, J. Alvarez, H. Sanchez, P. Ippolito, N. Tai, S. Litch, J. Eno, J. Golab, N. Vanderschaaf, J. Kahle, "A 2.2 W, 80 MHz superscalar RISC microprocessor," *IEEE Journal of Solid State Circuits*, vol. 29, no. 12, pp. 1440-1454, 1994.
- [7] D. Markovic, J. Tschanz, V. De, Transmission-gate based flip-flop. US Patent 6642765, Nov. 4, 2003.
- [8] S. Hsu, S. Mathew, M. Anders, B. Bloechel, R. Krishnamurthy, S. Borkar, "A 110 GOPS/W 16-bit multiplier and reconfigurable PLA loop in 90-nm CMOS," *IEEE Journal of Solid State Circuits*, vol. 41, no. 1, pp. 256-264, 2006.
- [9] R. Hossain, L. D. Wronski, A. Albicki, "Low power design using double edge triggered flip-flops," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 2, no. 2, pp. 261-265, 1994.
- [10] A. Strollo, E. Napoli, D. Caro, "Low-power flip-flops with reliable clock gating," *Microelectronics Journal*, vol. 32, no. 1, pp. 21-28, 2001.
- [11] M. Nogawa, Y. Ohtomo, "A data-transition look-ahead DFF circuit for statistical reduction in power consumption," *IEEE Journal of Solid State Circuits*, vol. 33, no. 5, pp. 702-706, 1998.
- [12] F. Klass, C. Amir, A. Das, K. Aingaran, C. Truong, R. Wang, A. Mehta, R. Heald, G. Yee, "A new family of semidynamic and dynamic flip-flops with embedded logic for high-performance processors," *IEEE Journal of Solid State Circuits*, vol. 34, no. 5, pp. 712-716, 1999.
- [13] P. Zhao, T. Darwish, M. Bayoumi, "Low power and high speed explicit-pulsed flip-flops," *Proceedings of IEEE Midwest Symposium on Circuits and Systems*, vol. 2, pp. 477-480, 2002.
- [14] H. Partovi, R. Burd, U. Salim, F. Weber, L. DiGregorio, D. Draper, "Flow-through latch and edge-triggered flip-flop hybrid elements," *Proceedings of IEEE International Solid State Circuits Conference*, pp. 138-139, 1996.
- [15] R. Heald, K. Aingaran, C. Amir, M. Ang, M. Boland, P. Dixit, G. Gouldsberry, D. Greenley, J. Grinberg, J. Hart, T. Horel, W. Hsu, J. Kaku, C. Kim, S. Kim, F. Klass, H. Kwan, G. Lauterbach, R. Lo, H. McIntyre, A. Mehta, D. Murata, S. Nguyen, Y. Pai, S. Patel, K. Shin, K. Tam, S. Vishwanathaiah, J. Wu, G. Yee, E. You, "A third generation SPARC V9 64-b microprocessor," *IEEE Journal of Solid State Circuits*, vol. 35, no. 11, pp. 1526-1538, 2000.
- [16] N. Nedovic, M. Aleksic, V. Oklobdzija, "Conditional techniques for low power consumption flip-flops," *Proceedings of IEEE International Conference on Electronics, Circuits and Systems*, vol. 2, pp. 803-806, 2001.
- [17] S. Naffziger, G. Colon-Bonet, T. Fischer, R. Riedlinger, T. Sullivan, T. Grutkowski, "The implementation of the Itanium 2 microprocessor," *IEEE Journal of Solid State Circuits*, vol. 37, no. 11, pp. 1448-1460, 2002.
- [18] B. Kong, S. Kim, Y. Jun, "Conditional-capture flip-flop for statistical power reduction," *IEEE Journal of Solid State Circuits*, vol. 36, no. 8, pp. 1263-1271, 2001.
- [19] S. Shin, B. Kong, "Variable sampling window flip-flops for low power high-speed VLSI," *IEEE Proceedings of Circuits, Devices and Systems*, vol. 152, no. 3, pp. 266-271, 2005.
- [20] B. Nikolic, V. Stojanovic, V. Oklobdzija, W. Jia, J. Chiu, M. Leung, "Improved sense-amplifier-based flip-flop: Design and measurements," *IEEE Journal of Solid State Circuits*, vol. 35, no. 6, pp. 876-884, 2000.
- [21] N. Nedovic, V. Oklobdzija, W. Walker, "A clock skew absorbing flip-flop," *Proceedings of IEEE International Conference on Solid State Circuits*, vol. 1, pp. 342-344, 2003.
- [22] A. Strollo, D. Caro, "Low power flip-flop with clock gating on master and slave latches," *Electronic Letters*, vol. 36, no. 4, pp. 294-295, 2000.
- [23] Y. Suzuki, K. Odagawa, T. Abe, "Clocked CMOS Calculator Circuitry," *IEEE Journal of Solid-State Circuits*, vol. 8, no. 6, pp. 462-469, 1973.
- [24] I. Sutherland, B. Sproull, D. Harris, Logical Effort: Designing Fast CMOS Circuits. Morgan Kaufmann: Los Altos, CA, 1998.
- [25] M. Alioto, E. Consoli, G. Palumbo, "General strategies to design nanometer flip-flops in the energy-delay space," *IEEE Transactions on Circuits and Systems—Part I*, vol. 57, no. 7, pp. 1583-1596, 2010.
- [26] V. Stojanovic, V. Oklobdzija, "Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems," *IEEE Journal of Solid-state Circuits*, vol. 34, no. 4, pp. 536-548, 1999.
- [27] S. Heo, K. Asanovic, "Load-sensitive flip-flop characterization," *Proceedings of IEEE Computer Society Workshop on VLSI*, pp. 87-92, 2001.
- [28] M. Alioto, E. Consoli, G. Palumbo, "Analysis and comparison in the energy-delay-area domain of nanometer CMOS flip-flops: part I—methodologies and design strategies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 5, pp. 725-736, 2011.
- [29] M. Alioto, E. Consoli, G. Palumbo, "Analysis and comparison in the energy-delay-area domain of nanometer CMOS flip-flops: part II—results and figures

- of merit,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 19 (2011), 737-750.
- [30] G. Palumbo, M. Pennisi, “Design guidelines for high-speed transmission-gate latches : Analysis & comparison,” *IEEE International Conference on Electronics, Circuits and Systems*, (2008), 145-148.
- [31] E. Consoli, G. Palumbo, M Pennisi, Reconsidering high-speed design criteria for transmission-gate-based master-slave flip-flops,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 5, pp. 284-295, 2011.
- [32] M. Alioto, E. Consoli, G. Palumbo, “From energy-delay metrics to constraints on the design of digital circuits,” *International Journal of Circuit Theory and Applications*, 40(2012), 815-834.
- [33] J. Chao, C. A. Johnston, “ Behaviour analysis of CMOS D flip-flops,” *IEEE Journal of Solid State Circuits*, vol. 24, no. 5, pp. 1454-1458, 1989.
- [34] H.Q. Dao, K. Nowka, and V. Oklobdzija, “Analysis of clocked timing elements for dynamic voltage scaling effects over process parameter variation” *IEEE International Symposium on Low Power Electronic Design*, pp. 56-59, 2001.

A Neuro-Fuzzy Classifier for Website Quality Prediction

Ruchika Malhotra
Delhi Technological University
Delhi, India
ruchikamalhotra2004@yahoo.com

Anjali Sharma
CSIR-National Physical Laboratory
New Delhi, India
anjali@nplindia.org

To improve the quality of websites, it is necessary to continually assess and evaluate web metrics and subsequently make improvements. In this research, we have computed nine quantitative web measures for each website using an automated Web Metrics Analyzer tool developed in JAVA programming language. The website quality prediction models are developed utilizing ANFIS-Subtractive clustering and ANFIS-FCM based FIS models, to classify the quality of website as good or bad. The models are validated using 10 cross validation on a collection of web pages of Pixel Awards web metrics collected through the tool. The results are analyzed using Area Under Curve obtained from Receiver Operating Characteristic (ROC) analysis. The results showed that both ANFIS-Subtractive and ANFIS-FCM have acceptable performance in terms of specificity and sensitivity. In addition, ANFIS-Subtractive and ANFIS-FCM clearly induces only two rules, which are much less than 512 rules generated by the normal ANFIS model. Hence ANFIS-Subtractive and ANFIS-FCM are the most comprehensible techniques tested in this work.

Keywords: *Web Page, Web Page Metrics, Website Quality, Neuro-Fuzzy, Machine Learning Techniques.*

I. INTRODUCTION

Web based applications (WAs) are the multi tier, distributed applications accessed through client side browsers in the heterogeneous environment by unlimited number of users with varied experience. While the complexity of WAs increases each day, the resources and testing times decrease [24]. As WAs become complex, there is a growing concern about their quality. Evaluating website quality is essential, but there are few ways to analyze and evaluate the quality of the website in quantitative form [4].

Web metrics are considered to be the important source of decision and are useful for estimating the quality of web engineering product [13]. As the websites are updated to cater the heterogeneous demands of users, there is a growing demand in the industry to identify the effect of additions and frequent modifications of web pages on the quality of the website.

In this research, we have developed adaptive Neuro fuzzy inference models based on subtractive clustering and fuzzy c-

means (FCM) to predict the quality of the websites. A comparative analysis and performance assessment of the website models have been developed to classify the quality of website as good or bad. The website is good or bad based on the judges rating of Pixel Awards for a specific website. The panel of judges evaluates each site against innovation, content, navigation, visual design, functionality and overall experience. Each site is reviewed by at least 10 judges and the list of nominees is created. In our research we have considered the award winning websites as good.

To create the website quality prediction models, nine quantitative web measures from different categories of Pixel Awards website are computed by the Web Metrics Analyzer tool developed in JAVA programming language.

The rest of the paper is organized as follows. Section II describes related work. Section III discusses our methodology. Section IV focuses on the data preprocessing, the evaluation metrics and model building. Section V describes the analysis of the results. Section VI draws the Conclusion.

II. RELATED WORK

In various research work carried out in the field of evaluating web page quality, most quantitative methods rely on statistically analyzing the user behavior in the server logs [6, 8, 11]. All the data that are required to identify usability problems is interpreted using traffic-based analysis which computes the number of pages per visitor or number of visitors per page and time-based analysis which computes page-view durations. This kind of analyzing data produces uncertainties since there are incomplete records of user patterns provided by the server logs. Another method proposed by G. Velayathan and S. Yamada for evaluating web pages of user interest automatically investigates various factors related to a user's browsing behavior such as number of scrolls, form input, search text etc. [16].

Many inspection-based approaches were proposed that depends on evaluating the static HTML based on fixed guidelines [12]. For example, J. Scholtz et.al [12] developed WebSAT (Web Static Analyzer Tool) which is used to check the accessibility issues i.e. whether there is a support for disabled users,

maintainability issues and other important issues of web pages. H. Thimbleby [15] compared quantitative web page to thresholds using a web authoring tool called Gentler. However, there are no clear thresholds established or proposed in guidelines for a large class of web page measures.

The most closely related work is done in Ivory et.al [15, 16] which provides an introductory analysis of a collection of web pages and also captures various web measures associated with the rated websites. However, this work does not apply various machine learning algorithms to predict the best suited model that can provide high accuracy.

The recent work related to our study is that of Mittal et.al [25] in which quality of a website is analyzed by applying a fuzzy technique on web metrics. However the study provides the formula for quality prediction using selected eleven metrics. The results are not validated and the study is not aimed to classify the website as good or bad.

III. RESEARCH METHODOLOGY

Our methodology employs the quantitative web page attributes (number of links, words etc.) to compare the goodness of the web pages and to construct a quality prediction model utilizing Subtractive and FCM clustering model for predicting the class of website as good or bad. Fig. 1 shows the flowchart of methodology.

The flowchart of methodology shows the 5 sections. The first section is an Empirical Data Collection which involves 2 steps. The first step is to select the websites nominated in different categories for the particular year from Pixel Awards for which metrics estimations are to be calculated. The next step is to use a crawler to download and store the web pages belonging to all these websites.

The second section is the web metrics analyzer which includes 2 modules: web page processor to preprocess the source files to make them suitable for metric calculation and metric calculator which calculate the metric values. The metrics selected for the present study are described in Table I.

The third section is the data processing, which includes techniques such as data preprocessing, data import, data transformation and dividing data into training and checking subsets.

The fourth section is the model building, which includes initialization of fuzzy systems, application of subtractive clustering and FCM of the FIS and training of the model. The model is validated using 10-cross validation technique.

The fifth section is the data result analysis which includes the analysis and comparison of data using machine learning algorithms. The comparison shows which algorithm gives better results compared to others. This predicts the model that can be used for further evaluations of the websites.

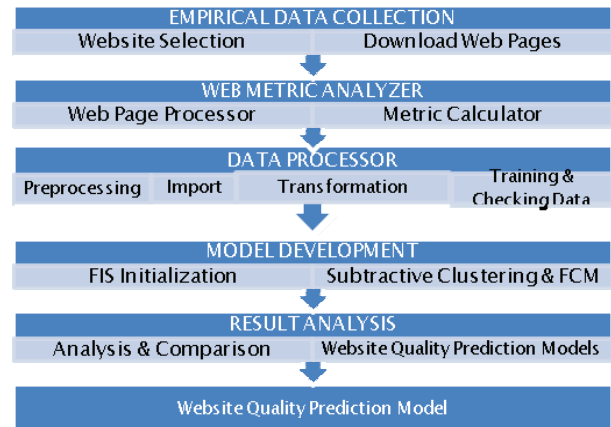


Fig. 1. Methodology Flowchart

IV. EMPIRICAL ANALYSIS

We have described the data set and its descriptive statistics in this section. The emphasis is laid upon the data preprocessing techniques and algorithms used for building the model. The evaluation criteria to quantify the accuracy of the constructed models are also discussed.

A. Data

We investigated the web pages collected from Pixel Awards website [11]. For our study, we collected datasets from 20 categories of Pixel Awards for the year 2009; using A1 Website Download Microsys tool [1]. This tool is configured to crawl 0-level i.e. homepages and 1-level pages from each site; thus, we collected homepages and 1-level web pages for each website. In this way, we have the dataset for 2678 web pages from 255 websites. There are 945 web pages in 79 websites nominated in the year 2009. Table 1 enlists the nine web metrics that we have selected as input for the development of models in this work.

The output of the model is the website categorization variable which takes two values: good or bad based on the judges rating of Pixel Awards for a specific website.

TABLE I. METRICS SELECTED FOR STUDY

Metrics	Description
Word Count	Total number of words on a page
Body Text Words	Total number of words falling in the display section of the body
Page Size	Total number of bytes i.e. size of the page and images
Table Count	Number of tables used on the web page
Graphics Count	Total number of images on a page
Division Count	Number of divisions on a page
List Count	Total number of ordered and unordered lists on a page
Number of Links	Total links on a page
Page Title length	Total number of words in the page title

B. Descriptive Statistics

The summary of the sample data set for all the metrics is described quantitatively with “min”, “max”, “mean” and “std dev” in Table II below.

TABLE II. DESCRIPTIVE STATISTICS OF 2009 DATASET

Web Metrics	Min	Max	Mean	SD
Table Count	0	388	17.57	56.17
Link Count	0	4092	915.54	1151.57
Graphic Count	0	1130	180.73	265
List Count	0	706	93.57	158.96
Page Title Length	1	368	79.73	89.90
Division Count	0	3434	696.77	941.52
Word Count	1	41158	7233.41	9333.51
Body Text Words	0	38359	6980.84	9092.20
Page Size	1	1501	322.37	374.12

C. Data Processing

Low quality data will lead to erroneous results. Therefore data preprocessing techniques such as data cleaning, data transformation and data reduction are applied to the dataset before modeling. Adjustments are made to outliers in the data by ensuring that unusually high values are removed from the dataset.

The total number of observations used for model training and checking is sixty. The quantitative description of the sixty vnumbers of datasets after preprocessing is given in Table III.

TABLE III. DESCRIPTIVE STATISTICS OF TRAINING DATASET

Web Metrics	Min	Max	Mean	SD
Table Count	0	25	2.48	5.93
Link Count	0	3170	532.6	772.24
Graphic Count	0	449	75.16	111.56
List Count	0	637	53.66	101.57
Page Title Length	1	368	60.07	83.75
Division Count	0	2107	347.87	533.46
Word Count	1	26614	4411.35	6252.6
Body Text Words	0	26244	4235.68	6100
Page Size	1	846	191.65	258

D. Algorithms Used

1) Fuzzy c-means (FCM) clustering

Fuzzy c-means (FCM) is a data clustering algorithm wherein each data point belongs to a cluster to some degree that is specified by the membership grade. This method was originally introduced by Jim Bezdek in 1981 [18].

The FCM algorithm attempts to partition a finite collection of n elements (1), into a collection of c fuzzy clusters (2), with the aim of minimization of the objective function given in (3), where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , c_j is the p -dimensional center of

the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center.

$$X = \{x_1, \dots, x_n\} \quad (1)$$

$$C = \{c_1, \dots, c_n\} \quad (2)$$

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, 1 \leq m \leq \infty \quad (3)$$

2) Subtractive clustering

Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. This method was proposed by Chiu in 1994 [19].

Consider a collection of n data points in (1), in an M -dimensional space. As each data point is a candidate for cluster centers, a *density measure* at data point x_i is defined by (4).

$$P_i = \sum_{j=1}^m \exp^{-\alpha \|x_i - x_j\|} \quad (4)$$

Where $\alpha = 4/r_a^2$ and $r_a > 0$ defines the neighborhood radius for each cluster centre.

The higher potential vector is selected as the first cluster center. Let x_1^* be the center of the first group and P_1^* its potential. Then the potential for each x_i is reduced according to (5).

$$P_i = P_i - P_1^* \exp^{-\beta \|x_i - x_1^*\|} \quad (5)$$

The cluster information is used for determining the initial number of rules and antecedent membership function for FIS.

3) Adaptive Neuro Fuzzy Inference System (ANFIS)

Fuzzy Logic and Fuzzy Inference System, first proposed by Zadeh [20], enables to take decisions based on vague, ambiguous, imprecise or missing data. There are two types of FIS: Mamdani and Sugeno [21, 22]. The Sugeno type is more compact and computationally efficient than Mamdani, therefore it is used for constructing fuzzy models with adaptive techniques. We have used the first order Sugeno fuzzy model.

In order to make the fuzzy model work appropriate in prediction mode, it is required to be adopted through a learning process using sufficient input output pattern of data. ANFIS is one of the most commonly used learning systems for first order Sugeno model and has following layers [23].

Layer 0: It consists of a plain input variable set. We have nine input variables.

Layer 1: The node function of every node i in this layer is of the form (6).

$$O_i^1 = \mu_{A_i}(x) \quad (6)$$

Where x is the input to the node and μ_{A_i} is the membership function(MF). In this study the triangular MF are utilized.

Layer 2: Every node in this layer multiplies the incoming signal from layer 1 and the output is the firing strength of the rule.

Layer 3: Every node in this layer determines the ratio of the i^{th} rule's firing strength to the sum of all rules firing strength and the output represents the normalized rule firing strengths.

Layer 4: Every node i in this layer is an adaptive node, which multiplies the output of layer 3 with the parameter set function.

Layer 5: It has a single node that computes the overall output as the summation of all incoming signals from layer 4.

In a nutshell, ANFIS takes the initial fuzzy model and adapts it by means of a hybrid approach of gradient descent back-propagation and mean least square optimization algorithm. At each epoch, the training error is reduced and training is stopped if either the predefined epoch number is reached or the error rate is obtained. We have used 100 epochs in this work.

E. Evaluation Criteria

1) Sensitivity and Specificity

Model performance is examined in terms of how well the two models are able to identify non-default cases. This is based on their respective sensitivity and specificity measures.

Sensitivity is the probability of a test to identify positive results i.e. True Positive Rate (TPR). The sensitivity of the model can be defined as the percentage of websites correctly predicted to be good, (7). Specificity is the probability of the test to identify the negative results, (8). The percentage of websites correctly predicted to be bad is called specificity (True Negative Rate i.e. TNR) of the model. Ideally, both the sensitivity and specificity should be high to predict good and bad websites.

$$\text{Sensitivity (TPR)} = \frac{\text{True positives}}{\text{True positives} + \text{False}} \quad (7)$$

$$\text{Specificity (1 - FPR)} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False pos}} \quad (8)$$

Where, TP = websites correctly predicted to be good,
TN = websites correctly predicted to be bad,
FP = websites incorrectly predicted to be good,
FN = websites incorrectly predicted to be bad.

2) Receiver Operating Characteristic (ROC) analysis

ROC curve analysis is an effective method for evaluating the performance of predicted models. In a ROC curve, the true positive rate (sensitivity) on the y-coordinate is plotted in the function of the false positive rate (1-specificity) on the x coordinate. It is known as a Relative Operating Characteristic curve, because it is a comparison of two operating characteristics (TPR & FPR). It allows us to select many cutoff points between 0 and 1, and to calculate sensitivity and specificity at each cut off point. The optimal cutoff point that maximizes both sensitivity and specificity can be selected from the ROC curve. Hence, one can easily determine the optimal cutoff point for a predicted model using the ROC curve.

3) Area Under Curve (AUC)

Area Under Curve is the value of the area under the ROC Curve. It is a combined measure of sensitivity and specificity and used to compute the accuracy of the predicted models. The perfect model scores the value of AUC as 1.

The accuracy of the model can be predicted by applying it to the different data sets. We therefore, performed a 10-cross validation of the models [14]. Each dataset is randomly divided into 10 equal subsets. Each time one of the 10 subsets is used as the test set and the other 9 subsets are used to form a training set. Model I and Model II are predicted with respect to dataset of 2009.

F. Model Building

The models are developed using clustering algorithms and trained by ANFIS. The structure of the models utilizing different clustering function and trained by ANFIS is depicted in Fig 2.

Given separated sets of input and output data, genfis2 is a function of MATLAB which uses a subtractive clustering algorithm for creation of the prediction model. The parameters of subtractive clustering are set with the following values: the range of influence is 0.5, squash factor is 1.25, accept ratio is 0.5 and rejection ratio is 0.15. The number of epochs is equal to 100. The root mean square error generated by the training and checking data for ANFIS Subtractive Model against a number of epochs is shown in Fig 3. The training error and checking error oscillate in the range from 0.34 to 0.39. The best model is achieved when the test error is minimal. This corresponds to the circle in Fig 3 at the 13th epoch. The optimum value of the training error is 0.345, obtained from the RMSE value of the curve at 13th epoch.

The MATLAB function genfis3 utilizes a FCM clustering algorithm and is used to create Model II. The developed model trained by ANFIS, extracts a set of rules that models the data behavior. The training and checking error curves for ANFIS FCM Model are shown in Fig 4. The value of training and checking error is in the range 0.36 to 0.51. The minimal test error

corresponds to the circle in Fig 4 at the 48th epoch. The optimum value of the training error is 0.361.

Fig. 2. FIS Clustering Model Structure

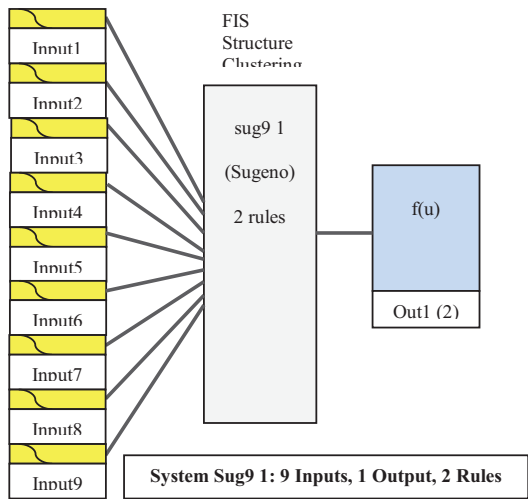


Fig. 3. Error Curve for ANFIS Subtractive Model

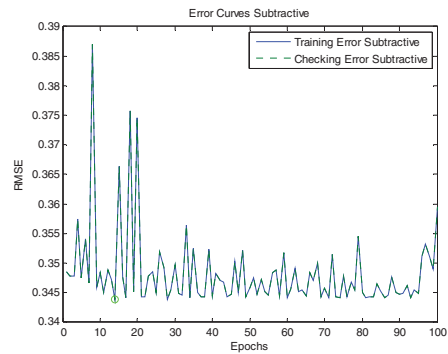
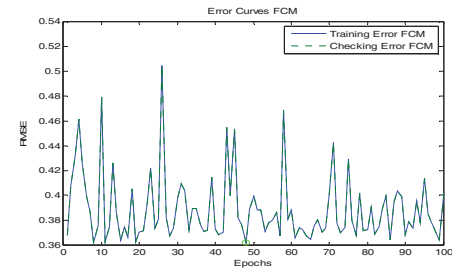


Fig. 4. Error Curve for ANFIS FCM Model



V. RESULTS AND ANALYSIS

The work in this paper examines the predictive performance of two models, ANFIS Subtractive and ANFIS FCM. A comparative examination of these two predictive models is conducted in terms of their predictive accuracies of website classification as good or bad. The value of each model’s sensitivity, specificity, AUC, cutoff point and number of fuzzy rules are summarized in Table IV. The ROC curve is plotted with sensitivity on y-axis and (1-specificity) on the x-axis and ROC analysis is used to find the cutoff point. The cutoff point is selected such that the sum of sensitivity and specificity is maximized as both sensitivity and specificity should be high to predict good and bad websites. The ROC curve for subtractive-ANFIS and FCM-ANFIS models are shown in fig 5 and fig 6 respectively.

It is observed that both subtractive-ANFIS and FCM ANFIS have acceptable performance in terms of sensitivity and specificity. The specificity, sensitivity and AUC for the subtractive-ANFIS is 94.1%, 79.1% and 0.901 while for FCM-ANFIS the values are 94.1%, 74.4% and 0.843 respectively. The cutoff value and number of fuzzy rules for both the models are same. In addition, ANFIS-Subtractive and ANFIS-FCM induce only two fuzzy rules, which are much less than 512 rules generated by the normal ANFIS model.

One of the major results of this study is that ANFIS clustering techniques can be used to predict the quality of the website. Both subtractive and FCM can be used to generate and adapt the FIS model. The subtractive clustering can be used to estimate the number of clusters and cluster centers, whereas FCM starts with the provided number of clusters.

TABLE IV. PERFORMANCE OF MODELS

Model	Sensitivity	Specificity	AUC	Cut-off point	No of Rule
ANFIS Subtractive	94.10	79.1	0.901	0.5	2
ANFIS FCM	94.10	74.40	0.843	0.5	2

Fig. 5. ROC Curve for ANFIS Subtractive Model

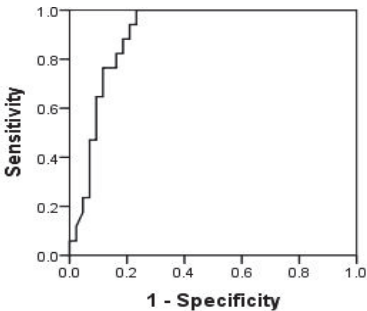
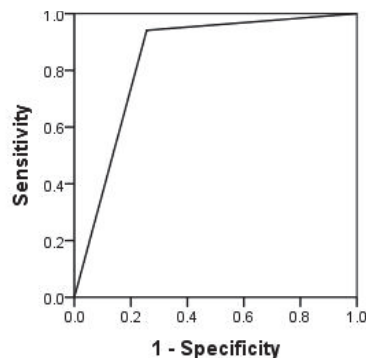


Fig. 6. ROC Curve for ANFIS FCM Model



VI. CONCLUSION

Web Applications have evolved into complex applications. The quality of the website is at stake due to short development cycles and frequent modification requests. In order to control the quality of the website, predictive models like the one developed in this study, which classify the website as good or bad is useful for industry. However, the work done is subjected to application of important web metrics as input to the model.

The AUC of subtractive ANFIS and FCM ANFIS are 0.901 and 0.843 respectively. The limitations of this study is that websites are categorized into two classes only, more number of classes can be included for website categorization.

This study confirms that construction of models using ANFIS is feasible, and useful in predicting the goodness of the websites. This research can be carried forward for all the levels of web pages in the website and some more web page metrics can be introduced that are more effective for the websites.

REFERENCES

- [1] A1 Website Download: <http://www.microsystools.com/products/website-download/>
- [2] Alsmadi, I., Al-Taani, A.T., Zaid, N.A.: Web Structural Metrics Evaluation. *Developments in E-systems Engineering (DESE)* (2010)
- [3] Chi, E.H., Pirolli, P., Pitkow, J.: The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. *Proceedings of ACM CHI 00 Conference on Human Factors in Computing Systems* (2000)
- [4] Ivory, M.Y., Sinha, R., Hearst, M.A.: Preliminary findings on quantitative measures for distinguishing highly rated information-centric web pages. *Proceedings of the 6th Conference on Human Factors and the Web* (2000)
- [5] Ivory, M.Y., Sinha, R., Hearst, M.A.: Empirically Validated Web Page Design Metrics. *Proceedings of the SIGCHI conference on Human factors in computing systems*, p.53-60, Seattle, Washington, United States (2001)
- [6] Khan, K.M.: Assessing Quality of Web Based System. *IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, p.763-769 (2008)
- [7] Leavitt, M.O., Shneiderman, B.: *Research Based Web Design and Usability Guidelines*. U.S. Department of Health and Human Services, Washington (2006)
- [8] Li, P., Yamada, S.: Automated Web Site Evaluation – An Approach Based on Ranking SVM. *International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, IEEE/WIC/ACM*, pp. 34-37 (2009)
- [9] Li, P., Yamada, S.: Extraction of Web Site Evaluation Criteria and Automatic Evaluation. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.14 No.4, pp. 396-401 (2010)
- [10] Nielsen, J.: *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, Indianapolis, IN (2000)
- [11] Pixel Awards | Web Awards Competition (2009). Available from <http://www.pixelawards.com>.
- [12] Scholtz, J., Laskowski, S., Downey, L.: Developing usability tools and techniques for designing and testing web sites. *Proceedings of the 4th Conference on Human Factors & the Web*, (1998)
- [13] Singh, Y., Malhotra, R., Gupta, P.: Empirical Validation of Web Metrics for Improving the Quality of Web Page. *International Journal of Advanced Computer Science and Applications (IJASCA)*, Vol. 2, No. 5, pp. 22-28 (2011)
- [14] Stone, M.: Cross-validators choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B (Methodological)*, 36, pp. 111-147 (1974)
- [15] Thimbleby, H.: Gentler: A tool for systematic web authoring. *International Journal of Human-Computer Studies*, 47, pp. 139-168 (1997)
- [16] Velayathan, G., Yamada, S.: Behavior-Based Web Page Evaluation. *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT Workshops)*, pp. 409-412 (2006)
- [17] Zorman, M., Podgorelec, V., Kokol, P., Babic, S.H.: Using machine learning techniques for automatic evaluation of Websites. *Proceedings of the Third International Conference on Computational Intelligence and Multimedia Applications ICCIMA '99*, pp. 169-173, New Delhi, India, IEEE Computer Society Press, September (1999)
- [18] Bezdec, J.C., 1981.: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York
- [19] Chiu, S., 1994.: Fuzzy Model Identification Based on Cluster Estimation, *Journal of Intelligent & Fuzzy Systems*, Vol 2, No. 3, Sept, pp 267-278.
- [20] Zadeh, L.A., 1965. Fuzzy sets, *Information and Control*, pp. 338-253.
- [21] Mamdani, E.H., Assilian, S., 1975: An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man Machine studies*, pp. 1-13.
- [22] Sugeno, M., 1985: *Industrial applications of fuzzy control*. Elsevier Science Pub., Co.
- [23] Jang, J-S. R., 1993: ANFIS: Adaptive-Network-based Fuzzy Inference System, *IEEE Transactions on Systems, Man and Cybernetics*, pp.665-685.
- [24] Nguyen, H.Q., 2000: *Testing Applications on the Web: Test Planning for Internet-Based Systems*. John Wiley & Sons, Inc.
- [25] Mittal, H., Sharma, M., Mittal, J.P., 2012: Analysis and Modelling of Websites Quality Using Fuzzy Technique. *Proceedings of the Second International Conference on Advanced Computing & Communication Technologies (ACCT)*, IEEE 2012, pp. 10-15.

A Variable Length Distributed Source Coding Algorithm for WSNs

K S Shivaprakasha¹, Muralidhar Kulkarni²

Abstract

Ubiquitous computing has become an integral part of the modern communication system. Wireless Sensor Networks (WSNs) are networks of small devices called sensors deployed in a geographical area to sense some physical entity. The sensed information will be then conveyed to the central entity called the Base Station (BS). WSN nodes being generally deployed in harsh environments, recharging or replacing of batteries become an infeasible task. Thus it is desirable to represent the information using as minimum number of bits as possible. One such technique to perform this is source coding. In this paper an attempt is made in proposing a novel algorithm called Variable Length Distributed Source Coding (VLDSC) which is an improvement over the existing Huffman coding technique. The results validate the betterment in the performance of the proposed algorithm.

Keywords

Distributed Source Coding, Huffman Coding, DSC using Syndromes, Variable Length Distributed Source Coding (VLDSC)

1. Introduction

Wireless Sensor Networks (WSNs) have become one of the most important research areas in the modern communication system. WSNs find applications in vast areas like military, disaster management, habitat monitoring, environmental monitoring etc [1, 2]. WSN consists of hundreds of small devices called motes. A mote consists of a sensor to sense the information, a processor to process the same, transmitting and receiving antennas, memory unit and power units. The sensed information will be communicated to the BS which in turn is connected to the physical world. A typical WSN is as shown in figure 1. BS being facilitated with the fixed infrastructure, is not energy constrained.

K S Shivaprakasha, Department of Electronics and Communication Engineering, National Institute of Technology Karnataka (NITK), Surathkal, INDIA.

Muralidhar Kulkarni, Department of Electronics and Communication Engineering, National Institute of Technology Karnataka (NITK), Surathkal, INDIA.

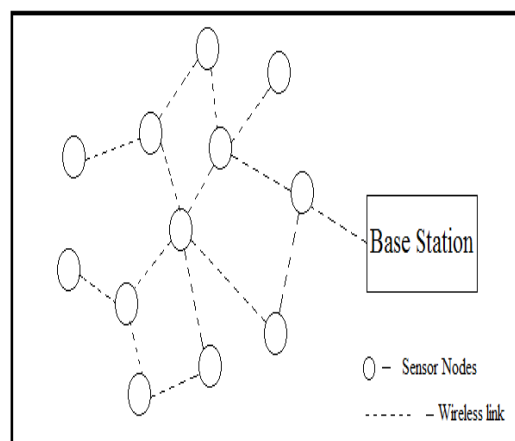


Figure 1: Basic Wireless Network

However motes suffer severe energy restrictions as recharging the batteries is difficult in WSNs. Thus the representation of the data should be done with minimum number of bits. Source coding is one of the promising techniques to achieve data compression before transmission [3].

Source Coding (SC) can be either centralized or distributed. In centralized approach there will be an encoder to compress the information and decoder at the receiver to retrieve the information. Whereas in distributed algorithms the compression is done in the distributed manner amongst motes with or without mutual exchange of information. Distributed approach suits WSN better as it also imparts load distribution amongst all nodes.

An ample number of Distributed Source Coding (DSC) algorithms have been proposed in the literature. The performance of the DSC algorithm is also dependent on the underlying routing algorithms. Most of the studies have concluded that the cluster based algorithms perform better [4]. In cluster based algorithms, the entire geographical area is divided into clusters consisting of a set of nodes with a respective Cluster Head (CH). CH monitors the data transmissions in the corresponding cluster.

Figure 2 presents a WSN incorporating a hierarchical architecture. Different clusters and the corresponding CHs and cluster members have been highlighted.

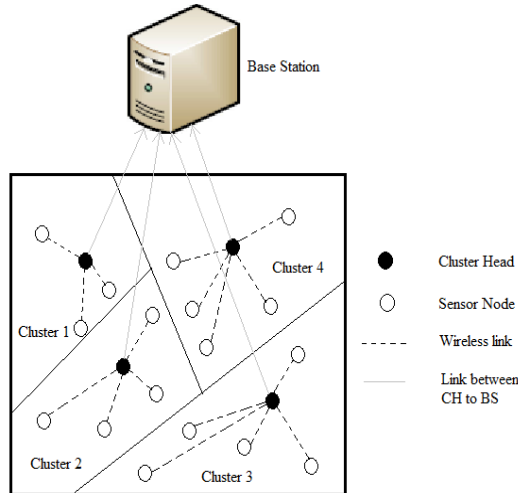


Figure 2: Basic CH Communication in DSC

Nodes with higher energy become the CHs. CHs collect the data from all its cluster members, combine the same and then transmit the fused data to the BS. SC can further improve the performance. Huffman variable length distributed source coding and Distributed Source Coding using Syndromes (DISCUS) algorithms were proved to offer a good compression ratio. However the algorithms can be improved for better efficiency. In this paper an attempt has been made in improvising Huffman variable length distributed source coding so as to assure a better compression ratio and thus enhancing the network lifetime. The performance is validated through simulations.

The rest of the paper is organized as follows: section 2 gives an insight on the previous works on DSC proposed in the literature. Section 3 presents a detailed description on the Huffman coding for WSNs. Section 4 details the proposed VLDSC algorithm. Simulation results and analysis have been carried out in section 5. Finally Section 6 gives the concluding remarks.

2. Previous Works

In WSNs, nodes are densely deployed; hence correlation exists amongst the data sensed by nodes. SC is a very important problem in information and communication systems. WSN devices are powered by batteries. Most of the WSNs are placed in hostile area, where battery recharging and replacing is very difficult. Hence better energy efficient data communication technique is required.

DSC is one such technique [3]. DSC generally shifts the complexity to the decoder side. In the cluster based communication, one of the nodes in the cluster is considered as the CH. The CHs gather the information from all nodes in the cluster and send it to the BS [5, 6]. Slepian and Wolf have given a limit on the rate of transmission and have theoretically showed that separate encoding (with increased complexity at the joint decoder) is as efficient as joint encoding for lossless compression [7, 8].

In [9], authors have proposed the design for two estimators based on the correlation information so as to achieve a practical DSC. DISCUS algorithm achieves DSC by modeling the source coding problem into a channel coding one and thus exploring the usage of syndrome bits to represent the side information [10, 11]. An improvement over the syndrome based DSC has been proposed in [12]. It not only offers a higher compression ratio but also is rate adaptive. It considers integer valued data for compression.

3. Huffman Algorithm for DSC

In [13] authors have proposed a novel algorithm using Huffman coding to encode the data to be sent by the sensor in a WSN. Let m_i be the data to be sent in the i^{th} time instant. Let r_i be the binary equivalent of the data consisting of R bits where R indicates the resolution of the ADC. Then the corresponding codeword d_i can be represented as a combination of two bit sequences $\langle s_i a_i \rangle$ where s_i indicates the number of bits n_i used to represent the data and a_i gives the corresponding representation.

The term d_i can be calculated as $d_i = r_i - r_{i-1}$, i.e. the difference in the measured entity in a sensor node. The representation of s_i for different values of d_i is as shown in the table 1. Table 1 is retained from the baseline JPEG algorithm.

Table 1: Huffman Variable Length Codes used

n_i	s_i	d_i
0	00	0
1	010	-1, +1
2	011	-3, -2, +2, +3
3	100	-7, ..., -4, +4, ..., +7
4	101	-15, ..., -8, +8, ..., +15
5	110	-31, ..., -16, +16, ..., +31
6	1110	-63, ..., -32, +32, ..., +63
7	11110	-127, ..., -64, +64, ..., +127
8	111110	-255, ..., -128, +128, ..., +255
9	1111110	-511, ..., -256, +256, ..., +511
10	11111110	-1023, ..., -512, +512, ..., +1023
11	111111110	-2047, ..., -1024, +1024, ..., +2047

a_i can be calculated as follows:

- If $d_i > 0$, a_i is the 2's complement representation of the sequence d_i
- If $d_i < 0$, a_i is the 2's complement representation of the sequence d_i-1
- If $d_i=0$, a_i is not transmitted

The transmitter transmits the tuples $\langle s_i, a_i \rangle$ for every d_i . Simulation results show a good compression achieved despite its simplicity.

4. Variable Length Distributed Source Coding (VLDSC) Algorithm

The algorithm discussed in section 3 offers a better performance and simplicity in design. However some modifications can be incorporated so as to improve its performance and thus enhance the network lifetime. This section provides an overview of the proposed VLDSC algorithm for WSNs.

The following assumptions were made in the proposed VLDSC:

- Nodes are assumed to be stationary.
- BS has the complete network information.
- BS does cluster formation and CH selection based on M-TRAC algorithm [14].
- CH allocates schedule for their cluster members based on TDMA.
 - Scheduling is done in such a way that node in the common transmission range of all other nodes in the cluster will be given the chance to transmit first. However it is always true as the transmission within the cluster is direct.
 - Other nodes listening to the transmission of the first sender, transmit the CH, the difference between its data and the data of the first node
 - The CH collects data from all members, encodes using the proposed algorithm and then sends the encoded sequence to the BS.
- Transmission between the nodes and CH is direct and that of CH to the BS is multihop.
- Nodes will have the continuous monitoring of data and the data is conveyed to the BS either
 - When there is a significant change in the sensing entity (event driven).

In this case only the node sensing the significant change initiates a request for transmitting and the normal M-TRAC algorithm is used for communication. As there is only one sender, no DSC is required.

- Periodically, in which case nodes transmit the data as per the TDMA policy. First node sends the n-bit data uncoded. Listening to the data transmitted by the first node, others transmit the difference to the CH. CH performs encoding as per the proposed algorithm.

As per our algorithm we consider to transmit the difference in the measured entity w.r.t the data transmitted by the first node. As the nodes are likely to be densely deployed, the measured entity will be correlated and thus the difference can be 0, ± 1 , ± 2 , ± 3 and so on. It is clear from our assumption that the difference being 0 or ± 1 would be more likely to occur than the differences being ± 2 , ± 3 and so on. Thus it is advantageous to encode the smaller difference with less number of bits and the length is an increasing function w.r.t the difference.

The following improvements over the algorithm proposed in [13] have been done. The algorithm considers the compression of the data from a single sensor node. The auto-correlation of the data sensed by a sensor is explored for the data compression. However it can be modified for a cluster based query based/time driven WSN where the data sensed by the nodes within a cluster are more likely to be correlated. Thus an aggregation has to be performed at the CH. CHs collect all data from its cluster members, say M_1, M_2, M_3, M_4, M_5 if there are 5 nodes in the cluster. All these 5 data are highly spatially correlated. Thus the first data can be sent as it is whereas for the other messages the difference sequences d_2, d_3, d_4 and d_5 can be transmitted to the CH. CH then performs the aggregation using the procedure as explained in the previous algorithm.

The proposed compression algorithm also has to assure a simple computational requirement as the computational energy requirement may override the amount of energy saved in the reduction of the size of the data representation. In the algorithm proposed in [13], each of the nodes has to compute the 2's complement of either d_i or d_i-1 before transmission. However this computation can be avoided if there is a priori tabulated code vectors, a_i for each of the difference vector d_i . It is as shown in table 2.

Thus if both the sender and receiver agrees upon this table during network initialization, the CH, need not has to compute anything. Instead it has to just transmit an entry corresponding to the difference from the table.

Table 2: Modified Huffman Variable Length Codes used

n_i	s_i	d_i	a_i
0	00	0	---
1	010	-1	0
		+1	1
2	011	-3	00
		-2	01
		+2	10
		+3	11
3	100	-7	000
		-6	001
		-5	010
		-4	011
		+4	100
		+5	101
		+6	110
		+7	111
...
...

Also the proposed algorithm is an improvement compared to DISCUS too on the following grounds. In DISCUS, it is assumed that the maximum Hamming distance amongst the data sensed by nodes as 1. As in general the data being sensed in a WSN is analog, in our proposed algorithm, we consider the analog representation of the data. Thus the difference in the sensed entity of the nodes in a cluster can be either 0 (no difference), ± 1 (difference of 1 either in positive or negative direction), ± 2 and so on.

In DISCUS, nodes transmit 'n' bit data to a central entity and the central entity does the DSC, by transmitting 'n' bits for the first data and just '(n-k)' bits of syndrome to represent the remaining data. The receiver then decodes the data from the syndrome as the maximum allowed Hamming distance is 1. However the transmission in the last mile (i.e nodes to a central entity say CH) is uncompressed.

In the proposed algorithm we have explored the advantage of using wireless medium for transmission in WSN. All nodes in the range receive the data transmitted by the first node as the first node to transmit data is assumed to be in the common range of all other nodes in the cluster. Thus instead of CH

receiving complete data from all nodes in the cluster and then compressing, we propose each of the nodes itself can directly transmit the compressed data by sending just the difference in the sensed entity. This achieves better energy efficiency by not only reducing the overall network traffic but also assuring a better load balancing as CH is relieved from getting overburdened.

In DISCUS, a fixed length (n-k) bits are used for sending the information and thus saving 'k' bits per data. In the proposed algorithm we propose a variable length coding where for every non negative value for 'l' there exists 2^l codes of length 'l'. Thus differences which are more likely to occur will be coded with lesser bits which will further add to the energy conservation in the network.

The encoding procedure is as follows

- The first node in the time schedule transmits its data to the CH.
- Hearing to the data transmitted by the first node, further nodes transmit the difference signal to the CH
- CH computes the pair $\langle s_i, a_i \rangle$ for all nodes in the cluster. Where s_i and a_i can be obtained from table 2.
- The fused information is communicated to the BS from the CH.
- Each node has to be active only for two time slots. One is for the first time slot to get the information about the first data sent by a node and the other is at its slot to transmit the difference signal.

5. Simulation Results and Analysis

To validate our algorithm, we have considered the sensirion [15] sensor nodes. Sensirion SHT11 module is used for the analysis. It measures the temperature and relative humidity. The module equipped with a 14-bit ADC. From the datasheetsof the SHT11, 12 bit resolution is used for humidity and 14 bit for the temperature.

We have considered 500 samples of temperature and humidity for the analysis. We considered a cluster of 10 nodes for the validation. The results are detailed in table 3.

Table 2: Compression Ratio in VLDSC

	Temperature	Humidity
Original Size	7000 bits	6000 bits
Compressed Size	2388 bits	1994 bits
Compression Ratio	65.88%	66.76%

It is seen from the table that the proposed VLDSC offers a very high compression ratio and thus assures better energy efficiency.

6. Conclusion and Future Scope

Energy aware data communication is the most vital criterion in the design of a WSN. DSC along with routing assures a better data representation in a sensor network. In this paper an improvement over the Huffman Coding for WSNs has been proposed resulting in VLDSC which offers a better compression. The proposed algorithm has been validated for its performance through simulations. Although the work highlights the advantage of using a DSC for WSNs, selection of appropriate scheduling policy also has an impact on its performance. As a part of future work, an appropriate scheduling algorithm has to be devised for the proposed algorithm. Also the performance of the proposed algorithm can be compared with any of the existing ones to further validate the results.

References

- [1] WaltenegusDargie, Christian Poellabauer, "Fundamentals of Wireless Sensor Networks Theory and Practice", John Wiley and Sons, 2010.
- [2] Krishnamachari, L., Estrin, D., & Wicker, S. "The impact of data aggregation in wireless sensor networks" In Distributed Computing Systems Workshops. Proceedings. 22nd International Conference on (pp. 575-578). IEEE, 2002.
- [3] Pier Luigi and MichaeGastpar, "Distributed Source Coding", Academic Press, Elsevier, 2009.
- [4] K.S. Shivaprakasha, MuralidharKulkarni, "Energy Efficient Routing Protocols for Wireless Sensor Networks: A Survey", International Review on Computers and Software (IRECOS), Vol 6, No 6, pp 929-943, 2009.
- [5] Xiong, Z., Liveris, A. D. & Cheng, S. "Distributed Source Coding for Sensor Networks", Signal Processing Magazine, IEEE, 21(5), pp 80-94, 2004.

- [6] Arjmandi, H., & Lahouti, F. "Resource Optimized Distributed Source Coding for Complexity Constrained Data Gathering Wireless Sensor Networks" Sensors Journal, IEEE, 11(9), pp 2094-2101, 2011.
- [7] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inform. Theory, Vol. IT-22, pp. 1-10, Jan. 1976.
- [8] Stankovic, V., Stankovic, L., & Cheng, S. "Distributed Source Coding: Theory and Application", Proceedings of European Signal Processing Conference, 2010.
- [9] Barceló-Lladó, J. E., Pérez, A. M., & Seco-Granados, G. "Enhanced Correlation Estimators for Distributed Source Coding in Large Wireless Sensor Networks" Sensors Journal, IEEE, 12(9), pp 2799-2806, 2012.
- [10] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," Proc. IEEE Data Compression Conference, Snowbird, UT, March 1999.
- [11] Pradhan, S. S., & Ramchandran, K. "Distributed source coding: Symmetric rates and applications to sensor networks", In Data Compression Conference. Proceedings. DCC 2000, pp. 363-372. IEEE, 2000.
- [12] Chen, J., Khisti, A., Malioutov, D. M., & Yedidia, J. S. "Distributed source coding using serially-concatenated-accumulate codes", In Information Theory Workshop, 2004. IEEE, pp. 209-214. IEEE, 2004.
- [13] Francesco Marcelloni, Massimo Vecchio, "A Simple Algorithm For Data Compression In Wireless Sensor Networks", IEEE Communications Letters, Vol. 12, No. 6, June 2008
- [14] K S Shivaprakasha, M Kulkarni, Nishant Joshi D, "Improved Network Survivability using Multi-threshold Adaptive Range Clustering (M-TRAC) Algorithm for Energy Balancing in Wireless Sensor Networks", Journal of High Speed Networks, Vol 19, pp 99-113, 2013.
- [15] SHT15 - Digital Humidity Sensor (RH&T). Available: <http://www.sensirion.com/en/products/humidity-temperature/humidity-sensor-sht15/>



K S Shivaprakasha received his BE (Electronics & Communication) from Bahubali College of Engineering, Visvesvaraya Technological University with IX rank to the University and MTech (Digital Electronics and Communication Systems) from Malnad College of Engineering, Visvesvaraya Technological University with I rank to the University in 2004 and 2007 respectively.

Presently he is pursuing his PhD at National Institute of Technology Karnataka, Surathkal in the field of Wireless Sensor Networks. He is currently working as Associate Professor in the Department of Electronics and Communication Engineering, Bahubali College of Engineering, Shravanabelagola, Karnataka, India. His areas of interest include Wireless Sensor Networks, Mobile Adhoc Networks, Information Coding Theory and Cryptography. He has more than 15 publications to his credit.



Muralidhar Kulkarni received his B.E. (Electronics Engineering) degree from University Visvesvaraya College of Engineering, Bangalore University, Bangalore, M. Tech (Satellite Communication and Remote Sensing) from Indian Institute of Technology, Kharagpur (IIT KGP) and

PhD from JMI Central University, New Delhi in the area of Optical Communication networks.

He has 31 years of experience which includes 5 years in industry and 26 years of teaching experience. He has held the positions of Scientist in Instrumentation Division at the Central Power Research Institute, Bangalore (1981-1982), Aeronautical Engineer in Avionics group of Design and Development team of Advanced Light Helicopter(ALH) project at Helicopter Design Bureau at Hindustan Aeronautics Limited(HAL), Bangalore (1984-1988), Lecturer (Electronics Engineering) at the Electrical Engineering Department of University Visvesvaraya College of He has 31 years of experience which includes 5 years in industry and 26 years of teaching experience.

He has held the positions of Scientist in Instrumentation Division at the Central Power Research Institute, Bangalore (1981-1982), Aeronautical Engineer in Avionics group of Design and Development team of Advanced Light Helicopter(ALH) project at Helicopter Design Bureau at Hindustan Aeronautics Limited (HAL), Bangalore (1984-1988), Lecturer (Electronics Engineering) at the Electrical Engineering Department of University Visvesvaraya College of Engineering, Bangalore (1988-1994) and Assistant Professor in Electronics and Communication Engineering (ECE) Department at the Delhi College of Engineering (DCE), Govt. of National Capital territory of Delhi, Delhi (1994-2008). He has served as Head, Department of Information Technology and Head, Computer Center at the Delhi College of Engineering (University of Delhi), Delhi. Currently, he is a Professor and Head in the Department of Electronics and Communication Engineering (ECE) Department, National Institute of Technology Karnataka (NITK), Surathkal, Karnataka, India.

He is currently the Coordinator of the Centre of Excellence for Wireless Sensor Networks, Dept. of Electronics and Communication Engineering, National Institute of Technology Karnataka. Dr. Kulkarni's teaching and research interests are in the areas of Digital Communications, Fuzzy Digital Image Processing, Optical

Communication and Networks, and Wireless Sensor Networks. He has published several research papers in the above areas, in national and International journals of repute. For various contributions his Biography has been listed in the Marquis, Who's Who in Science & Engineering (2008). He has also authored/coauthored four very popular books in Microwave & Radar Engineering, Communication Systems, Digital Communications and Digital Signal Processing.

An FRAM-Based Nonvolatile Logic MCU SoC Exhibiting 100% Digital State Retention at $V_{DD} = 0$ V Achieving Zero Leakage With <400 -ns Wakeup Time for ULP Applications

Sudhanshu Khanna, *Member, IEEE*, Steven C. Bartling, Michael Clinton, *Member, IEEE*, Scott Summerfelt, John A. Rodriguez, *Member, IEEE*, and Hugh P. McAdams

Abstract—This paper presents a nonvolatile logic (NVL)-based 32-b microcontroller system-on-chip (SoC) that backs up its working state (all flip-flops) upon receiving a power interrupt, has zero leakage in sleep mode, and needs less than 400 ns to restore the system state upon power-up. Nonvolatile Fe-Cap-based mini-arrays backup the machine state and allow the chip to wake up instantly after a power cycle. Without NVL, a chip would either have to keep all flip-flops powered, resulting in high standby power, or waste energy and time rebooting after power-up. NVL allows systems to use leakier processes to achieve higher performance/lower dynamic power while still having zero leakage in the sleep mode. Optimized system, architecture, and circuit techniques are presented that make NVL practical by adding only 3.6% to the SoC area. Since nonvolatile elements are added to the SoC, reliability and testability have to be key features of the design. This is the first NVL SoC with measured NVL bitcell read signal margin data and extensive test and debug capabilities. The chip is fabricated in a commercial 130-nm low-leakage process and uses a single 1.5-V power supply.

Index Terms—Ferroelectric RAM (FRAM), low power, micro-controllers, nonvolatile logic (NVL).

I. INTRODUCTION

ADVANCES in low-power VLSI design over the last two decades have made ICs pervasive in our lives. We are surrounded by smart systems like ambient sensors, home automation, and implantable/wearable fitness and medical devices. These embedded systems have varied applications and system requirements. They also have lower volumes and system cost than consumer products like laptops or cell phones. Hence, programmable solutions like low-power 16–32-b micro-controllers (MCUs) are preferred over ASICs or subthreshold systems which have a limited use or performance range.

Lowering power consumption is the key design challenge in this application space. To reduce dynamic power techniques like

clock gating, dynamic voltage scaling, low-voltage operation, efficient RTL design, and optimized standard cell libraries are employed [1], [2]. However, since these applications sleep for a majority of their lifetime, reducing standby power is even more critical. For this reason, state-of-the-art ultralow-power (ULP) SoCs are developed using older technology nodes like 90 nm or 130 nm. In ULP MCU SoCs, standby power is dominated by digital logic and memory leakage and the on-chip power-management quiescent current. To achieve the standby power goals, SoCs implement complex power management schemes. Multiple sleep modes with varying degrees of power gating are used. This eliminates leakage from the digital sections of the SoC and allows the associated LDOs to be turned off as well. However, power gating causes flip-flops to lose their saved state. This means that, every time the SoC comes back from sleep, the MCU has to boot up again. The boot-up process takes thousands of cycles, hundreds of microseconds, and hundreds of nanojoules of energy. The boot-up process is costly for programmable SoCs because these SoCs are built for a range of applications and cost points and are configured at boot. Trim and calibration for power management and analog is also done at boot. This factory boot is then followed by the customers boot-up code. For simple applications like a temperature sensor, boot-up becomes the major component of the overall energy and time cost. Applications that need lower latencies would not be feasible at all due to boot-up time. To prevent having to boot-up after sleep mode, a technique called retention is used where the slave stage of each flip-flop is on a separate “always-on” supply which is not power-gated. This helps decrease wakeup time and energy by eliminating need for boot-up but it increases the standby power due to the “always-on” retention flip-flops and the LDO generating the “always-on” supply. In this manner, different sleep modes trade off sleep mode leakage, wakeup time, and design volatility.

We present an MCU SoC that achieves zero sleep mode leakage while still having an ultrafast 400-ns wakeup time. We achieve 100% digital state retention. No boot-up is required after a power cycle. Before going into sleep mode, data from all flip-flops and latches in the system are transferred to custom ferroelectric-capacitor (FeCap)-based nonvolatile memory arrays called NVL arrays. The SoC has traditional ferroelectric RAM (FRAM) as a unified data and program memory. Ferroelectric capacitors are programmable capacitors that can retain their

Manuscript received April 28, 2013; revised August 14, 2013; accepted September 02, 2013. This paper was approved by Guest Editor Byeong-Gyu Nam.

S. Khanna, S. C. Bartling, S. Summerfelt, J. A. Rodriguez, and H. P. McAdams are with Texas Instruments, Dallas, TX 75243 USA (e-mail: skhanna@ti.com).

M. Clinton was with Texas Instruments, Dallas, TX 75243 USA. He is now with TSMC, Austin, TX 78759 USA.

Digital Object Identifier 10.1109/JSSC.2013.2284367

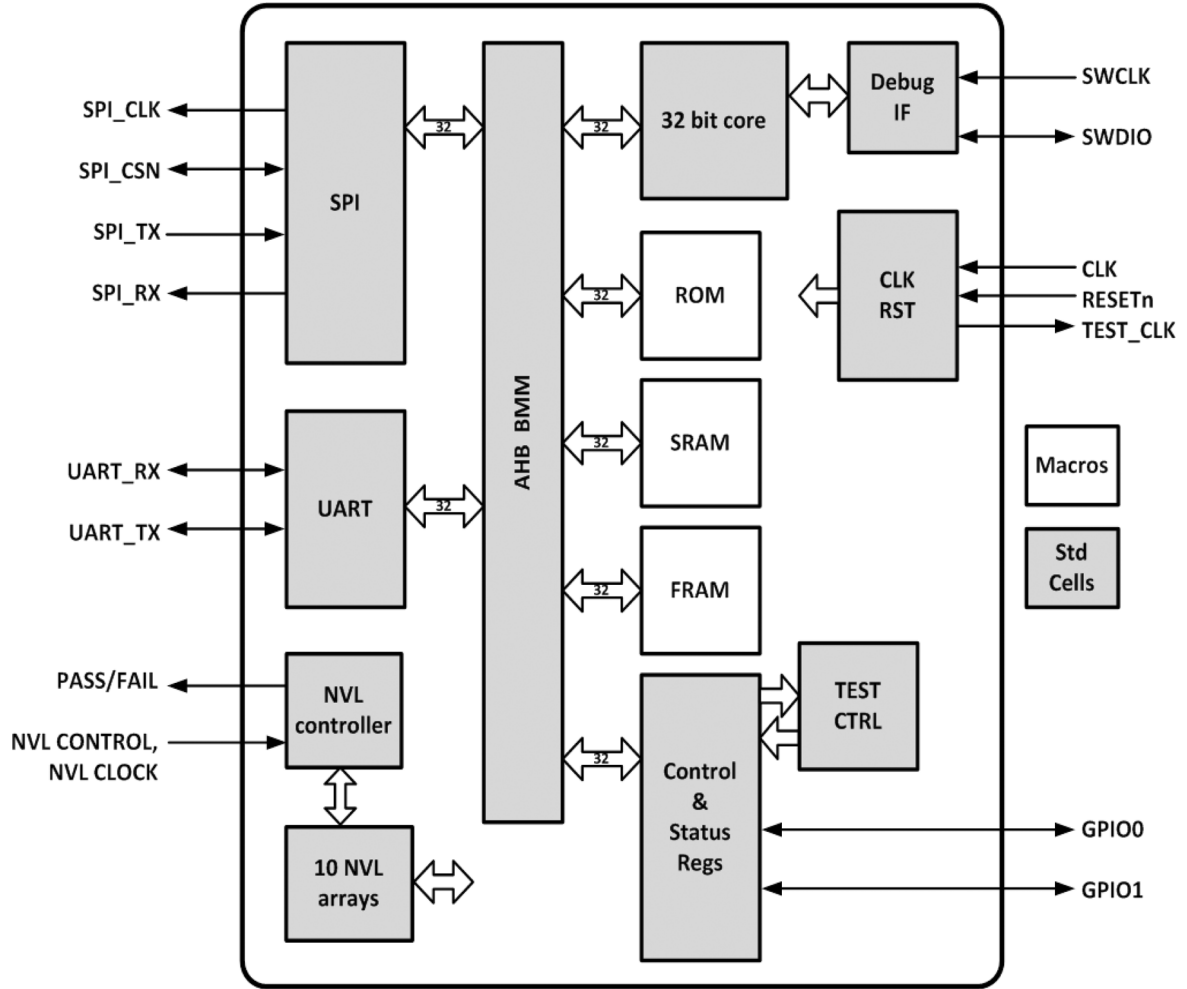


Fig. 1. . SoC architecture with major system components. NVL arrays are dispersed in the physical standard cell area.

programmed state without the need for a power source. FRAM integration into regular CMOS process has been previously demonstrated [3] with high yields and only 2 additional masks and is already used in high volume MCU products [1], [2]. Thus during sleep mode, no power is required to maintain system state. When the chip comes back to functional mode from a sleep mode, data from NVL arrays is transferred back to the flip-flops and latches, and the SoC can resume operation from where it left off. This SoC has the fastest ($8\times$), lowest energy ($8\times$), and lowest area overhead (3.6%) NVL implementation presented to date.

The concept of an NVL SoC is not new. Multiple nonvolatile flip-flop topologies [4]–[6] have been reported with the intent of enabling an NVL SoC. These designs embed ferroelectric [4], [5] or MRAM [6] elements inside a flip-flop to enable non-volatile state retention. However, all these are demonstrated at the flip-flop level. There is only one previous NVL SoC implementation [7] reported in literature. However, it has 90% area overhead, a slow $3\text{-}\mu\text{s}$ wakeup time, and no bit-level testability for the nonvolatile flip-flops.

We introduce the system architecture in Section II. Section III provides an introduction to FeCaps and their application in memory design. Section IV describes how the NVL arrays interact with the flip-flops in the digital core. NVL bitcell and

array design is described in Section III-A. System design, physical implementation, and EDA flow is discussed in Section V. Section VI describes the various use cases for an NVL SoC and quantifies the benefits. Finally, Section VII concludes the paper and describes future work.

II. SYSTEM ARCHITECTURE AND NVL FUNCTIONAL DESCRIPTION

Fig. 1 shows the system architecture of our NVL MCU SoC. It consists of a commercial 32-b processing core, UART and SPI peripherals, serial wire debug interface, FRAM for program and data, ROM test firmware, and SRAM for test purposes. A central NVL controller controls the operation of the NVL arrays distributed throughout the SoC. We rely on off-chip components to regulate the power supply for this version of the chip.

Fig. 2 shows a functional diagram of the NVL SoC. When the supply detection (SD) circuit detects a supply droop, the switch at the output of the LDO is turned off to prevent spurious voltages from reaching the SoC. The on-board supply capacitor holds enough charge to support the remaining work. To tolerate unplanned power interrupts, the minimum size of the supply capacitor is dictated by the energy needed to complete a data backup. In this SoC, we use an external capacitor but

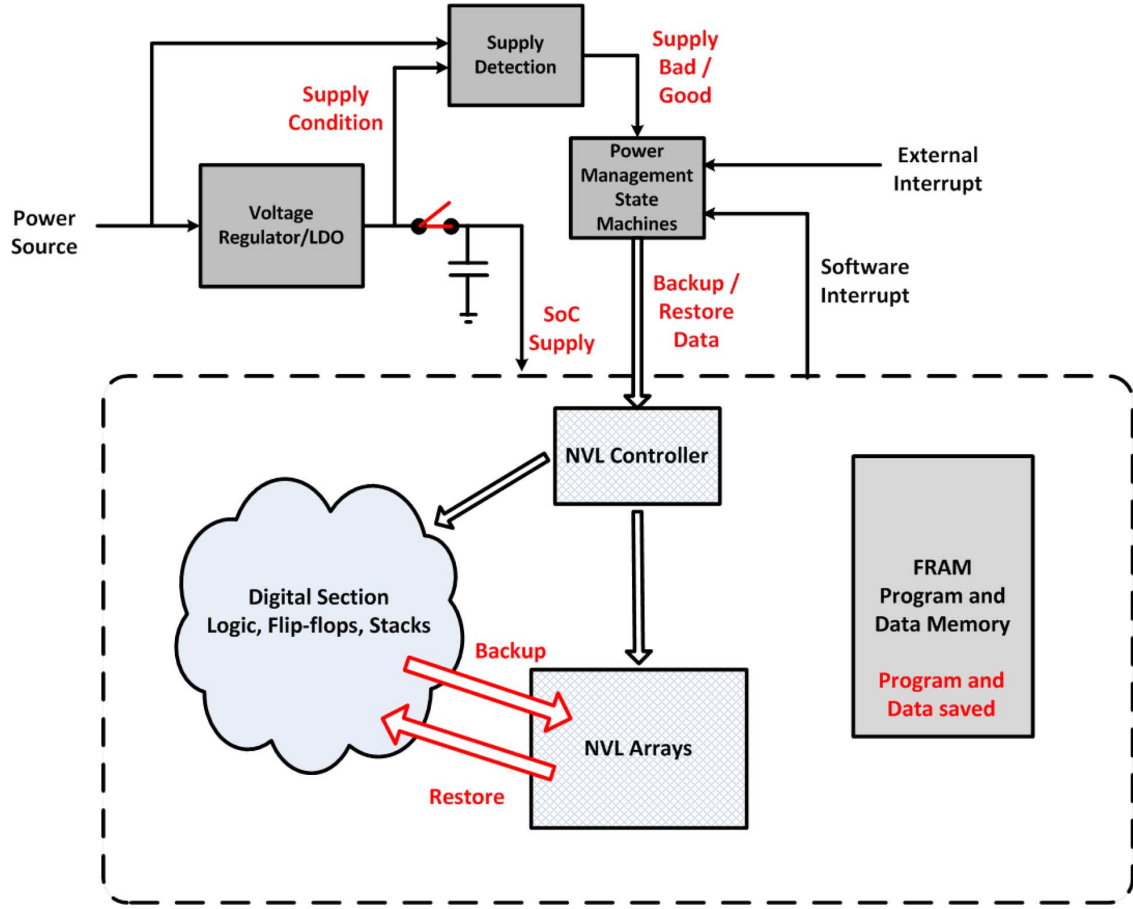


Fig. 2. NVL SoC functional diagram. For this version of the NVL SoC, voltage regulation, supply detection and PM state machines are off-chip.

future SoC would have a package based capacitor implementation. On the other hand, if a system only has planned software based power interrupts the supply capacitor size would be determined by conventional dynamic IR drop requirements. Then the SD activates the power management (PM) state machines with a supply bad signal. Following this, data from all flip-flops and latches in the SoC are backed up into NVL arrays. Since the program and data memory are already nonvolatile, no action needs to be taken for the FRAM. In this way, the entire system state is saved before power loss. In the case of software based power interrupt, the same sequence is followed, except that the supply switch is turned off last. The NVL controller orchestrates data backup and restore, and controls communication between the system flip-flops and the NVL arrays. When power is resumed, following a similar procedure, data from NVL arrays is restored back into the system flip-flops, and system operation beings from where it left off. No boot-up is required.

Though we have embedded SRAM, it is only used for a FRAM test in this SoC. SRAM could be used in future NVL SoCs during normal operation if it is configured as a write-through or read-only cache. The SRAM data would not be backed up using NVL. By invalidating the “valid bits” while waking up from a sleep mode, the SRAM can be used in an NVL SoC to improve memory speed and power consumption during normal operation, while having NVL functionality.

The key metrics of an NVL SoC implementation are:

- 1) reliability: the added nonvolatile bits must be highly robust and testable for signal margin;
- 2) system-level area overhead: NVL must not have a huge area overhead;
- 3) NVL energy cost and wakeup time: if an NVL implementation takes as much time/energy to wake-up as the boot-up code then it doesn't have much benefit;
- 4) regular operation metrics: NVL should not impact regular system performance or power;
- 5) process and design complexity: additional masks, high-voltage supplies add to SoC cost.

To make NVL compatible with CMOS processing, ferroelectric nonvolatile elements were chosen rather than flash. Flash requires a dozen extra masks as compared with only two for ferroelectric nonvolatile elements. Also, ferroelectric nonvolatile elements require a single 1.5-V supply for read and write. No charge pump or voltage boosting is required.

All NVL implementations reported so far [4]–[8], propose adding nonvolatile elements at the flip-flop level. Another possible architecture is to use a large centralized NV-RAM array to backup all flip-flop data. Fig. 3 shows these architectures conceptually. Flip-flop-level NVL implementations [7] have the potential to achieve the lowest wakeup time because each nonvolatile element sits next to the flip-flop it backs up. However, nonvolatile elements are typically qualified to yield well in the protected and carefully studied environment of a

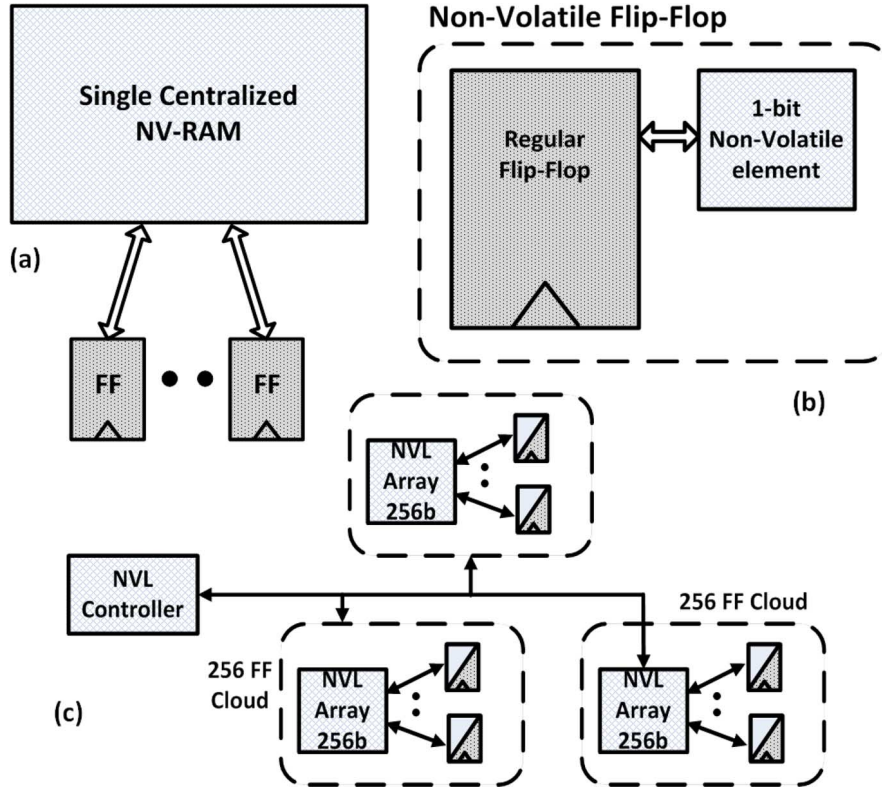


Fig. 3. Possible NVL architectures. (a) A centralized architecture where a single large NVRAM backs up all FF data. (b) Flip-flop-level NVL implementation where a nonvolatile element is added to each FF. (c) Our hybrid mini-array architecture.

memory array. Dispersing nonvolatile elements (like flash bits or FeCaps) throughout the digital logic adversely impacts the yield, since the surroundings of the FeCaps in active and metall layers are not predictable. To make things worse, parity and ECC cannot be implemented since the nonvolatile elements are not arranged in a regular array fashion. Further, read and write infrastructure for the nonvolatile elements has to be replicated in each flip-flop making the NVL area overhead large. For example, the work in [7] has 90% area overhead due to an NVL flip-flop that is five times the size of a regular flip-flop. Further, low-resistance analog reference voltages required for NVL test, and dedicated NVL supply voltages [8] have to be routed throughout the digital section of the chip to every flip-flop.

The other option is to use a large centralized NVRAM to backup the flip-flop data as shown in Fig. 3(b). For example, it is possible to use part of the FRAM that already exists in our MCU SoC to implement NVL. Such an implementation helps save area by using a dense memory array to backup flip-flop (FF) data. Also, no additional testability or reliability concerns exist because the same FRAM array is used as regular memory and for NVL. However, the primary concern in a centralized NVL implementation is the limited bandwidth a typical NVRAM offers. For example, the FRAM we use in our SoC can read/write 64 b of data at 8 MHz. This translates to a 5- μ s wakeup time for our 32-b MCU (with ~ 2600 FFs). If a more complex MCU core like ARM Cortex M3 or MSP430 is used, the FF count and wakeup time would be four times the above value since the FRAM read/write bandwidth would remain unchanged. In contrast our NVL SoC has a wakeup time of 400 ns, and this number

is independent of the flip-flop count. The limited bandwidth and unscalable nature of a centralized scheme makes it impractical.

In our NVL SoC, a hybrid approach is followed [Fig. 3(c)]. We use custom 256-b mini-arrays of FeCap-based bitcells to backup FF data. Multiple such arrays are located in the digital area with standard cells surrounding them in a doughnut-like shape. For larger SoCs, the preferred implementation would be to disperse the mini-arrays throughout the logic cloud with each array serving 256 neighboring FFs. This hybrid approach allows us to amortize read/write circuitry over 256 b. An analog reference and the NVL power supply only need to be routed to the arrays, and a solid mesh distribution throughout the chip is not needed. Testability and bit-level screening features applicable to a large NVRAM array can be exactly implemented for such a mini-array. This is critical for making a viable product. Also, being in a relatively well-controlled array environment drives bit yield higher. Thus, the mini-array architecture serves as a healthy compromise between a centralized NVRAM and FF-level NVL implementation.

A. NVL Use Cases and Advantages

For the first generation, we envision two primary NVL use cases. The most common way to use NVL would be to store a boot image one time before deployment into NVL arrays and to restore this image to system FFs every time the SoC wakes up. In this use case, NVL is used as a boot replacement. The second use case is "compute through power loss," where before a power interrupt FF data is backed up into NVL arrays and is restored after power is back, enabling computation to start *exactly* where

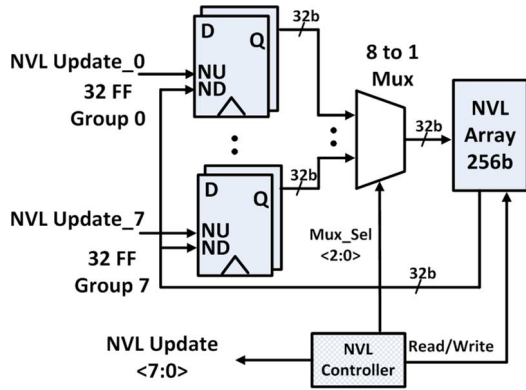


Fig. 4. NVL array connects to 256 NVL FFs using an 8:1 multiplexor. The NVL controller synchronizes the data transfer from and to the NVL array to NVL FFs.

it left off. An NVL SoC offers the following benefits over traditional SoCs.

- 1) Zero leakage replacement for retention FFs: Retention FFs and associated LDOs need to be powered during sleep mode to retain the system state. In an NVL system, these can be turned off, lowering standby power from the 2537 FF at 125C in our SoC from 1.7 μ A to zero.
- 2) Enabled use LVT instead of HVT: Current ULP MCU designs use HVT FFs [2] to lower current consumption of the “always on” retention FFs. However, using HVT limits the maximum performance of the design or limits the minimum operating voltage. With NVL, the entire logic domain can be turned off. This helps decouple leakage and performance. SoCs can now use LVT logic throughout the design, improve performance and/or operating voltage and also avoid the design complexity of a mixed V_t design.
- 3) Fast wakeup time: NVL reduces system wake-up time from 100’s of microseconds to a few microseconds by eliminating boot-up.
- 4) Ease of use: ULP MCU SoCs implement multiple sleep modes that tradeoff wakeup time, standby power, and design volatility. With NVL, we achieve low leakage, fast wakeup time and 100% state retention with a minimum number of sleep modes. Ease of use is not quantifiable but is a large advantage for products.

III. NVL FLIP-FLOP DESIGN

Here, we discuss how data are transferred between NVL arrays and flip-flops in the SoC. To enable fast backup and restore, NVL arrays have direct access to the system flip-flops. Fig. 4 shows how a group of 256 FFs connect to an NVL array. The NVL array is organized into eight rows of 32 b each. Thus, FFs are grouped into bunches of 32. Each 32-b FF group writes to and reads from a row in the NVL array. The NVL controller shown in Fig. 5 controls the select pins of the write multiplexor and the control signals of the NVL array. The eight rows of the NVL array are sequentially written under the control of the NVL controller.

Conventional flip-flops have a single data-input port. To insert data from NVL arrays back into a flip-flop, we added a

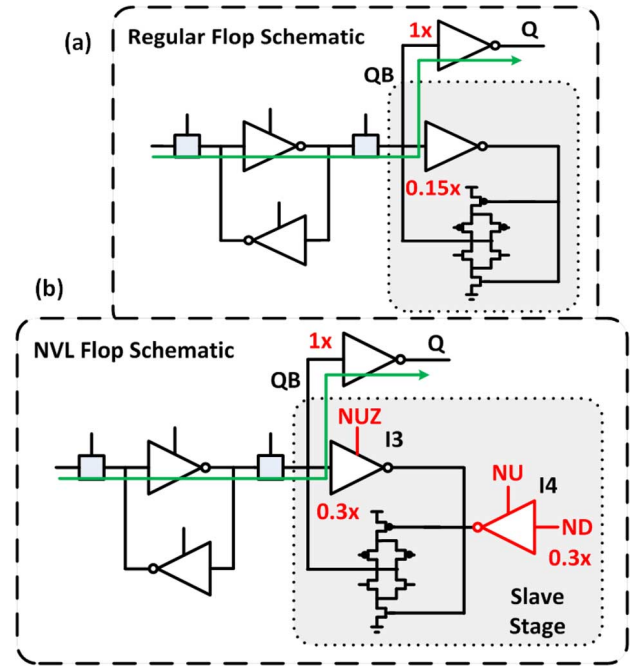


Fig. 5. Flip-flops need to be modified to be NVL compatible. (a) Conventional retention FF with shaded slave latch that is powered by a separate “always on” supply. (b) The NVL FF is built on top of a retention FF.

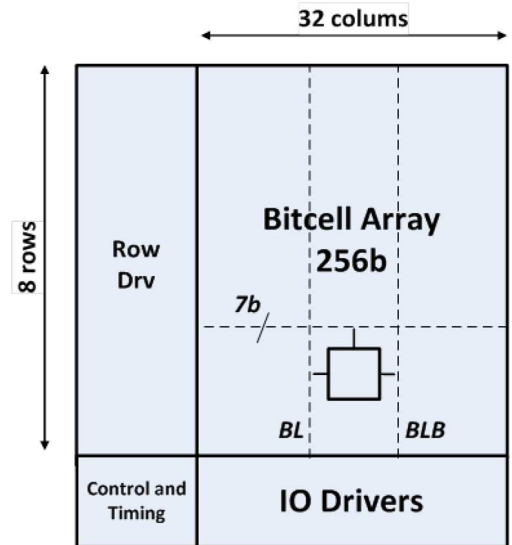


Fig. 6. NVL array architecture.

second data-input port (ND) and an enable port (NU) to a regular FF. As shown in Fig. 4, the 32-b data output bus of the NVL array is connected to all eight FF groups. When row 0 of the NVL is being read, NU ports of the FF group 0 are enabled. In this fashion, all eight rows are sequentially read. Again, the NVL controller generates the NU signals and synchronizes data transfer between the NVL array and FFs. We call the modified flip-flop an NVL flip-flop.

Fig. 5(a) shows the schematic of a retention flip-flop. The difference between a regular FF and a retention FF is that the slave latch (shaded) of a retention FF is powered by a separate supply. This allows the primary logic supply to be turned off during sleep mode to save leakage power. At the same time, the state

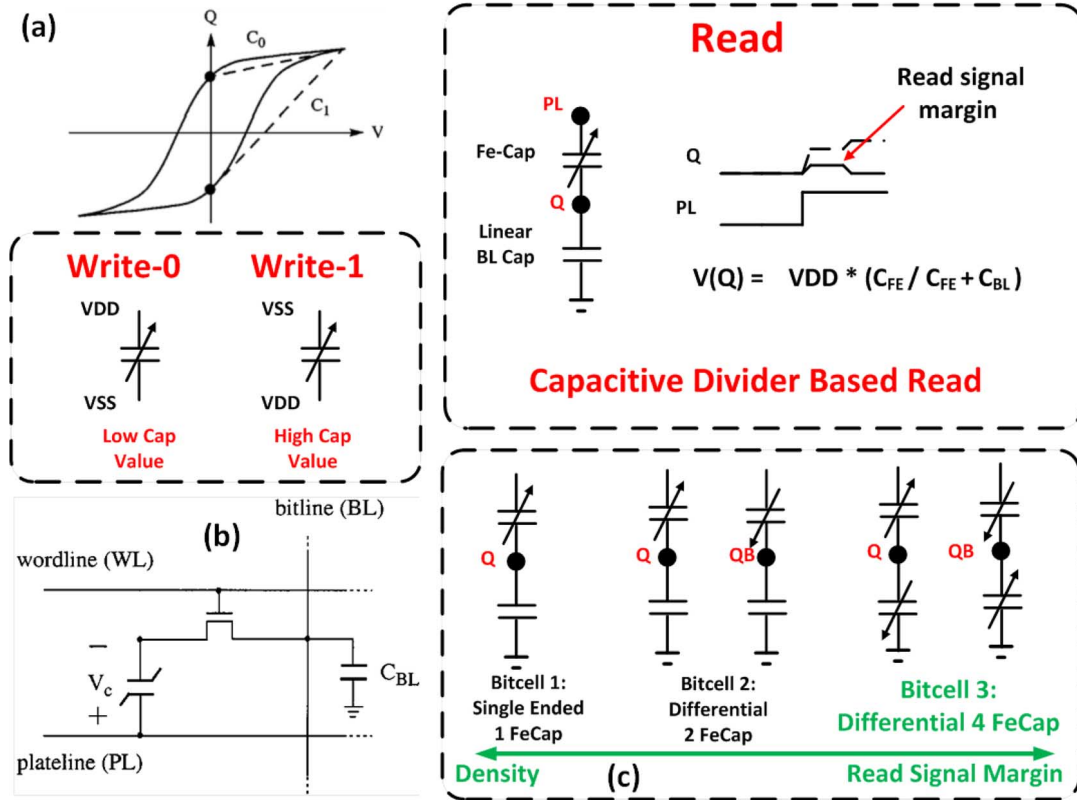


Fig. 7. Ferroelectric capacitor behavior and FRAM bitcells. (a) FeCap Q - V curve showing hysteresis. FeCap capacitance depends on the last state of polarization. (b) A common FRAM bitcell. Reading such a bitcell involves a capacitor divider. (c) Different FRAM bitcells with increasing area and read signal margin.

of the FF is retained by only keeping the minimum-sized slave latch powered. The NVL FF builds on top of a retention FF. Two new ports ND and NU are added to the slave stage. During regular operation, $NU = L$ and $ND = X$. Since I4 is turned off and I3 is on, the NVL FF behaves like a regular retention FF. After coming back from a sleep mode, data must be transferred from the NVL array into the NVL FF. To accomplish this, NU is pulsed high for a cycle while ND has valid data. RET is kept high at this time to allow ND to propagate to QB (CLK maybe X because the logic domain may be off while restoring data). When NU goes low, the latch closes and ND is latched as QB. Since inverted version of ND is available at Q, an inverter is placed in the path between the NVL array and the NVL FF. We carefully added circuits only to the slave latch so that they are not on the critical path and are minimum sized. The NVL FF is 19% larger, 2% slower and has 7% higher power than a retention FF. At the SoC level, impact on performance and power is not even perceptible. Moreover, these numbers are for minimum-drive-strength FF's. For higher drive strengths these NVL overhead values are even lower.

IV. NVL BITCELL AND NVL ARRAY DESIGN

The NVL SoC uses 256-b FRAM-based mini-arrays to backup FF data. Implementing ECC is impractical in such a small array. Hence, we use a custom bitcell that has much higher signal margin than a regular FRAM bitcell [3]. Fig. 6 shows the array architecture. The NVL array consists of the bitcell array, row drivers, IO circuits, and a control and timing block.

A FeCap is a nonlinear capacitor with hysteretic behavior, as shown in Fig. 7(a). It retains the charge even when voltage across its terminals is zero. The next time a voltage is applied across its terminals, the capacitance depends on the polarity of charge retained. This forms the basis of nonvolatility. Reading an FRAM bitcell involves activating a capacitor divider and sensing the mid-rail voltage that develops at the intermediate node as shown in Fig. 7. Fig. 7 also shows how the read signal margin increases as the bitcell is made increasingly complex. Bitcell 1 uses a single FeCap and the BL cap as a load. A reference voltage is used by the sense amplifier to sense the mid-rail voltage at Q. Bitcell 2 is a differential version of bitcell 1 and gives twice the margin. Bitcell 3 uses a FeCap as the load rather than the BL cap. This doubles the margin again because now data is stored in all four capacitors in complementary fashion. We use the configuration shown in bitcell 3 since robustness is critical to our design. In future generations we would use bitcell 1 or 2 to lower area and read/write power.

Fig. 8 shows the NVL bitcell. C1-2 and C1-2' are the four complementary FeCaps. A sense amplifier (SA) is embedded inside the bitcell. This is done for enhanced read signal margin. If the SA is placed outside the bitcell like an SRAM, the analog voltage developed at Q and QB has to be passed onto the bitlines. Unlike an SRAM, the voltages at Q and QB are based on a capacitor divider and are not actively driven. Thus, as soon as TG opens, charge sharing causes signal loss of 13% as shown in Fig. 9. Since the SA is inside the bitcell, we need to pass a full swing signal onto the bitlines. To enable this, a CMOS pass gate is used rather than an NMOS or PMOS only. Using a CMOS pass

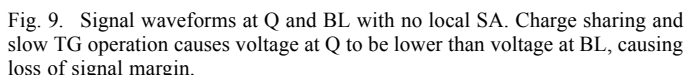
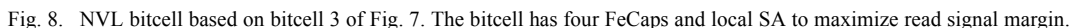


Fig. 10 shows the bitcell read and write waveforms. In the default state, PL1-2 are low, CLR is high, the pass gates are disabled and the SA is disabled. During a read operation, PL1 is pulsed high. This causes voltages to develop on Q and QB according to the C1/C2 capacitor ratio. Next, the SA is enabled which causes Q and QB to develop full rail signals which is eventually transferred to the BL and read latch. The test read is described later in this section.

During a write operation, store drivers in the array periphery are enabled and drive BL and BLB to the correct logic. In the waveform sequence, first PL1 and PL2 are pulsed high. Q and QB couple high due to capacitance of the FeCaps. Next, the pass gates are enabled and Q and QB are driven to the correct logic values. Then the SA is enabled which assists the store driver in holding Q and QB. At this time, the FeCaps on “0” side get written. Then the PL1 and PL2 are driven back low. FeCaps are highly capacitive (~ 150 fF each), so any change on PL1-2

The row drivers drive the seven bitcell control signals described above. The IO driver section shown in Fig. 11 contains the store driver, read latch, scan chain, and test-related hardware. BL is driven into the read latch L1. The read latch is paired with a master latch L2 to make a scan flop. In normal mode, L2 is disabled, and in scan mode, LCH_EN is disabled. The store drivers are used to drive full swing signals on BL and BLB. They receive input from either the input 32 bit data port or from the scan out (SQ) node.

BL disturb TGs are used to insert an analog reference voltage onto bitlines during a test read in order to measure read signal margin. As shown in Figs. 10 and 12, before the PL1 is pulsed high, the reference voltage `vnnode_a` is transferred to the bitline on the “0” side. Then, when the PL1 is pulsed high, the voltage that develops on Q is sum of the reference voltage, and the voltage that would have developed on Q in a normal read. This lowers the differential that the SA sees. The maximum reference voltage with which the read is successful is the read signal margin for that bitcell.

For high-volume production, a fast on-chip test is needed to screen out bad NVL arrays. To make this possible, dedicated hardware to write all 0s/1s is included in the IO drivers. Once the all 0s/1s are written, in the following test read cycle, the parallel test read logic AND and OR gates determine if data was correctly read. This way, a fast on-chip BIST is implemented for the NVL arrays. During such a test, the reference voltage (vfnod_a) is set to an empirical value set to provide sufficient signal margin against packaging and time-dependent degradation. Adding these test and margining circuits is prohibitive at the flip-flop level. Hence, the hybrid min-array approach was found to be most suitable to implement NVL.

The NVL controller generates all of the control signals for the NVL array. The controller uses a synchronous state machine to generate these signals using a separate high-frequency NVL clock. NVL does not require an accurate clock source, and

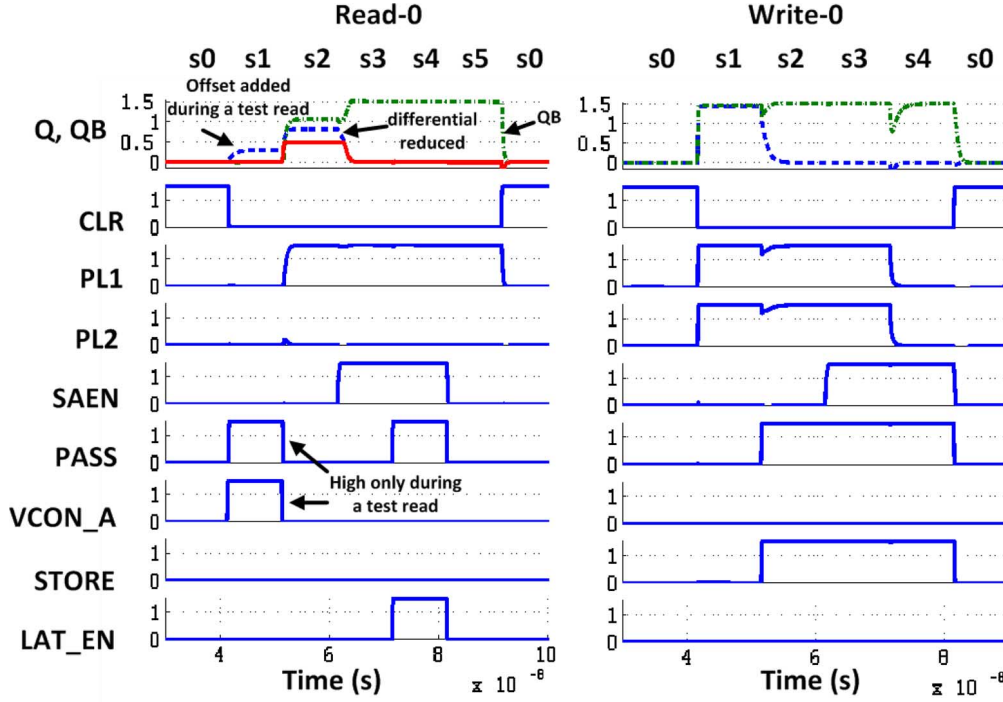


Fig. 10. NVL array read “0” and write “0”. Key states like PL1 high to SAEN high (s2), SAEN high pulse (s3) during read and FeCap dc bias states (s3-4) during write can selectively be made multi-cycle to provide higher robustness. Dotted lines show modification during a test read.

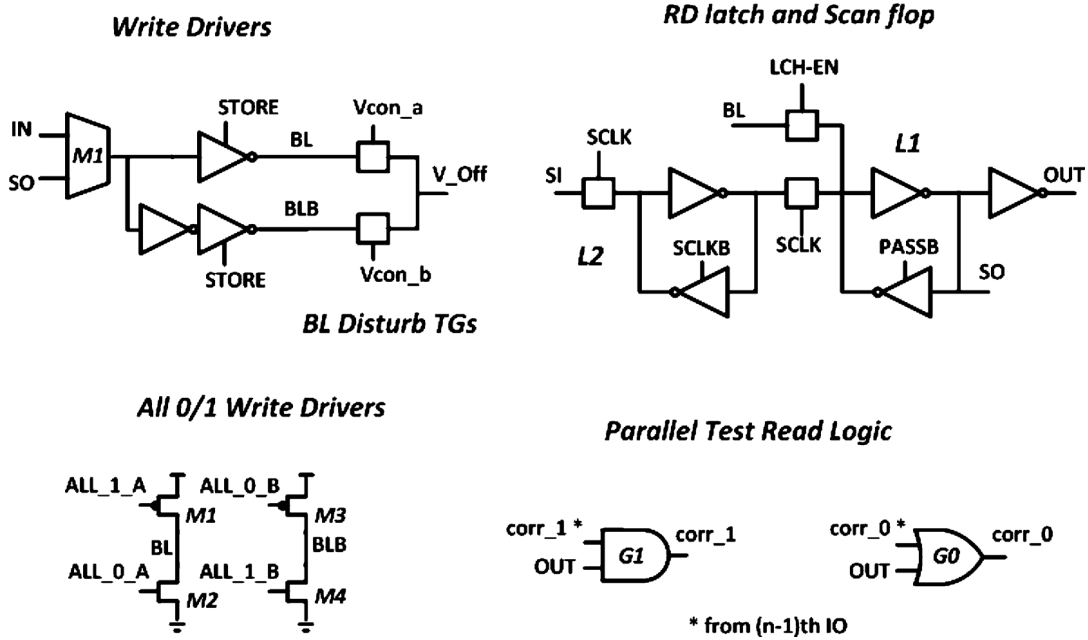


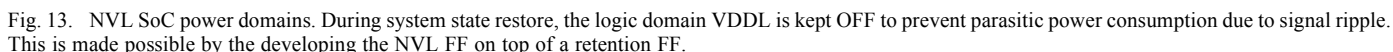
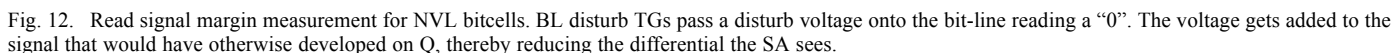
Fig. 11. IO circuits used in the NVL array periphery. Read/write circuits are amortized over eight rows. Scan flop helps provide for a dedicated NVL scan chain. BL disturb TGs, and all 0/1 write and read circuits help implement fast on-chip NVL test. Implementing all of these circuits in every FF is prohibitive, making the mini-array architecture the most appropriate for a practical NVL implementation.

a simple ring oscillator is acceptable. For this chip, however, clock is provided externally. The NVL controller can read/write to all of the NVL arrays in parallel or serial order depending on the peak power/wake-up time tradeoff applicable at the system level. For this SoC, we use a 32×8 array architecture. A maximum of eight rows was chosen to limit wakeup time to less than 500 ns. Thirty-two columns helped get a near-square form factor for the array. Lower limit on array size is set by array

efficiency: at 256 bits, we are at 65%. The higher limit would be set by the wakeup time requirement. For future versions, we plan to make the NVL array compiler-based so that SoC design teams can select NVL array size.

V. NVL SYSTEM DESIGN AND IMPLEMENTATION

NVL can be applied to any digital system, programmable core, ASIC or FPGA. NVL has just two system requirements:



Inserting NVL into a design is simple and design-independent. NVL is inserted into a design in the post-synthesis netlist using a Tool Control Language (TCL) script as the first step after loading the design into a place and route tool. Fig. 13 shows

power domains in our NVL SoC. There are three power domains: VDDN (for NVL controller and NVL arrays), VDDR (for the slave stage of all FFs), and VDDL (for all of the remaining logic, master stage of all FFs, and memories). It shows the power supply conditions during different modes. In functional mode, VDDL and VDDR are on. VDDN can be turned off to save active mode leakage. Before going into sleep mode, VDDN is turned on and all FF data is backed up. Then in sleep mode, all the supplies can be turned off. When the chip comes back from sleep mode, all supplies can be turned on while data is

TABLE I
NVL SoC METRICS

Regular SoC Operation	
Technology Node	130nm, HVT
Clock Frequency	8MHz @ 1.5V
Energy Consumption	75uW/MHz
NVL Sleep Mode Lkg Power	0.000nW
Retention Mode Lkg Power	280nW @ 85C
(only slave latch powered)	1740nW@125C
Area Numbers (mm ²)	
SoC Area	4.4
10 NVL Arrays	0.12
NVL Controller (A)	0.005
NVL Flop overhead:	19% per flop
For 2537 flip-flops (B)	0.02
NVL gate overhead (C)	0.0075
Routing overhead estimate	0.0081
(0.25 * (A+B+C))	
Total overhead due to NVL	0.16
SoC Level Area Impact	3.63%
NVL Operation	
NVL Clock Frequency	125MHz @ 1.5V
Chip Backup Cost (Parallel)	7.2nJ, 320ns
Chip Restore Cost (Parallel)	2.4nJ, 384ns
Breakeven Time (Parallel)	34ms @ 85C
(based on leakage above)	5.45ms @ 125C
Chip Backup Cost (Serial)	19.5nJ, 3200ns
Chip Restore Cost (Serial)	16.9nJ, 3840ns

TABLE II
NVL SoC COMPARISON

Metric	This Work	Wang2012
Impact on regular FF operation	2% Perf 6.9% Power 19% Area	0% Perf No Pwr Data 500% Area
NVL Arch, NVM Element	Mini Arrays, Fecaps	FF Level, FeCaps
SoC Area Ovhd	3.6%	90.0%
Signal Margins	550mV min	NVL bit has no test capability
Backup Time	320ns	7us
Backup Energy/Bit	2.8pJ	14.4pJ
Restore Time	384ns	3us
Restore Energy/Bit	0.95pJ	5pJ

being restored from the NVL arrays to the FFs. However, as data from different rows and different arrays is sequentially restored, logic values ripple through the VDDL domain, as different logic cones may be interconnected. This is parasitic power consumption. Our measurements show that the ripple energy is 1.3 and 5.2 nJ while accessing arrays in parallel and series respectively. To prevent this parasitic power consumption, we use an innovative but simple design technique: we shutdown the logic domain while restoring data back to the FFs. This is made possible because we use retention FFs instead of normal FFs to build NVL FFs. Retention FF have a separate supply for the master and slave latches as shown in Fig. 5. Since NVL arrays only talk with the slave latch, FFs can be restored, and there is no logic ripple because the logic domain is off. Only after all the FFs have been restored, VDDL is turned on. As VDDL rises to 1.5 V, all nodes rise to their final values without ripple. As an additional benefit, building on top of retention FFs allows the SoC to use regular retention instead of NVL for very short sleep times.

VI. RESULTS AND CHIP SUMMARY

Here, we present simulation and measurement results for our NVL SoC. The SoC has been implemented in a 130-nm low-leakage HVT process. It uses a single 1.5-V nominal supply voltage for all blocks in the design. The design was timing closed at 8 MHz at the weak corner (weak process, 1.35 V, 85 °C). The SoC consumes 75 μ A/MHz at 1.5 V while running from SRAM and 170 μ A/MHz while running from FRAM.

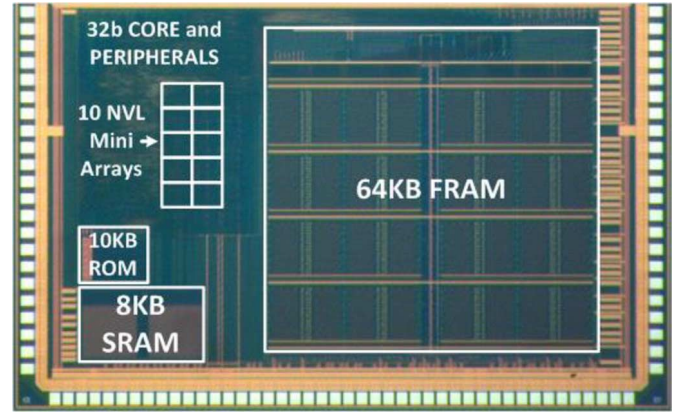


Fig. 14. NVL SoC die photograph.

These numbers were measured while running a program that increments a counter on an LCD display connected to the SoC's SPI interface.

The maximum NVL clock frequency is 125 MHz. Table I reports the chip-level backup and restore costs with all arrays being accessed in parallel/series (including NVL clock and controller energy and FF update energy which was not included in numbers in [9] due to measurement limitations). The high frequency NVL clock and high bandwidth due to parallel NVL array access helps us achieve a sub-400-ns wakeup time. NVL has a one-time energy cost of backing up and restoring FF data. However, without NVL the sleep mode leakage due to the powered retention FF is 280 nW at 85C and 1.74 μ W at 125 °C. This results in a breakeven time of 34 and 5.45 ms, respectively. This is much smaller than typical sleep times in most ULP applications. Moreover, we have not taken into account potential saving made possible by turning off the on-chip LDO that is usually part of an MCU SoC. Table II shows a comparison of this work with the only other NVL SoC reported so far.

Our measurement results show that the minimum NVL bitcell margin is 550 mV, which is almost double the margins obtained from a comparable FRAM bitcell (type 2 in Fig. 7). Fig. 14

shows the die photograph. Digital logic consumes only a small portion of the overall SoC area, but is the major contributor to standby power. Though NVL increases digital area significantly, the SoC area only grows by 3.6%. However, NVL impact on standby power reduction is significant. Since our SoC only has no power regulation, we can achieve zero standby power. In other example implementations of a complete SoC [2] the Real Time Clock (RTC), and some power management may be always on and consuming power. In such systems, the benefit of NVL in reducing standby power is less, but still achieves a 40% reduction in system standby mode power consumption according to our system level estimates.

VII. CONCLUSION

We demonstrate an NVL SoC that helps achieve zero standby power with a 400-ns wakeup time. The novel mini-array-based architecture helps achieve a practical NVL implementation by minimizing area overhead to 3.6% while ensuring testability of each NVL bitcell. Parallel access to all mini-arrays makes the wakeup time fast and independent of SoC flip-flop count. Bit-cell design decisions help achieve the high signal margin and robustness that is crucial for commercial success. A novel NVL flip-flop design with dual ported slave latch allows use of NVL with $< 2\%$ impact on FF performance. Ripple power is minimized by clever use of retention FFs rather than regular FFs. The system uses a single 1.5-V supply, fully CMOS-compatible process technology, and simple design methodologies. The above system, architecture, and circuit techniques help in realizing a production-ready NVL SoC. By making zero standby power a practical reality, we believe that such an SoC would help enable energy-constrained systems such as energy-scavenging wireless microsystems a practical reality.

REFERENCES

- [1] M. Zwerg, A. Baumann, R. Kuhn, M. Arnold, R. Nerlich, M. Herzog, R. Ledwa, C. Sichert, V. Rzehak, P. Thanigai, and B. O. Eversmann, "An 82 μ A/MHz microcontroller with embedded FeRAM for energy-harvesting applications," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 20–24, 2011, p. 334, 336.
- [2] A. Baumann *et al.*, "A MCU platform with embedded FRAM achieving 350 nA current consumption in real-time clock mode with full state retention and 6.5 μ s system wakeup time," in *Proc. IEEE VLSI Symp.*, Jun. 2013, pp. 202–203.
- [3] T. S. Moise *et al.*, "Demonstration of a 4 Mb, high density ferroelectric memory embedded within a 130 nm, 5 LM Cu/FSG logic process," in *Proc. Int. Electron Devices Meet.*, Dec. 8–11, 2002, pp. 535–538.
- [4] S. Masui, W. Yokozeki, M. Oura, T. Ninomiya, K. Mukaida, Y. Takayama, and T. Teramoto, "Design and applications of ferroelectric nonvolatile SRAM and flip-flop with unlimited read/program cycles and stable recall," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 21–24, 2003, pp. 403–406.
- [5] J. Wang, Y. Liu, H. Yang, and H. Wang, "A compare-and-write ferroelectric nonvolatile flip-flop for energy-harvesting applications," in *Proc. Int. Conf. Green Circuits Syst.*, Jun. 21–23, 2010, pp. 646–650.
- [6] S. Yamamoto, Y. Shuto, and S. Sugahara, "Nonvolatile flip-flop using pseudo-spin-transistor architecture and its power-gating applications," in *Proc. Int. Semiconductor Conf. Dresden-Grenoble*, Sep. 24–26, 2012, pp. 17–20.
- [7] Y. Wang, Y. Liu, S. Li, D. Zhang, B. Zhao, M. F. Chiang, Y. Yan, B. Sai, and H. Yang, "A 3 μ s wake-up time nonvolatile processor based on ferroelectric flip-flops," in *Proc. ESSCIRC*, Sep. 17–21, 2012, pp. 149–152.

- [8] M. Qazi, A. Amerasekera, and A. P. Chandrakasan, "A 3.4 pJ FeRAM-enabled D flip-flop in 0.13 μ m CMOS for nonvolatile processing in digital systems," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 17–21, 2013, pp. 192–193.
- [9] S. C. Bartling, S. Khanna, M. P. Clinton, S. R. Summerfelt, J. A. Rodriguez, and H. P. McAdams, "An 8 MHz 75 μ A/MHz zero-leakage non-volatile logic-based Cortex-M0 MCU SoC exhibiting 100% digital state retention at $V_{DD} = 0$ V with < 400 ns wakeup and sleep transitions," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 17–21, 2013, pp. 432–433.



Sudhanshu Khanna (M'13) received the B.Tech. degree in electronics and communication engineering from Delhi College of Engineering, Delhi, India in 2005, and the M.S. degree from the University of Virginia, Charlottesville, VA, USA, in 2011. His thesis was on low-power SRAM design.

He is currently a Senior Design Engineer with Texas Instruments, Dallas, TX, USA. From 2005 to 2008, he was a Design Engineer with Texas Instruments, India, where he worked on IO circuit design. His research interests include SRAM design,

FRAM design, and low-power system design methodologies. In his current role at Texas Instruments, Dallas, he is working on nonvolatile processors that can retain their state across a power interrupt and wake up instantly without needing to boot-up. He has authored and coauthored over 20 peer-reviewed publications and has filed 33 U.S. patents.

Mr. Khanna was the winner of the Student Design Contest at IEEE ISSCC/DAC 2011 and was elected Texas Instruments Innovator in Action for 2013.



Steven Bartling received the B.S.E.E. degree from the University of Texas at Austin, Austin, TX, USA, in 1985, and the M.S.E.E. degree from the Georgia Institute of Technology, Atlanta, GA, USA, in 1987.

He is currently the Technology Development Manager (SMTS) with the MCU Business Unit at Texas Instruments, Dallas, TX, USA. His current research interests include ultralow-power logic and memory circuits, processor architectures, and system design. In the past, he has held key roles in the development of RISC 29000, PowerPC, and x86

compatible CPU's. He is inventor or coinventor on 54 U.S. and international patents.



Michael Clinton (M'13) received the B.S.E.E. degree from Lehigh University, Bethlehem, PA, USA, in 1982, and the M.S.E.E. degree from the University of Vermont, Burlington, VT, USA, in 1991.

In 1982, he joined IBM's Microelectronics Division, Essex Junction, VT, USA, and worked on dynamic memory (DRAM) circuit design. While at IBM, he worked on DRAM designs ranging in density from 64 k to 512 Mb. In 2002, he joined Texas Instruments (TI), Dallas, TX, USA, where he was involved with SRAM compiler designs in the

65-nm, 45/40-nm, and 28-nm technologies. During this time, his work focused on power management and read/write assist circuit techniques. While at TI, he was also involved with FRAM design for both high-speed applications and compilers. In 2013, he joined TSMC, Austin, TX, USA, where he is currently a Deputy Director working on high-speed memory design.

Mr. Clinton is a member of the IEEE Solid-State Circuits Society. He is currently a member of technical program committee for the ISSCC and was previously a member of the Symposium on VLSI Circuits technical program committee. He has presented papers or participated in tutorials, forums and panels discussing SRAM design and variability.



Scott Summerfelt received the B.S. degree in metallurgical engineering and the B.S. degree in physics and mathematics from Iowa State University, Ames, IA, USA, both in 1984, and the M.S. and Ph.D. degrees in materials science and engineering from Cornell University, Ithaca, NY, USA, both in 1990. His dissertation focused on interfaces and microstructures in spinel-wustite systems.

He is a Fellow with the Analog Technology Development (ATD) Group, Texas Instruments, Dallas, TX, USA. In 1990, he joined Texas Instruments, Dallas, TX, USA, to work on (Ba, Sr)TiO₃ for DRAM and later Pb(Zr, Ti)O₃ for FRAM. His research interests have focused on high-dielectric and ferroelectric materials for memory applications including electrical properties, electrode materials, processing, design rules, cell design, ferroelectric spice models, and integration issues. He has helped create and optimize the FRAM process integration approach since its beginning in 1996 when it started production in 2007. He has authored and coauthored over 34 refereed and 40 proceeding publications, two book chapters, and 159 U.S. patents.



John A. Rodriguez (M'94) received the Ph.D. degree in electrical and computer engineering from Rice University, Houston, TX, USA, in 1999.

He is a Distinguished Member of the Technical Staff in Analog Technology Development, Texas Instruments, Dallas, TX, USA, where he researches the reliability physics of ferroelectric memory devices embedded within advanced CMOS processes. He has coauthored over 30 conference and journal publications and holds 18 U.S. patents.

Dr. Rodriguez was a recipient of the 2002 IEEE International Reliability Physics Symposium Outstanding Paper Award, the 2005 EOS/ESD Symposium Best Presentation Award, and the 2012 Semiconductor Research Corporation Mahboob Khan Outstanding Industry Liaison Award.



Hugh McAdams received the B.A. degree in physics from Rice University, Houston, TX, USA, in 1968.

He is a Distinguished Member Technical Staff with the Analog Technology Development (ATD) Group, Texas Instruments (TI), Dallas, TX, USA. He joined TI in 1970 and designed DRAMs until TI exited the memory business in 1998. He later rejoined TI in 2000 and began working on the design of ferroelectric memories. He has helped define the architecture and design many of the embedded FRAM macros used in various business units within

TI. He is an author of several refereed and proceeding publications on DRAMs and FRAMs and holds over 70 U.S. patents.

Ant Colony Based Load Flow Optimisation Using Matlab

Kapil Upamanyu

kapil.upamanyu@gmail.com

Keshav Bansal

keshav.dtu@gmail.com

Miteshwar Singh

miteshwar@yahoo.com

Department of Electrical Engineering

Delhi Technological University, Shahbad Daultapur, Main Bawana road, New Delhi-110042, India

Abstract: This paper presents a solution to the network constrained optimisation problem using Ant Colony Optimisation (ACO) algorithm. This algorithm consists of artificial agents, called ants, which cooperate among themselves to find an optimal solution to the Constrained Load Flow (CLF) problem. These ants communicate with each other using pheromone matrix, pheromone being a chemical released by real ants for finding the shortest path from food source to their nest. The study revolves around finding the optimum settings of three tap changing transformers, which gives minimum power losses while maintaining a constant demand. The above method is applied to the standard IEEE 30-bus system and results are compared with the solutions obtained from conventional methods. The MATLAB code is developed for the same, and compared with the conventional approach.

Index Terms – Ant colony Optimisation (ACO), Constrained Load Flow (CLF), Genetic Algorithms (GA)

I. INTRODUCTION

The CLF problem deals with the offline adjustment of the power system control variables in order to satisfy physical and operating constraints [1]. Every year an immense amount of power is lost due to transmission, generation and distribution imperfections and non-idealities. The objective is to minimise the transmission losses by changing the tap setting of switching transformers. This gives us a constrained load flow problem to which ACO optimisation is applied.

Various methods can be followed for this purpose, conventional being testing all possible combinations of tap settings of transformers and finding the combination which corresponds to minimum power losses. But this method is associated with a large computational time and a need for a large computational memory.

Genetic Algorithms (GA) offer a great advantage of less computational time. GA has been used to evolve computer programs for specific tasks, and to design other computational structures [2]. GA involves heuristic methods such as swarm particle optimisation and ACO.

Heuristic methods do not guarantee the optimal solution to the given problem but the algorithm can be fine tuned so as to obtain a near optimum solution with a high accuracy and low computational time. The significant part of heuristic algorithms comprises metaheuristic methods, which differ from the classical methods in that they combined the stochastic and deterministic composition. It means that they are focused on global optimization, not only for local extremes [5]. The big advantage of metaheuristic is that they are built not only for solving a concrete type of problem, but they describe general algorithm in that, they show only the way, how to apply some procedures to become solution of the problem.

The essential trait of ACO algorithms is the combination of a priori information about the structure of a promising solution with a posterior information about the structure of previously obtained good solutions [3]. This metaheuristic algorithm is developed on the natural behaviour of ants of transporting their food from the source to their nest through shortest path. To find the shortest path, moving ants lay some pheromone on the ground, so an ant encountering a previously trail can detect it and decide with high probability to follow it. As a result, the collective behaviour that emerges is a form of a positive feedback loop where the probability with which an ant chooses a path increases with the number of ants that previously chose the same path [4].

In power systems, the ACS has been applied to solve the optimum generation scheduling problems [6], [7] and the optimum switch relocation and network reconfiguration problems for distribution systems [8], [9]. It is rather difficult to find a single search space, configuration, and a parameter set of an ACS algorithm that can satisfy every optimization problem. The ACS algorithm proposed in this paper formulates the CLF problem as a combinatorial optimization problem, implying the settings of the three tap changing transformers are combined to obtain an optimum solution to the CLF.

II. CLF PROBLEM

The load flow problem can be expressed by two sets of non linear equations:

$$\begin{aligned} Y &= g(X, U) \\ Z &= f(X, U) \end{aligned} \quad [10]$$

Where

- Y Vector of nodal power injections
- Z Vector of constrained variables (power flows, reactive Powers of PV buses, etc.);
- X State vector (voltage angles and magnitudes);
Control vector (transformer tap settings)
- U Voltage and power at PV buses

The objective of the Constrained Load Flow problem is to maintain some or all elements of vectors X & Z within given operating limits and hence determine the combination of the control variables which satisfies such limits and provides the system with minimum power losses.

III. BASIC CONCEPTS OF ACS

Ant Colony Algorithms are based on the principle of stimulating the behaviour of real ants. [11][12] ACS finds its basis on the fact that as a group the ants are able to find the shortest path to their destination using simple communication methods. For real life ants the communication medium is the chemical, pheromone which they leave as a trail marker on the ground. The pheromone evaporates over time unless an additional amount is deposited, thus indicating that a greater number of ants prefer this path. In other words the trail with the greatest pheromone levels is indicative of high degree of optimisation.

Consider the minimisation of the CLF Power Loss Problem (S, P, C) where S is the matrix of all possible solutions, P is the objective function to be minimised and C is the matrix of constraints to be adhered to [13]. For the solution of the CLF problem artificial ants create feasible solutions by following all relevant possible routes. The probability of an ant choosing node j is:

$$P(c_{h+1} = j | X_h) = \begin{cases} \frac{F_{ij}(\tau_{ij})}{\sum_{(i,j) \in N_i^k} F_{il}(\tau_{il})}, & \text{if } (i, l) \in N_i^k \\ 0, & \text{otherwise} \end{cases}$$

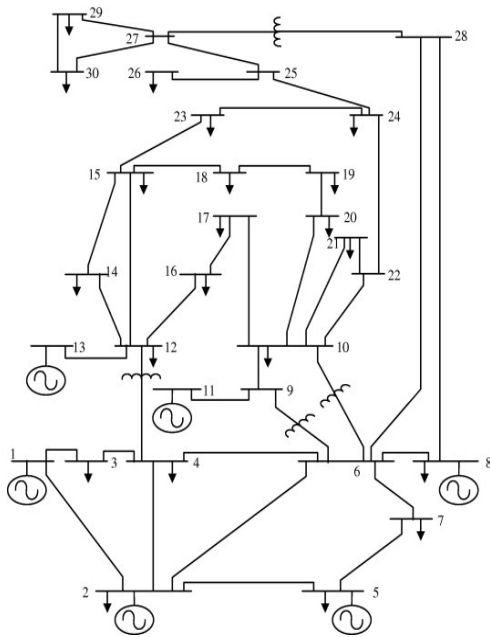
Where (imp) belong to N_i^k satisfying the constraints matrix C.

IV. ANT COLONY ALGORITHM FOR CLF

1. Create the search space that represents the discrete settings (states) of the control variables (Stages)
2. Insert the pheromone matrix according to the nodes of the Ant System Graph where n is number of states and m is the number of stages
3. Initialize the pheromone matrix to a common value
4. Allow M ants to run across the various possible combinations of the states and stages
5. Calculate Power Losses
6. Obtain the control variable settings corresponding to minimum power losses.
7. Update the Pheromone matrix first locally then globally corresponding to tap settings.
8. Repeat for a number of Iterations or till the time convergence occurs
9. Run a set of M new ants in each iteration
10. Display the Result
11. Obtain results 10 times and compare results and their variance.

V. METHODOLOGY USED

In this system we have considered all the parameters fixed, except the three transformer tap settings between line 6 & 9, line 6 & 10, line 4 & 12. The objective is to minimize the power losses associated with this system by changing tap settings of these three transformers. We have used Newton Raphson Method to determine the losses associated with different combinations of settings of these three transformers. Each transformer tap is varied in steps of 16, each step indicating a value of 0.01. The range of turn-ratio varies from 0.9 to 1.05. Hence there are 4096 (16*16*16) possible combinations of transformer settings corresponding to which losses can be calculated. Our task is to find out that combination which gives minimum power loss in the minimum possible iterations with the maximum possible accuracy.



After the initial matrices corresponding to bus voltage and power levels have been defined the pheromone probability matrix is initialised in MATLAB. Initially all pheromone matrix elements are given a constant value and a random ant distribution function is run in the first iteration.

```

for t=1:16 % Loop for running the 16 Artificial Ants
ran=rand; % Using Random function in first Iteration
p=0;
n=0;
for r=1:al %Taking a Random Probability based Value for Tap A
p=p+pa(r)/sum(pa); %Determining Probability for choosing path

    if ran<p && n==0
        sa=r;
        n=n+1;
    end
end
a(sa);
ran=rand;
p=0;
n=0;

```

As the number of iterations increases the coefficient of pheromone updation is correspondingly changed under the assumption that we are nearing the optimum solution and hence greater weightage is given to the results which are obtained in the later stages of the process.

```

if q<5 % No. Of Iterations

```

```

    cp=2000; %Coefficient of Pheromone Updation
else if q<40
    cp=300000; %Coefficient of Pheromone Updation
else
    cp=300000; %Coefficient of Pheromone Updation
end
else if f==0
    cp=270000; %Coefficient of Pheromone Updation

```

After each set of artificial Ants is run and a set of power losses is obtained, the pheromone probability matrix is updated while also taking into account the evaporation matrix.

```

pa=.999*pa; %Evaporation Rate for Pheromone A
pb=.998*pb; %Evaporation Rate for Pheromone B
pc=.998*pc; %Evaporation Rate for Pheromone C
cp;
if q~=1 %Global Updation Loop
    pa(s(1))=pa(s(1))+.1*cp*f; %Updating Pheromone Level of A
    pb(s(2),s(1))=pb(s(2),s(1))+.40*cp*f; %Updating Pheromone Level of B
    pc(s(3),s(2))=pc(s(3),s(2))+.35*cp*f; %Updating Pheromone Level of C
    if f==0 %Updating Pheromone Levels when previous set minimum value again comes as a result
        pa(s(1))=pa(s(1))+.1*cp*.0007;
        pb(s(2),s(1))=pb(s(2),s(1))+.25*cp*.0007;
        pc(s(3),s(2))=pc(s(3),s(2))+.28*cp*.0007;
    end

```

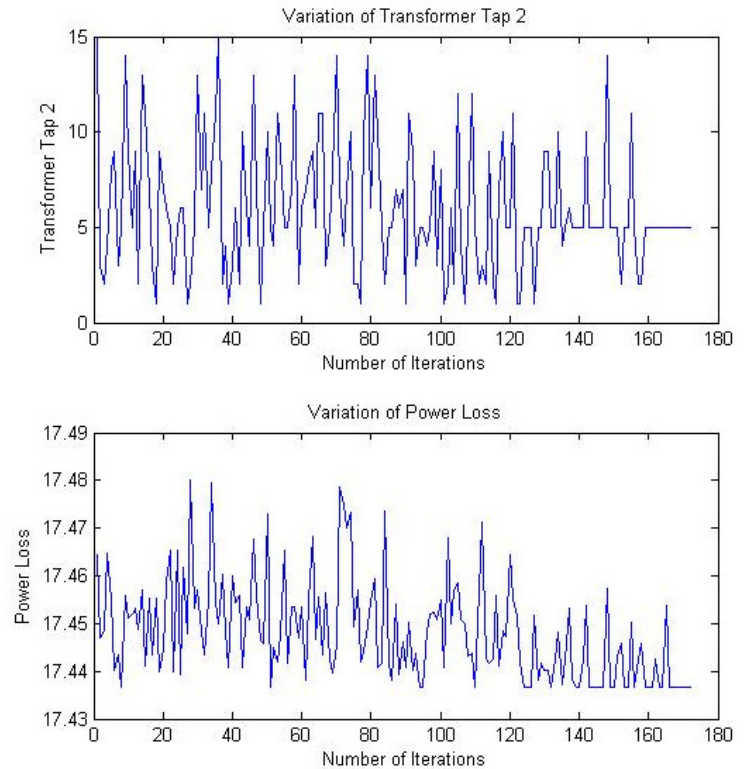
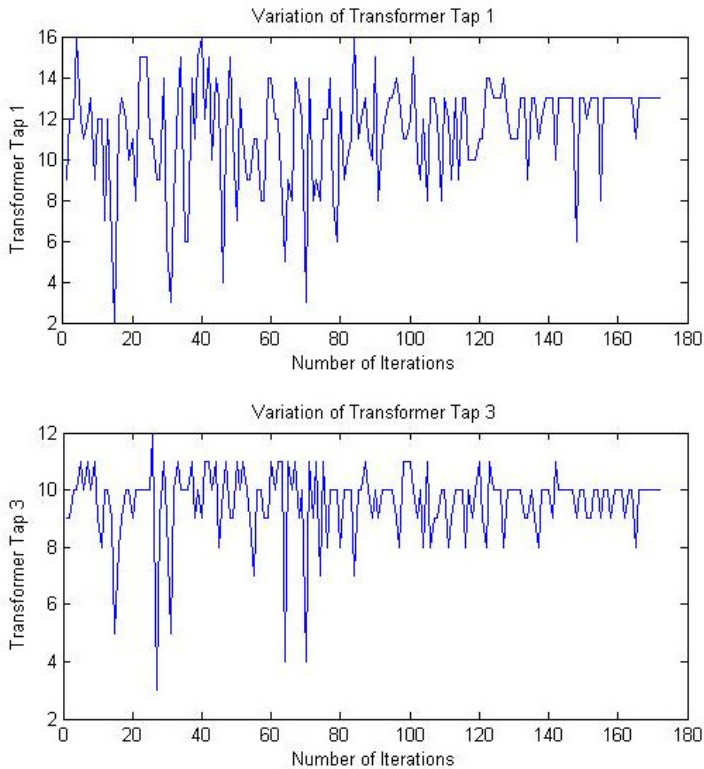
Similar to the constraints which are put on the various parameters of the Load Flow System in case of a fall of pheromone probability levels below a minimum value or a rise above maximum values the matrix components are reset to the standard values.

```

for h=1:al
    if pa(h)<=.01 %If Pheromone A falls below a minimum value
        pa(h)=.01;
    end
    if pa(h)>=500 %If Pheromone A exceeds a maximum value
        pa(h)=500;
    end

```

VI. RESULTS



The Minimum Power Loss obtained for the IEEE 30 Bus System was 17.4367 pu and it corresponded to a tap transformer setting of 1.03 pu on Transformer A, 0.95 pu on Transformer B and 1.00 pu on Transformer C. The process took an average of 174 iterations when run 10 times as compared to 4096 iterations which it would have taken had we applied conventional problem solving algorithms.

VII. CONCLUSION

The procedure implemented gives the minimum power losses combination within an error of 0.05% Power Loss. However we have only considered 3 Tap transformers with 16 discrete levels each. If the number of tap transformers is increased along with more variable elements like Variable Capacitors, variable Generator Buses etc, a greater degree of efficiency is possible. The ACO algorithm enables us to find the optimum solution to the CLF problem in just 174 iterations whereas by conventional methods the Newton Raphson Load Flow analysis would have to be run for the stipulated 4096 times.

However a major flaw with the ACO Algorithm for CLF

of iterations and reduction in accuracy.

REFERENCES

- [1] Ant Colony System-Based Algorithm for Constrained Load Flow Problem John G. Vlachogiannis, Nikos D. Hatziaargyriou, Senior Member, IEEE, and Kwang Y. Lee, Fellow, IEEE
- [2] Ant Colony Optimization Algorithm Nada M. A. Al Salami
- [3] M. Dorigo, Luca M. G., "Ant Colony system: A Cooperative learning approach to the Travelling Salesman Problem", IEEE transaction on evolutionary computation, Vol. 1, No. 1, 1997.
- [4] M. Dorigo and G. Di Caro, "The Ant Colony Optimization metaheuristic", in New Ideas in Optimization, D. Corne et al., Eds., McGraw Hill, London, UK, pp. 11-32, 1999.
- [5] Solving Travelling Salesman Problem using Ant Colony Ivan Brezina Jr. Zuzana Čičková
- [6] I.-K. Yu, C. S. Chou, and Y. H. Song, "Application of the ant colony search algorithm to short-term generation scheduling problem of thermal units," in Proc. Int. Conf. Power Syst. Technol., vol. 1, 1998, pp. 552-556.
- [7] S.-J. Huang, "Enhancement of hydroelectric generation scheduling using ant colony system based optimization

- approaches,” IEEE Trans. Energy Convers., vol. 16, no. 3, pp. 296–301, Sep. 2001.
- [8] J.-H. Teng and Y.-H. Liu, “A novel ACS-based optimum switch relocation method,” IEEE Trans. Power Syst., vol. 18, no. 1, pp. 113–120, Feb. 2003.
- [9] Y.-J. Jeon, J.-C. Kim, S.-Y. Yun, and K. Y. Lee, “Application of ant colony algorithm for network reconfiguration in distribution systems,” in Proc. IFAC Symp. Power Plants Power Syst. Control, Seoul, Korea, Jun. 9–12, 2003, pp. 266–271.
- [10] W. Kellermann, H. M. Zein El-Din, C. E. Graham, and G. A. Maria, “Optimization of fixed tap transformer settings in bulk electricity systems,” IEEE Trans. Power Syst., vol. 9, no. 3, pp. 1126–1132, Aug. 1991.
- [11] M. Dorigo, “Optimization, learning and natural algorithms,” Ph.D. dissertation, Politecnico de Milano, Milan, Italy, 1992.
- [12] M. Dorigo, V. Maniezzo, and A. Coloni, “Ant system: optimization by a colony of cooperating agents,” IEEE Trans. Syst., Man, Cybern. B, vol. 26, no. 1, pp. 29–41, Feb. 1996.
- [13] Phermone Models in Ant Colony Optimisation(ACO) E Foundas, A Vlochas Department of Informatics University of Piraeus Greece

Application of Adaptive Neuro-Fuzzy Inference System for Predicting Software Change Proneness

Akshit Peer

Department of Computer Engineering
Delhi Technological University
Delhi – 110 042, India
akshit.peer@gmail.com

Ruchika Malhotra

Department of Computer Engineering
Delhi Technological University
Delhi – 110 042, India
ruchikamalhotra2004@yahoo.com

Abstract—In this paper, we model the relationship between object-oriented metrics and software change proneness. We use adaptive neuro-fuzzy inference system (ANFIS) to calculate the change proneness for the two commercial open source software systems. The performance of ANFIS is compared with other techniques like bagging, logistic regression and decision trees. We use the area under receiver operating characteristic (ROC) curve to determine the effectiveness of the model. The present analysis shows that of all the techniques investigated, ANFIS gives the best results for both the software systems. We also calculate the sensitivity and specificity for each technique and use it as a measure to evaluate the model effectiveness. The aim of the study is to know the change prone classes in the early phases of software development so as to plan the allocation of testing resources effectively and thus improve software maintainability.

Keywords—ANFIS; bagging; change proneness; logistic regression; random forest; receiver operating characteristic (ROC) curve; sensitivity; specificity.

I. INTRODUCTION

The software maintenance phase is one of the most important, expensive and time consuming of all the phases in the software development life cycle. The maintenance phase consumes nearly 40% to 70% of the total cost spent during the entire lifecycle of the software [1]. Due to this reason, software maintainability is still considered a challenging task in the software industry. Software maintainability means the ease with which a software system or component can be modified to correct faults, improve the performance or adapt to a changed environment [2]. The requirement to change the software may arise in order to correct the discovered faults, adapt to a changed environment, improve the software performance or enhance the understandability of the source code. As the duration and cost involved in maintaining the software is very high, it becomes essential to monitor the software metrics right from the design phase. This will help the developer in improving upon the various aspects of software development such as design, coding and

implementation which could otherwise lead to increased maintenance cost in the later part of software development life cycle. In the literature, a variety of software metrics [3-7] have been proposed which can be effectively used to predict the software maintenance behavior. In fact, the object-oriented software metrics are becoming increasingly popular due to the resulting ease of maintenance and reduced maintenance effort. Moreover, theoretical studies have shown that there is a strong evidence of link between object-oriented software metrics and their maintainability.

Recently, there has been a significant interest on the use of statistical methods and artificial neural networks (ANN) for predicting the software maintenance effort. In this paper, we make the use of adaptive neuro-fuzzy inference system (ANFIS) to predict the software maintainability in terms of change proneness. ANFIS is a Sugeno-based fuzzy inference system and it integrates the principles of neural networks as well as fuzzy logic into a single framework. The inference system corresponds to a set of fuzzy IF-THEN rules with learning capability to approximate the non-linear functions. The functions are tuned using either a backpropagation method alone or in combination with the least squares type of method [8]. Change proneness can be understood as the probability that the particular part of the software would change. The change in the software is calculated in terms of the number of lines added, deleted or modified in the recent version classes with respect to the previous version classes. We employ the present model to predict the change proneness for two open source software using the object-oriented design metrics suite given by Chidamber and Kemerer [4]. We compare the results of ANFIS with other machine learning methods like logistic regression, bagging and random forest. To the best of our knowledge, ANFIS gives the best results for predicting change proneness as compared to the other machine learning methods reported in the literature till date.

II. RESEARCH BACKGROUND

In the literature, several techniques have been proposed for predicting the software change proneness. In 2007, Sharafat and Tahvildari [9] proposed a probabilistic model in which they used the change history as well as the source code of the software to determine the probability of change. In order to account for the probability that the change would propagate

from one class to another, they analyzed the dependencies obtained from the UML diagram of the system. Han et. al [10] used behavioral dependency measurement (BDM) for predicting the change proneness in UML 2.0 models. The correlation between the change coupling (interdependence between various artifacts of software) and defects for three large software systems was analyzed by Ambros et al. [11]. Chaumon et al. [12] proposed a change impact model in order to investigate the impact of changes made to the classes of a system. According to their model, impact of change depends on the type of change as well as on the type of link (inheritance, aggregation) between the classes. They concluded that the object-oriented design metrics are important in determining the software changeability. Zhou et al. [13] based their study on three size metrics viz. source lines of code (SLOC), number of methods implemented in a class (NMIC) and the sum of the number of parameters implemented in class (NumPara). They analyzed the confounding effect of the class size on relationship between object-oriented metrics. A confounding variable can be defined as the one which leads to misjudgment of the actual relationship between the dependent and independent variable. In spite of growing interest for relationships between object-oriented metrics and change proneness among the researchers, Lu et al. [14] suggested that there was a need to reexamine such a relationship. This is mainly due to the fact that most of the studies on change proneness use small number of object-oriented metrics that too on relatively fewer systems. In view of this, it becomes relatively difficult to generalize the relationship between object oriented metrics and change proneness for all the systems.

Apart from the methods discussed above, many soft computing techniques have been proposed and employed by the researchers to predict the software maintenance effort. It is to be noted that the soft computing approaches are gaining attention of researchers mainly because of their ability to predict the maintainability with accuracy and precision. Yuan et al. [15] and Gyimothy et al. [16] used fuzzy subtractive clustering and regression respectively to predict the fault proneness of the software. Aggarwal et al. [17] also developed a fuzzy model for predicting the software maintainability. Malhotra and Khanna [18] evaluated and compared the performance of machine learning methods with statistical methods using ROC analysis. The study, based on three open source software, showed that the results of machine learning methods were comparable to statistical techniques. Kaur et al. [19] first suggested the use of ANN, FIS and ANFIS to predict the software maintainability in terms of continuous dependent variable. However, none of the machine learning methods discussed above employs FIS or ANFIS for predicting the software change proneness. In this paper, we demonstrate the use of ANFIS on metric suite of Chidamber and Kemerer for predicting software change proneness.

III. SOFTWARE METRICS

In this section, we introduce the object-oriented metrics used in the prediction and model evaluation. The independent variables i.e. the object-oriented metrics used in the present study were first proposed by Chidamber and Kemerer [4]. The each metric of the suite is briefly described in Table I.

TABLE I. INDEPENDENT VARIABLES

Metric	Definition
Coupling between objects (CBO)	A count of the number of other classes to which it is coupled and vice versa.
Response for a class (RFC)	A count of methods implemented within a class and the number of methods accessible to an object class due to inheritance.
Lack of cohesion (LCOM)	The number of pairs of local methods in a class using no attribute in common.
Number of children (NOC)	The number of immediate subclasses of a class in a hierarchy.
Depth of inheritance (DIT)	The maximum number of steps from the class node to the root of the tree and is measured by the number of ancestor classes.
Weighted methods per class (WMC)	Count of sum of complexities of all methods in a class.
Lines of code (SLOC)	The number of lines of code.

IV. DATASET CHARACTERISTICS

This study makes the use of two datasets consisting of the Chidamber and Kemerer metric suite for the open source software written in Java language. The first dataset, *Frinika*, contains metric data of 248 classes whereas the second dataset, *CheckStyle*, contains the metric data of 693 classes. The values of these metrics have been made available from the Understand your Code [20]. Software maintainability has been predicted in terms of the dependent variable *change proneness* i.e. the probability of occurrence of change after the software release. The change in lines of code for a particular class has been calculated by counting the number of lines of code that are added, deleted or modified in that class. The method of box-plots has been used to remove the outliers from the datasets. Table II and III describe the dataset characteristics in terms of statistical distribution of the seven metrics as they appear in the datasets for both the software systems – *Frinika* and *CheckStyle*.

V. RESEARCH METHODOLOGY

A. Adaptive Neuro-Fuzzy Inference Systems

Adaptive neuro-fuzzy inference system (ANFIS) constructs a fuzzy inference system (FIS) for the given input/output parameters whose membership function parameters are adjusted using either a back propagation algorithm alone or in combination with a least squares type of method [8, 21]. This adjustment allows the Fuzzy systems to learn from the data they are modeling. ANFIS is applicable to the modeling situations where it is difficult to discern what the membership functions should look like simply from looking at data. Rather than choosing the parameters associated with a given membership function arbitrarily, these parameters could be chosen so as to tailor the membership functions to the input/output data in order to account for these types of variations in the data values. For the present analysis, we use

TABLE II. DESCRIPTIVE STATISTICS OF FRINIKA DATASET

Metric	Maximum	75%	Median	25%	Minimum	Mean	Standard Deviation	Skewness
CBO	11	4	2	0	0	2.70	2.61	0.96
RFC	27	6	3	2	0	5.06	5.12	1.96
LCOM	100	60	50	0	0	37.79	30.24	0.01
NOC	0	0	0	0	0	0	0	-
DIT	3	2	2	1	1	1.63	0.59	0.32
WMC	30	9	4	2	0	6.62	6.09	1.69
SLOC	181	56	31	17	3	42.15	35.43	1.59

TABLE III. DESCRIPTIVE STATISTICS OF CHECKSTYLE DATASET

Metric	Maximum	75%	Median	25%	Minimum	Mean	Standard Deviation	Skewness
CBO	1	0	0	0	0	0.16	0.40	1.78
RFC	14	6	3.50	0	0	4.50	5.28	0.98
LCOM	5	2	1.50	0	0	1.67	1.86	0.93
NOC	85	52	6	2	0	25.16	35.23	0.95
DIT	92	7	4	0	0	17.83	36.48	1.75
WMC	10	2	1.50	0	0	2.50	3.78	1.56
SLOC	172	69	41.50	23	1	51.24	36.98	1.17

the generalized bell-shaped membership function (`gbellmf`) of the ANFIS toolbox in MATLAB. The number of membership functions is set to two and the number of epochs is fixed to be 40. The `genfis1` function of MATLAB is used to generate a Sugeno-type fuzzy inference system.

B. Logistic Regression

Logistic regression (LR) is used to predict the dependent variable (change proneness) from a set of independent variables (OO metrics) to determine the percent of variance in the dependent variable explained by the independent variable [22-27]. We try to predict the probability of occurrence of an event by fitting the data into logistic curve. LR is a statistical method for dealing with the formulation of the mathematical model depicting relationship among variables, which can be used for the values of each independent variable. Multivariate LR is done to construct a prediction model for the change proneness of classes. In this method, metrics are used in combination. The multivariate LR formula can be defined as:

$$prob(X_1, X_2, \dots, X_n) = \frac{e^{(A_0 + A_1 X_1 + \dots + A_n X_n)}}{1 + e^{(A_0 + A_1 X_1 + \dots + A_n X_n)}} \quad (1)$$

where X_i , $i = 1, 2, 3, \dots, n$ are the independent variables and $prob$ is the probability of detecting whether the class has changed or not.

C. Bagging

Bagging improves the classification models by creating numerous versions of the training sets. This classification

method trains a new function for each of the various similar training sets. This accumulation standardizes over the versions when we predict a numerical result [28]. In case of predicting a class the result of majority is selected. In order to create multiple versions, we create bootstrap duplicates of the learning set. These sets are then used as the new learning set. Bagging has numerous benefits such as substantial gains in classification accuracy over other regression methods, reduction in variance and non-overfitting of training data. The success of Bagging is due to instability of the prediction method. If disturbing the learning set can cause critical changes in the predictor constructed, then Bagging can give quite good results. Bag size percent of 100, 10 iterations and REP tree as classifier were chosen for our analysis.

D. Random Forest

Random forests (RF) was proposed by Breiman [29] and is a combination of tree predictors such that each tree depends on the value of a random vector. For all the trees in the forest, this random vector is sampled independently and with the same distribution. Thus, a random forest is a classifier containing many decision trees. The class which is output, is the mode of all the classes output by individual trees [4]. The number of trees in RF have been kept to be 10 for the present analysis. Advantages of RF include simplicity, high parallelizability, fast learning and robustness to noise and outliers. The tools described above have been realized using the default settings of WEKA, an open source data analysis software developed by the University of Waikato [30].

VI. ANALYSIS AND EMPIRICAL RESULTS

This section describes the analysis performed to predict whether there is a change in the SLOC for a class or not. We use the metric suite of Chidamber and Kemerer described in Section III to perform the analysis. Apart from employing the machine learning based ANFIS to predict the maintenance effort, we also apply the conventional machine learning techniques such as logistic regression, bagging and random forest. We use the following measures to evaluate the performance of each prediction model:

The sensitivity and specificity are important measures of performance of binary classification test. In the context of the present datasets, the percentage of classes correctly predicted to be change-prone is known as sensitivity (or true positive rate) of the model. The percentage of the classes predicted not to be change-prone correctly is called the specificity (or true negative rate) of the model. For a prediction model to be accurate, both the sensitivity and specificity of the model should be as high as possible.

The receiver operating characteristic curve, also known as ROC curve has been used to evaluate the performance of the present classifier model. The ROC curve plots sensitivity on the y-coordinate and 1-specificity on the x-coordinate. While constructing ROC curves, we selected many cutoff points between 0 and 1, and calculated sensitivity and specificity at each cut off point. The optimal choice of the cutoff point (that maximizes both sensitivity and specificity) can be selected from the ROC curve [24,31]. We use the area under the ROC curve (AUC) to estimate of the model accuracy. For a model to be accurate, the value of AUC should be close to one [31].

In order to predict the accuracy of the model, it is necessary to perform the cross-validation. Cross validation can be defined as applying the model to a data set other than the one from which the model was originally derived. In general, the dataset is divided into two parts – training set and testing set – wherein training set is used to derive the model and testing set is used to validate the model. Various methods such as k -fold cross validation and leave-one out validation have been proposed in the literature. For the present change-proneness prediction, we use the k -fold cross validation technique with $k=10$ since leave-one technique would be computationally intensive because of the large number of dataset instances involved in the study. In the 10-fold cross validation technique, the dataset is divided into 10 approximate equal parts, where each part in turn is used for testing and the remaining nine parts are used for training [32].

A. Validation Results for Frinika

Table IV shows the validation results on *Frinika* software using four different approaches – ANFIS, logistic regression, bagging and random forest. It can be seen that ANFIS has the greatest AUC of 0.752 and the corresponding values of sensitivity and specificity are 70.6% and 70.3% respectively at a cutoff point 0.4637. It can also be interpreted that bagging gives comparative results to ANFIS. Though the sensitivity

and specificity for bagging are same as that of ANFIS, the AUC is 0.726 which is slightly less than that of ANFIS. If we compare RF and LR, the AUC value is greater for RF due to which it is expected to give better performance. The sensitivity and specificity values for RF are equal (62.4%) and the cutoff value is 0.4285. The ROC curves for *Frinika* are shown in Fig. 1.

TABLE IV. VALIDATION RESULTS FOR FRINIKA

Method	Sensitivity	Specificity	Cutoff Point	AUC
ANFIS	70.60	70.30	0.4637	0.752
Logistic Regression	61.20	61.40	0.4115	0.671
Bagging	70.60	70.30	0.4430	0.726
Random Forest	62.40	62.40	0.4285	0.704

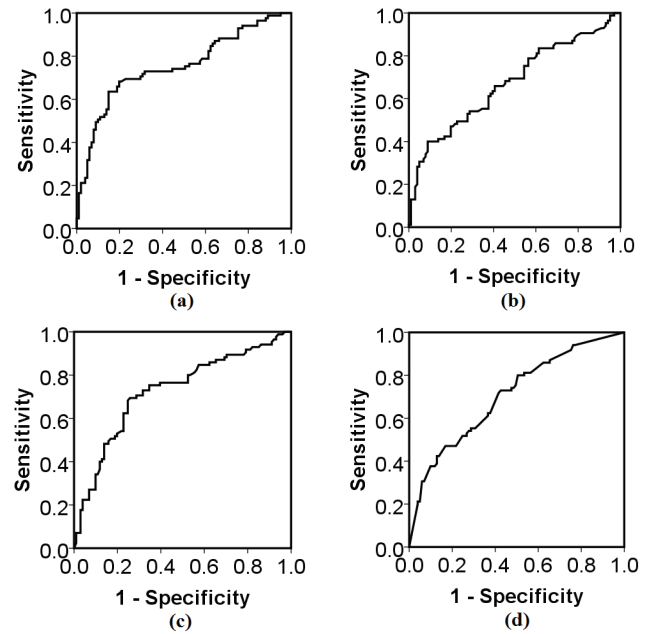


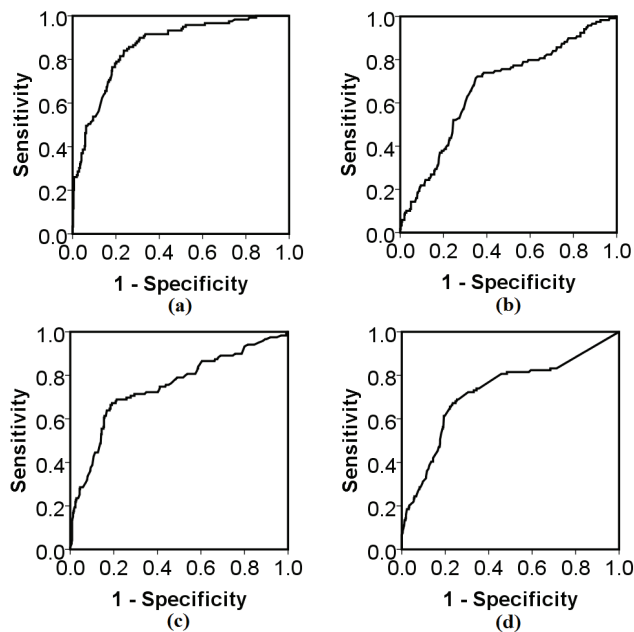
Fig. 1. ROC curves for *Frinika* (a) ANFIS (b) Logistic Regression (c) Bagging (d) Random Forest

B. Validation Results for CheckStyle

Table V shows the validation results for *CheckStyle* software using the four different approaches discussed above. Again, the results of ANFIS are the best with AUC of 0.860 and the corresponding sensitivity and specificity values being 79.0% and 78.9% respectively. Bagging and RF give the similar results in this case with the AUC for bagging (0.748) being slightly higher than the RF (0.725). LR gives the lowest performance of all the four approaches with AUC, sensitivity and specificity being 0.671, 67.2% and 66.7% respectively. Fig. 2 shows the ROC curves for *CheckStyle* using four approaches.

TABLE V. VALIDATION RESULTS FOR CHECKSTYLE

Method	Sensitivity	Specificity	Cutoff Point	AUC
ANFIS	79.00	78.90	0.3054	0.860
Logistic Regression	67.20	66.70	0.2230	0.671
Bagging	70.60	70.50	0.1615	0.748
Random Forest	72.30	69.50	0.1920	0.725

Fig. 2. ROC curves for *CheckStyle* (a) ANFIS (b) Logistic Regression (c) Bagging (d) Random Forest

VII. CONCLUSION

We have investigated the relationship between the object-oriented metrics and change proneness for a class. We have analyzed and compared the performance of ANFIS, bagging, logistic regression and random forest for two commercial open source software systems – *Frinika* and *CheckStyle*. The ROC analysis shows that of all the four techniques, ANFIS is the best with highest AUC for both the software systems. The AUC for *Frinika* using ANFIS is 0.752 whereas it is 0.860 for *CheckStyle* by using the same algorithm. There is a probability that the change prone class will require more effort during maintenance phase of the software development life cycle. Knowing the change prone classes at the early phases of software development will help the software practitioners in planning and managing the testing resources effectively.

Though this study is limited to two datasets only, we can carry out the same study using different datasets to obtain more generalized results. As a part of further research, we plan to carry out the cost-benefit analysis of the present model in order to determine its economic viability.

REFERENCES

- [1] Y. Singh and R. Malhotra, Object-oriented software engineering, 1st ed., Prentice Hall India, New Delhi, April 2012.
- [2] IEEE Standard for Software Maintenance, IEEE Std 1219-1993, pp. 1-45, June 1993.
- [3] W. Li, S. Henry, D. Kafura, and R. Schulman, "Measuring object-oriented design," *Journal of Object-Oriented Programming*, vol. 8, pp. 48-55, 1995.
- [4] S. Chidamber and C. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on Software Engineering*, vol. 20, pp. 476-493, June 1994.
- [5] M. Lorenz and J. Kidd, Object-oriented software metrics, 1st ed., Prentice Hall, New Jersey, July 1994.
- [6] B. Henderson-Sellers, Object-oriented metrics: Measures of complexity, 1st ed., Prentice Hall, New Jersey, December 1995.
- [7] K.K. Aggarwal, Y. Singh, A. Kaur, R. Malhotra, "Empirical study of object-oriented metrics," *Journal of Object Technology*, vol. 5, pp. 149-173, November-December 2006.
- [8] anfis and the ANFIS Editor GUI – MATLAB and Simulink – MathWorks India, <http://www.mathworks.in/>.
- [9] A.R. Sharafat and L. Tahvildari, "A probabilistic approach to predict changes in object-oriented software systems," 11th European Conference on Software Maintenance and Reengineering, pp. 27-38, March 2007.
- [10] A.R. Han, S.U. Jeon, D.H. Bae, and J.E. Hong, "Behavioral dependency measurement for change-proneness prediction in UML 2.0 design models," 32nd Annual IEEE International Computer Software and Applications Conference, pp. 76-83, July-August 2008.
- [11] M.D'Ambros, M. Lanza, and R. Robbes, "On the relationship between change coupling and software defects," 16th Working Conference on Reverse Engineering, pp. 135-144, October 2009.
- [12] M.A. Chaumun, H. Kabaili, R.K. Keller, and F. Lustman, "A change impact model for changeability assessment in object-oriented software systems," Third European Conference on Software Maintenance and Reengineering, pp. 130-138, March 1999.
- [13] Y. Zhou, H. Leung, and B. Xu, "Examining the Potentially Confounding Effect of Class Size on the Associations between Object-Oriented Metrics and Change-Proneness," *IEEE Transactions on Software Engineering*, vol. 35, pp. 607-623, September-October 2009.
- [14] H. Lu, Y. Zhou, B. Xu, H. Leung, and L. Chen, "The ability of object-oriented metrics to predict change-proneness: a meta analysis," *Empirical Software Engineering*, vol. 17, pp. 200-242, June 2012.
- [15] X. Yuan, T.M. Khoshgoftaar, E.B. Allen, and K. Ganesan, "An application of fuzzy clustering to software quality prediction," 3rd IEEE Symposium on Application Specific Systems and Software Engineering Technology, pp. 85-90, March 2000.
- [16] T. Gyimothy, R. Ferenc, and I. Siket, "Empirical validation of object-oriented metrics on open source software for fault prediction," *IEEE Transactions on Software Engineering*, vol. 31, pp. 897-910, October 2005.
- [17] K.K. Aggarwal, Y. Singh, P. Chandra, and M. Puri, "Measurement of software maintainability using a fuzzy model," *Journal of Computer Sciences*, vol. 1, pp. 538-542, 2005.
- [18] R. Malhotra and M. Khanna, "Investigation of relationship between object-oriented metrics and change proneness," *International Journal of Machine Learning and Cybernetics*, vol. 137, April 2012.
- [19] A. Kaur, K. Kaur, and R. Malhotra, "Soft computing approaches for prediction of software maintenance effort," *International Journal of Computer Applications*, vol. 1, pp. 69-75, 2010.
- [20] Understand your Code, <http://www.scitools.com/>.
- [21] J. Shing and R. Jang, "ANFIS: Adaptive network based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, pp. 665-685, May-June 1993.
- [22] V.R. Basili, L.C. Briand, and W.L. Melo, "Validation of object oriented design metrics as quality indicators," *IEEE Transactions on Software Engineering*, vol. 22, pp. 751-761, October 1996.

- [23] K.K. Aggarwal, Y. Singh, A. Kaur, and R. Malhotra, "Empirical analysis for investigating the effect of object-oriented metrics on fault proneness: a replicated case study," *Software Process: Improvement and Practice*, vol. 16, pp. 39-62, August 2008.
- [24] Y. Singh, A. Kaur, and R. Malhotra, "Empirical validation of object-oriented metrics for predicting fault proneness," *Software Quality Journal*, vol. 18, pp. 3-35, March 2010.
- [25] D.W. Hosmer and S. Lemeshow, *Applied logistic regression*, 2nd ed., John Wiley and Sons, New Jersey, 2000.
- [26] A.B. Musa, "Comparative study on classification performance between support vector machine and logistic regression," *International Journal of Machine Learning and Cybernetics*, vol. 4, pp. 13-24, February 2013.
- [27] G. Sidorov, M. Koeppen, and N.C. Cortés, "Recent advances in machine learning techniques and applications," *International Journal of Machine Learning and Cybernetics*, vol. 2, pp. 123-124, July 2011.
- [28] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123-140, August 1996.
- [29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, October 2011.
- [30] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, pp. 10-18, November 2009.
- [31] K. El Emam, S. Benlarbi, N. Goel, and S.N. Rai, "A validation of object-oriented metrics," *NRC Publications Archive*, October 1999.
- [32] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," 14th International Joint Conference on Artificial Intelligence, vol. 2, pp. 1137-1143, 1995.

APPLICATION OF SUPPLY CHAIN MANAGEMENT IN A FOOD PROCESSING COMPANY

Parvesh Antil¹, Anish Kumar²,

¹Assistant Professor, MAE Department, NIEC, New Delhi, (India)

²ME Department, Delhi Technological University, (India)

ABSTRACT

Supply chain is an important activity in most of the industries and is the central focus of many industries. Effectiveness and efficiency of a supply chain effect companies to a great extent. With the growing interest in it, a lot of work has been done in this field in the past two decades. . In this work tried to explain various drivers and important factors in supply chain management with the help of a case study. We have tried to explore various aspects of supply chain with the help of a case study of a food processing company. With the help of a case study we have tried to probe into various issues and factors of a supply chain and supply chain management. The case study has been developed with a more physical view into supply chain..

Keywords: Supply Chain, Supply Chain Management.

I INTRODUCTION

From the raw material stage to final delivering of product, all the entities involved in supplying a product to a customer is called supply chain. Manufacturing unit, marketing, transportation department, supplier, warehouse, retailer and even the customer are all parts of supply chain.

Supply chain management is the coordination of production, inventory, location, and transportation among the participants in a supply chain to achieve the best mix of responsiveness and efficiency for the market being served. It is a holistic approach and strategy to add value to the supply chain organisation. The SCM approach involves people, technology, activities, information directly or indirectly related to the product/services, involved in supplying the product/services from supplier to customer (sunil et al.).

Eg. here we have considered the case of a milk and milk products processing organization. The customer walks into the retail outlet, the supply chain begins with the customer and his/her need for the product. The retailer who has stocked inventory of different products in quantities in accordance with the daily demand and the life of each product. The retailer is supplied by the regional distributor or directly from the company. The distributor too is stocked by the company directly or the distributing unit of the company, which gets its stock from manufacturing unit. The manufacturing unit too need raw material like unprocessed milk, nuts and other additives in the milk, other materials which are required in daily plant running. The company gets its supplies from its suppliers and acts as customer to them thereby forming another supply chain which is procurement chain to it.

Thus supply takes place at different levels efficient management directly effects the customer satisfaction, sale, cost of product. This makes it a very important function to deal with.

Aim of Supply chain management is to recognize the strategic nature of coordination between trading partners, create sourcing, making and delivery process of products/services and logistics functions seamlessly across the supply chain to gain a competitive edge in the market. The objective of supply chain management is to be efficient and cost-effective. Thus, the emphasis is on taking a systems approach to supply chain management.

II LITERATURE BACKGROUND

SCM practices have been globally accepted by enterprises as an important factor to survive in the present global scenario. Numerous researchers have explored SCM on various issues viz. Definition, dimensions, performance measurement, strategy evaluation etc. Therefore a vast literature giving an insight into this evolving concept is there.

“A supply chain is the alignment of firms that bring products or services to market.”—from Lambert, Stock, and Ellram in their book *Fundamentals of Logistics Management* (Lambert, Douglas M., James R. Stock, and Lisa M. Ellram, 1998, *Fundamentals of Logistics Management*, Boston, MA: Irwin/McGraw-Hill, Chapter 14)².

Various definitions of supply chain are:-

“A supply chain consists of all stages involved, directly or indirectly, in fulfilling a customer request” Chopra and Meindl³.

“The systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole.” Mentzer et al⁴.

“A supply chain encompasses all activities associated with the flow and transformation of goods from the raw material stage, through to the end user, as well as the associated information flows” Handfield & Nichols⁵.

“Supply chain can be defined as integrated process wherein a number of various business entities (i.e., suppliers, manufacturers, distributors, and retailers) work together in an effort to: (1) acquire raw materials, (2) convert these raw materials into specified finished final products, and (3) deliver these final products to retailers”-Benita M. Beamon, “Supply Chain Design and Analysis: Models and Methods⁶”.

Acc. To Council of Supply Chain Management Professional – “Supply Chain Management encompasses the planning & management of all activities involved in sourcing & procurement, conversion, and all logistics management activities.”

SCM is a concept that has evolved since 50's and 60's, in 1950's, 1960's most manufacturers emphasised on mass production to minimize cost. Excessive inventory was maintained to take care of bottleneck operations. Information and technology sharing was seen as risk and not at all good for the organization. In place of Supply chain the term logistics was used which emphasized only on transportation function. In 1970's material requirement planning (MRP) was developed and managers realized new material management concepts to improve performance. In the 1980's due to global competition manufacturers were forced to offer low-cost, high performance and flexible products. JIT and more management programs were used to improve the efficiencies

of production and improve performance. The term Supply Chain Management came into use in 1990's when managers further extended improvement strategies to include strategic suppliers and logistics functions. The concept of supply chain has evolved to a great extent. Manufacturers and retailers are embracing the concept of SCM to improve efficiencies and effectiveness to a great extent (Tan, 2002)⁷.

Various entities perform different functions in a supply chain. Here we have discussed the following seven :- demand forecast, facility location, inventory, transport, information. Demand forecast is an important function which is heavily dependent on information sharing and connectivity in a supply chain and is a strategically important decision.. According to Vives⁸(1994) demand is directly influenced by market competition, nature of goods and degree of product differentiation. Yue et. al.⁹ assess the benefits of sharing demand forecast information in a manufacturer-retailer supply chain, consisting of a retail channel and a direct channel. Both the manufacturer and the retailer set their price based on their forecast of the primary demand. Information also has great effect on demand forecast. Information sharing can be very valuable to both the manufacturer and the retailer, especially in situations where the accuracy of the retailer forecast is low, the accuracy of the manufacturer forecast is high, and the correlation between forecasts is low. It also assesses the effects of demand information sharing in a manufacturer-retailer supply chain. (Ren et. al.)¹⁰.

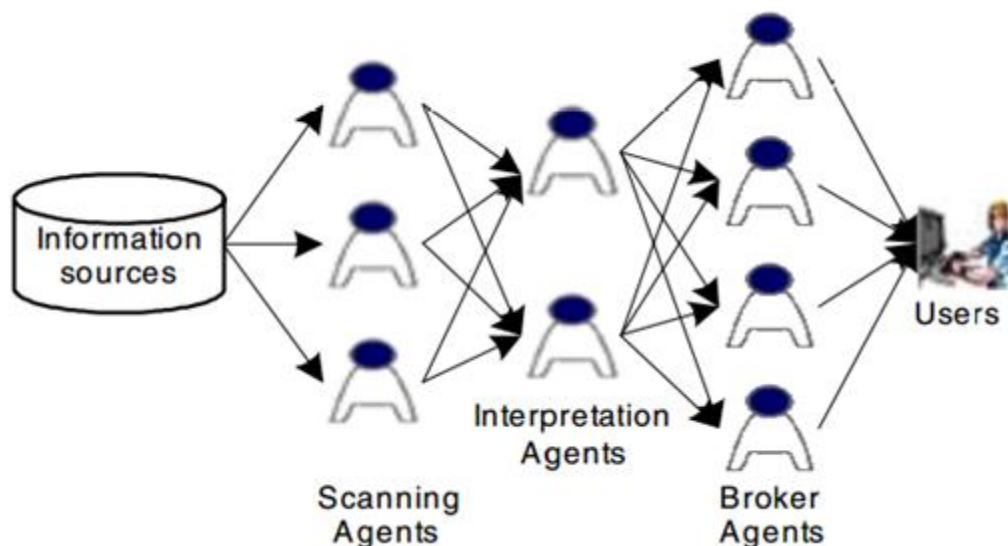
Facility location has great strategic and financial importance in a supply chain. It has been extensively studied. A crucial aspect of many practical location problems regards the existence of different types of facilities, each one of which playing a specific role (e.g., production, warehousing), and a natural material flow (that is, a hierarchy) between them (Melo et al)¹¹. Simchi-Levi et al.¹² state that "the strategic level deals with decisions that have a long-lasting effect on the firm.". Although typically no location decisions are made on the tactical or even operational level, a number of issues are strongly related to them such as inventory control policies, the choice of transportation modes and capacities, warehouse layout and management, and vehicle routing (among others) (Melo et. al.)¹³.

Inventory accounts for significant capital investment in an organization. Blackburn(1991)¹⁴ Shorter and shorter product life cycles as well as growing innovation rates make demand extremely variable and the collection of statistics (which are required by stochastic models) less and less reliable, this makes efficient inventory management very important. Giannoccaro et. al.¹⁵ Inventory decisions in SCs are made independently at each stage (as it is often the case), they are usually based on the local inventory status and local performance objectives (local policies). The paper proposes to define an inventory management policy that is both easy to be implemented and near optimal for the whole SC. It presents a methodology to define a supply chain (SC) inventory management policy, which is based on the concept of echelon stock and fuzzy set theory. It presses for an periodic review policy then a continuous review policy, and asks for clear definition of inventory policy. Firms with better capital intensity tend to have lower numbers of days holding their inventories. Sahari et. al.¹⁶ has investigated into the correlation between the inventory day, return on assets and capital intensity as a measure of supply chain performance. It finds a significant negative relationship between inventory days and return on assets. Sahin et. al.¹⁷ explores the impact of radio frequency identity – RFID technology on supply chain operations and finds significant cost improvement and reduced inventory data record inaccuracies with its use.

Logistics acts as blood through a supply chain, it is the physical connection between different levels in a supply chain. Liu¹⁸ has presented a new management system of material supplies based on the modern Logistics supply

chain theory for the enterprise material supply management. Logistics outsourcing is one of the growing trends in supply chain management. The good choice of service suppliers is the key to success in logistics outsourcing. Pankaj et al.²⁰ has studies the challenges for logistics sector in Indian market. Evolving technologies, infrastructure, changing government policies play a significant role in the changing logistics sector in India. Logistics in India²¹ report by KPMG too provides a great insight in to the present scenario in Indian logistics industry.

Information sharing plays a vital role to match demand and manufacture, for demand forecast in supply chain. However, the benefits of sharing information among supply chain members are not always the same. They depend on the supply chain structure (e.g., serial or distributive systems) and its operational characteristics (e.g., demand patterns and costs involved) choi²². Yen²³ proposes Information sharing chain to balance demand and supply in a supply chain. Similar to a supply chain, an information supply chain (ISC) fulfills users' information requirements by a network of information-sharing agents (ISA) that gather, interpret, and satisfy the requirements with proper information.



An Information Supply Chain

Fig 1- Shuang Sun and John Yen²³, "Information Supply Chain: A Unified Framework for Information-Sharing"

III The Case

ABC Fruit and Vegetable Pvt. Ltd. , is a 30 year old organization . It generates annual revenue of 2000 Cr. and provides employment to 1500 employees. It is an IS/ISO-9002, Is-15000 HACCP and IS-14001 EMS certified organization. Setting high quality standard for itself, ABC has its own Quality Assurance Laboratory, certified by National Accreditation Board for Testing and Calibration Laboratory (NABL)-Department of Science and Technology, Government of India.

ABC has a wide range of products in markets like Liquid milk , Ice creams, Flavoured Milk, Dahi , Lassi, Ghee, Butter, Cheese, Range of edible oils, range of Fresh Fruits and vegetables. With this wide range of products, it has a national level sales and distribution network. ABC markets 3 million ltrs of milk daily in the markets of Delhi, Mumbai, Saurashtra and Hyderabad. It has a 66% market share in the branded milk sector in

Delhi, where it sells 2.3 million litres of milk daily. It sells 1 million litres Bulk Vended Milk and 1.3 million Polypack milk daily in Delhi-NCR. ABC conducts its operations through 14000 retail sellers and 1000 exclusive outlets of ABC. This all involves a lot of, marketing, quality management, supplier-customer relationship management, logistics services. Bringing milk from all the collection points to the plant and that too the volume of almost 2 million litres every day requires great logistics and supplier relationship management.

ABC sale of has a daily of yogurt of 10000 kg which are sold in packed 200gm poly packs . Therefore 50000 poly packs every day. Other cultured products like lassi, mishit dohi, flavoured yogurt too have significant sales ABC also has also manufactures and markets a wide range of ice-cream. ABC is the third largest ice-cream manufacturer of ice cream in the country and has a market share of 20% presently. ABC has a daily sale of 5000 ltr per day which peaks up to 12000 ltrs per day in peak summer time. . The case here considered is based on all three chains of ABC.

IV SUPPLY CHAIN

4.1 Demand Forecast

In ABC, cultured chain process cycle is of 3 days, which starts with demand forecasting for two days forward. Demand is forecasted on the basis of last few days' sale, trend in the last week sales, weather conditions, seasoned demand. These factors contribute to the demand in Weather conditions directly affect the demand for curd, in hotter conditions demand increases as the temperature decreases demand too decreases. Last year sales are taken into account to forecast the sales for festive seasons like diwali, holi, marriage season etc. Inventory level is required to be kept at minimum because a daily sale forecast is done the inventory left for every day is adjusted in the demand prediction for the next day. Because the shelf life of curd is 8 days at 4°C temperature therefore it is required that each day production be sold on that every day and if not then on the next day this makes forecasting a very significant factor in cultured chain.

Demand forecast = Last day Demand + Trending Factor + 10% safety buffer

A 10% safety buffer is added to the equation to make sure that stock out may not happen. As left over stock may be sold on the next day. On the basis of this demand forecast an invoice is generated to the production unit which is at a different facility.

For ice-cream shelf life is higher and it is only supplied on the basis of demand.

Milk has fixed routes and fixed daily supply, Supply can be daily changed in accordance with the demand. The retailer daily gives his demand for the next day, thus no forecast is required. But a good information system is required that makes sure that demand is correctly invoiced and supplied.

4.2 Facilities

Facilities are the actual physical locations in the supply chain network where product is stored, assembled or manufactured. ABC has two type of facilities- 1) manufacturing facility 2) distribution facility.

Manufacturing units are strategically placed such that it is close to the receiving point of unprocessed milk. Milk is procured from 8 different states in the country. Transportation of milk from different states takes place by insulated trucks and a daily train as well. Also capital cost to buy facilities is also a major factor. Distribution facilities are placed such that distribution cost is minimum whereas providing maximum access to the maximum customers in the target area. Distribution facility is to be placed at the heart of the distribution area..

Having a distribution unit at a strategically chosen location gives following benefits:-

- I. Responsive supply chain, making it more responsive to complaints, short notice orders
- II. Less transportation cost to due nearness to most of locations.
- III. Centralized distribution facility in proximity to customer.
- IV. Ease of marketing to individual parties.
- V. Better time efficiency of transportation.
- VI. More numbers of on time order are met.

4.3 Inventory

Inventory is a major cost factor in supply chain management. Being a continuous supply Chain inventory is kept at minimum. Daily unprocessed milk is procured and processed and distributed to all the booths and retail points. Any stock left is supplied on the next day. A storage capacity of 11 lakh ltrs of milk is maintained at 2-8°C.

Inventory of curd, lassi and other cultured product is maintained at 2-4°C. Inventory of ice-cream is maintained at -20°C temperature. Major inventory cost is incurred due to maintenance of such low temperature. A minimum stock level is always maintained to avoid any chance of stock out. Inventory is only maintained at the retail outlets to avoid supply on a daily basis. Outlets are supplied once in a week or twice in a week basis depending on their demand. The retail point owner gives his demand to the transporter or directly to company via its demand placing helpline. The inventory cost is directly incurred by the booth owner although this factor is directly included in the margin given to the retailer.

4.4 Transportation

Transportation refers to the movement of product from one location to another as it makes way from the beginning of a supply chain to the customer. Transportation is an important supply chain as raw material is not supplied from and consumed at the same place, also finished product is not produced and consumed at the same time. Transportation Cost is a major driver in ABC, which contribute upto 10% cost of the final product. Success of a supply chain is largely dependent on transportation efficiency.

Transportation efficiency = distance travelled by vehicle in loaded condition/ total distance travelled .

ABC has used its transportation to achieve its strategic goal. ABC uses a very responsive transportation with a goal to decrease its transportation and receiving cost while ensuring that product availability matches customer

demand. Routes for transportation are fixed and are daily generated by SAP software. The routes are subjected to change in case of any short notice demand, which makes the logistics of ABC more responsive.

4.5 Transport and logistics

ABC has opted for the methodology of outsourcing the outbound logistics. The poly-pack distribution is totally outsourced and loose milk distribution is partially outsourced.

ABC earlier had a fleet of 105 vehicles and 200 Driver-cum-Salesman(DCS) for loose milk distribution, which has now been reduced to mere 32 vehicles. This has led to substantial capital saving over the years showing in the balance sheets.

But the results has not been very satisfactory. Outsourcing has led to a great downfall in the service level of logistics. Brand ABC has also taken a setback because of it. There is lesser reliability and lesser confidence in the outsourced logistics.

The outsourcing vehicles are not maintained very well by their owners. General hygiene is being compromised. There is lesser dedication and lack of belongingness. Disciplinary issues are also there. Quality of outsourcing drivers is not good. Responsiveness has gone down which has lead to a loss of business too.

Although outsourcing has led to great savings on balance sheets but the loss of brand value and other factors are significant too. This is why ABC has not gone for 100% outsourcing. 15 new vehicles have been bought recently and more are on orders. A reassessment of outsourcing option is being done.

4.6 Information flow

Information deeply affects every part of a supply chain in many different ways. Information serves as a connection between various stages of a supply chain, allowing them to coordinate and maximize total supply chain profitability.

It is crucial to each stage in a supply chain eg. a manufacturing schedule is maintained, it uses daily demand information to forecast future demand and manufacture the product in accordance so that right amount of product is available at right time an position.

Such demand forecast is excessively used in milk and curd production in ABC , where daily demand is monitored to create the production schedule of 2 days ahead.

A warehouse management system uses information to create visibility of the warehouse's inventory as in the case of inventory if ice-cream in ABC.

Information management in ABC is totally through SAP software. In SAP Inventory level, stock level, Dispatch quantity, received quantity , date and time of dispatch and receiving, cost price and selling price of product, product code all are saved. A product is considered dispatched or received only when it is entered in SAP.

This has improved accuracy and smart information management system.

Demand information also form a major part of information system, Daily demand of different product from retailers and short term demands are received through a demand department. A special call-in demand centre in formed that takes in demand from retailer and enter the product code and demand of a product into the system. This makes the supply chain of ABC highly responsive and flexible to customer demands. Any complaints and queries are also addressed here.

4.7 Supplier Selection

ABC selects its suppliers from all over the country. Local Dairies, cooperative dairy societies, individual local suppliers are all include in it. Some local dairies supply their milk directly to ABC and therefore use the goodwill of ABC to sell their product. these dairies have to comply with the standards set by ABC for their product. Total transportation, packaging and quality is their responsibility.

4.8 Retailers

ABC has developed an extensive franchise network which perform key role of providing the customer service in ABC supply chain. Responsibilities of a ABC retailer are:-

Operate and manage store

Hire and pay staff

Order supplies

Provide Customer service

Maintain ABC outlet appearance.

ABC is responsible the following functions to its outlets:-

Pay for system operations, it is added to the margin of retailer

Provide ordering system

Develop supply and merchandise

Accounting services

Time to time training and educating the retailer as they represent ABC to the customer

Install and remodel facilities

V CONCLUSION

Supply chain is a very dynamic topic; different viewpoints can be used to explain it. We can explain supply chain from the view point of methodologies and strategies, from SCM practices, factors affecting SCM, physical entities of supply chain. Here in this work various entities of a supply chain are explained with the help of a case study of the supply chain of milk processing company. Supply chains of three products each varying with

inventory, product life and practices of supply chain. Methodologies that are used have been explained. Different entities of a supply chain have been explained at their very operations level, with the help of a case study of a food processing company.

REFERENCES

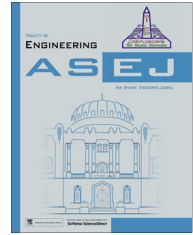
1. Sunil Chopra, Peter Meindl, D. V. Kalra, Supply Chain Management- Strategy, Planning, And Operation (third edition, Pearson Prentice Hall, 2007).
2. Lambert, Douglas M., James R. Stock, and Lisa M. Ellram(1998), Fundamentals of Logistics Management, Boston, MA: Irwin/McGraw-Hill, Chapter 14.
3. Sunil Chopra and Peter Meindl (2001) book, Supply Chain Management: Strategy, Planning, and Operation, Prentice Hall of India.
4. Mentzer J. T. , DeWitt V, Keebler K. S., Min S., Nix N. W. and Smith. C. D, Defining Supply Chain Management, Journal of Business Logistics, 2001.
5. Handfield R. B. and Nichols E. L., Introduction to Supply Chain Management, Prentice-Hall, New Jersey, 1999, pp. 1-183.
6. Benita M. Beamon, Supply Chain Design and Analysis: Models and Methods, International Journal of Production Economics (1998), Vol. 55, No. 3, pp. 281-294.
7. Keah Choon Tan(2002), Supply Chain Management: Practices, Concerns and Performance Issues; Journal of Supply Chain Management, 38, 1; ABI/INFORM Global pg. 42.
8. Xavier vives, Duopoly Information Equilibrium: Cournot and Bertrand, Journal of Economic Theory 34, 74-91(1984).
9. Xiaohang Yue, John Liu, Demand forecast sharing in a dual-channel supply chain, European Journal of Operational Research 174, (2006) 646–667.
10. Z. Justin Ren, Morris A. Cohen, Teck H. Ho, Christian Terwiesch, Sharing Forecast Information in a Long-term Supply Chain Relationship, August 2006.
11. M.T. Melo, S. Nickel, F. Saldanha-da-Gama, Facility location and supply chain management – A review, European Journal of Operational Research 196 (2009) 401–412.
12. Simchi-Levi, Kaminsky, and Simchi-Levi, Designing and Managing the Supply Chain, McGraw-Hill, London, 1999.
13. T. Melo, S. Nickel, F. Saldanha da Gama, Facility Location and Supply Chain Management – A comprehensive review, 2007.
14. J.D. Blackburn, Time-based Competition: The Next Battleground in American Manufacturing, Irwin, Homewood, IL, 1991.
15. Ilaria Giannoccaro, Pierpaolo Pontrandolfo, Inventory management in supply chains:a reinforcement learning approach, Int. J. Production Economics 78 (2002), 153-161.
16. Salawati Sahari, Michael Tinggi and Norlina Kadri, inventory management in malaysian construction firms: impact on performance, SIU Journal of Management, Vol.2, No.1 (June, 2012).
17. Evren Sahin, Yves Dallery, A literature review on the impact of inventory data record inaccuracies on inventory management and the potential of the rfid technology to tackle this issue.

18. Lanqing Liu, research on the management system of enterprises using modern logistics supply chain theor , Procedia Engineering 24, (2011), 721 – 725.
19. Jianliang Peng, Selection of Logistics Outsourcing Service Suppliers Based on AHP , Energy Procedia 17, (2012), 595 – 601.
20. Pankaj Chandra, Nimit Jain, The Logistics Sector in India: Overview and Challenges, March 2007, W.P. No.2007-03-07, IIM Ahemdabad.
21. Logistics in India, September 2010, KPMG.
22. Hyun-cheol, Paul Choi, Information Sharing in Supply Chain Management: A Literature Review on Analytical Research, Volume 8, Number 1, pp 110-116, California Journal of Operations Management © 2010 CSU-POM.
23. Shuang Sun , John Yen, Information Supply Chain: A Unified Framework for Information-Sharing.



Ain Shams University
Ain Shams Engineering Journal

www.elsevier.com/locate/asej
www.sciencedirect.com



MECHANICAL ENGINEERING

Development of reliability index for combined cycle power plant using graph theoretic approach

Nikhil Dev ^{a,*}, Samsher ^b, S.S. Kachhwaha ^c, Rajesh Attri ^a

^a Department of Mechanical Engineering, YMCA University of Science and Technology, Faridabad, Haryana, India

^b Department of Mechanical Engineering, Delhi Technological University, Delhi, India

^c Department of Mechanical Engineering, School of Technology, Pandit Deendayal Petroleum University, Gandhinagar, Gujrat, India

Received 22 October 2012; revised 5 August 2013; accepted 15 September 2013

KEYWORDS

CCPP;
Systems;
Graph theory;
Digraph;
Permanent function;
Matrix

Abstract A systematic approach based on graph theory and matrix method is developed ingeniously for the evaluation of reliability index for a Combined Cycle Power Plant (CCPP). In present work CCPP system is divided into six subsystems. Consideration of all these subsystems and their interrelations are rudiment in evaluating the index. Reliability of CCPP is modeled in terms of a Reliability Attributes Digraph. Nodes in digraph represent system reliability and reliability of inter-relations is represented by edges. The digraph is converted into one-to-one matrix called as Variable System Reliability Permanent Matrix (VPM-r). A procedure is defined to develop variable permanent function for reliability (VPF-r) from VPM-r. Reliability index of CCPP system is obtained from the permanent of the matrix by substituting numerical values of the attributes and their inter-relations. A higher value of index implies better reliability of the system. The proposed methodology is illustrated step-by-step with the help of two examples.

© 2013 Production and hosting by Elsevier B.V. on behalf of Ain Shams University.

1. Introduction

Reliability analysis is an innate aspect of combined cycle power plant design and plays considerable role throughout the plant operation in terms of expenses (operating and maintenance) and optimal maintenance scheduling of its

equipments. Reliability may be defined as the ability of an equipment, component, product, system, etc., to function under designated operating state of affairs for a specified period of time or number of cycles [1]. For a large and complex electricity generating system such as CCPP, reliability is the probability of generating electricity under operational conditions for a definite period of time. Reliability of a CCPP is function of maintenance (scheduled or forced) cost, which in turns depends upon the Mean Time Between Failures (MTBF) and Mean Time To Repair (MTTR) of equipments or systems, and which are further dependent on complexity in design, state, age of the equipment or system and to some extent on the availability of spare parts.

Recurring failures that lead to complete power plant outage need repair and proactive maintenance to invigorate power

* Corresponding author. Tel.: +91 9711812394.

E-mail addresses: nikhildevgarg@yahoo.com (N. Dev), sam6764@yahoo.com (Samsher), sskachhwaha@rediffmail.com (S.S. Kachhwaha), rajeshattri2005@gmail.com (R. Attri).

Peer review under responsibility of Ain Shams University.



Production and hosting by Elsevier

plant performance and reduction monetary losses. Downtime losses and maintenance cost of a CCPP can be reduced by adopting a proper mix of maintenance and repair strategies. In the worst situation, unavailability of an equipment or system affects whole plant and plant trips in this case. But in general, the failure of an equipment or system may not affect the complete plant and therefore its criticality is at some intermediate value. In that case reliability of system comes down and its effect on reliability of other systems is also observed. The criticality level decides the importance of the equipment or system and choice of appropriate maintenance and repair strategy so that reliability may be maintained up to a mark.

In the literature both qualitative and quantitative methods for assessing the reliability of complex systems are available. The most commonly used qualitative methods are Fault Tree Analysis (FTA), Failure Modes, Effects and Criticality Analysis (FMECA), Failure Modes and Effects Analysis (FMEA), Root Cause Analysis (RCA), Root Cause Failure Analysis (RCFA), Fish Bone Analysis (FBA), Event Tree Analysis (ETA), and Predictive Failure Analysis (PFA). Block diagram analysis, Markov chain, and Monte Carlo simulation are some of the quantitative methods of reliability analysis available in the literature.

Various attempts have been made by researchers in developing procedures for the evaluation of the reliability of various systems [2–10]. The two-state Markov model is the mainly used outage model in power system reliability analysis [11].

Eti et al. [12] integrated reliability and risk analysis for maintenance policies of a thermal power plant. Need to integrate RAMS (reliability, availability, maintainability and supportability) centered maintenance along with risk analysis was stressed, although results expected or obtained with the application of those concepts were not explained.

A staircase function was introduced by Ji et al. [11] to approximate the aging failure rate in power systems and a component renewal process outage model based on a time-varying failure rate was proposed. The model reflected the effects of component aging and repair activities on the aging failure rate.

Markov method was used by Haghifam and Manbachi [13] to model reliability, availability and mean-time-to-failure indices of combined heat and power (CHP) systems based on interactions between electricity generation, fuel-distribution and heat-generation subsystems. The proposed model can be useful in feasibility studies of CHP systems and in determining their optimal design, placement and operational parameters.

Carpaneto et al. [14] carried out Monte Carlo simulation for identifying long, medium and short-term time frames by incorporating uncertainty at large-scale and small-scale for cogeneration system. Availability coefficient assumed to be independent of year, scenario and control strategy was defined for unavailability of the CHP units, due to scheduled maintenance and reliability aspects, taking into account. Large-scale uncertainty referred to the evolution of energy prices and loads and relevant to the long-term time frame was addressed within multi-year scenario analysis. Small-scale uncertainty relevant to both short-term and medium-term time frames was addressed through probabilistic models and Monte Carlo simulations [15].

Mohan et al. [2] calculated RTRI (real-time reliability index) for a SPP (steam power plant) using graph theory. Integration of systems and subsystems and interaction among

them were considered for the reliability analysis and the proposed methodology can be applied for obtaining availability and maintainability; including optimum selection, bench marking, and sensitivity analysis of SPP. Tang [16] proposed a new method based on the combination of graph theory and Boolean function for assessing reliability of mechanical systems. Graph theory was used for modeling system level reliability and Boolean analysis for interactions. The combination of graph theory and Boolean function bring into being an effective way to evaluate the reliability of a large, complex mechanical system. Garg et al. [5] developed a graph theoretical model to compare various technical and economical features of wind, hydro and thermal power plants.

Performance analysis of coal based steam power plant boiler was carried out by Mohan et al. [3] using graph theory and step-by-step methodology for the evaluation was also proposed. Further graph theory was applied to calculate real-time efficiency index (RTEI) defined as the ratio of the values of variable permanent system structure function (VPF) in real-time (RT) situation to its achievable design value [4] and in this connection graph theory was used to recommend the an appropriate maintenance strategy for power plants [6].

The reliability and availability of a CCPP depend on the perfect operation of all its systems (e.g., gas turbine, heat recovery steam generator, steam turbine and cooling system) [17]. So far researchers evaluated combined cycle power plant system reliability only at system level without making an allowance for the interactions of systems, and subsystems. Therefore, there is a need for extending the compass of reliability analysis for combined cycle power plants by taking care of interaction among different systems and subsystems.

A number of approaches and methodologies developed by researchers are available in the literature to model the various systems and their elements. Graph theory is one of such methodologies. It synthesizes the inter-relationship among different parameters and systems to evaluate score for the entire system. Because of its inherent simplicity, graph theory and matrix method have wide range of applications in engineering, science and in numerous other areas [22]. Several examples of its use have appeared in the literature [2–6,21–24] to model the various systems.

This paper presents a mathematical model using graph theoretic systems approach that enables the prediction of CCPP reliability in terms of an index by taking into account various systems and interactions between them.

2. System structure graph of a combined cycle power plant

System structure development is imperative for understanding and analysis of its performance [25] and a combined cycle power plant is no exception. System structure is of two types: abstract and physical. Abstract structure involves performance contributing events and their interrelations or interdependencies. The physical structure of a system implies subsystems, assemblies, components and their interconnections. A CCPP is a combination of a Compressed Natural Gas (CNG) fired gas turbine with Heat Recovery Steam Generator (HRSG) and a steam powered turbine. These plants are very large, typically rated in the hundreds of mega-watts. Combined cycle power plant considered for the present analysis is shown in Fig. 1.

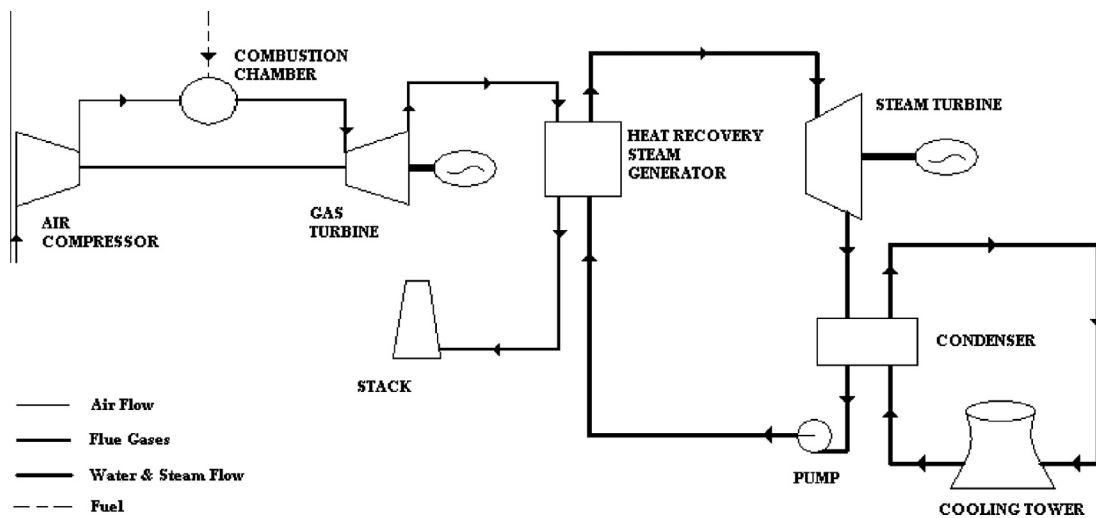


Figure 1 Schematic flow diagram of combined cycle power plant.

Ambient air at Normal Temperature and Pressure (NTP) is compressed by the air compressor to decrease its volume. Air at elevated temperature and pressure is directed to the combustion chamber. The compressed air is mixed with CNG from the fuel supply system to produce hot combustion gas in combustion chamber at constant pressure. Hot combustion gas enters the gas turbine where power is generated. HRSG is the link between the gas turbine and the steam turbine process, whose function is to transfer heat energy from exhaust gases to high pressure water and produces high pressure steam. The steam is separated in the boiler drum and supplied to the super heater section and boiler condenser section. The super heated steam produced in the super heater then enters into steam turbine through the turbine stop valve. After expansion in the turbine the exhaust steam is condensed in the condenser. In the cooling water system, heat recovered from the steam turbine exhaust is carried by the circulating water to the cooling tower, which rejects the heat to the atmosphere.

For the Graph Theoretic Analysis (GTA) large and complex system such as combined cycle power plant must be divided into small subsystems for the convenience of analysis. GTA takes care of inheritance and interdependencies of the subsystems. Further it gives a quantitative measure of system reliability which is helpful in comparing the present reliability with the design value. Six subsystems identified for a CCPP are as follows:

1. Air compressor system (S_1).
2. Combustion chamber system (S_2).
3. Gas turbine system (S_3).
4. Heat recovery steam generator system (S_4).
5. Steam turbine system (S_5).
6. Water system (S_6).

Division of combined cycle power plant in these subsystems is based on the working of different components and subsystems can be divided further into sub-subsystems. Combined cycle power plant is a combination of gas turbine cycle and steam turbine cycle. Gas turbine cycle comprises of air compressor system, combustion chamber system and gas turbine system. Output of gas turbine cycle that is flue gases at high temperature is the driving energy for steam turbine cycle.

Therefore, steam turbine cycle is dependent upon the gas turbine cycle and unavailability of gas turbine cycle trip the whole plant. In case steam turbine cycle is unavailable then flue gases may be sent directly to atmosphere through bypass stack. In this way gas turbine cycle is not dependent on the steam turbine cycle. Failure of a particular component does not mean that whole plant is not working but its reliability is decreased with respect to design value. In case of failure of air compressor or combustion chamber or gas turbine, whole plant is tripped. If any of the HRSG, steam turbine and water system fail then steam cycle will not be working. Therefore, taking in view of this combined cycle power plant is divided into the six subsystems as explained above. Components or systems affecting their reliability are considered their part. Reliability of these subsystems and their interaction will decide the reliability of CCPP as they are connected with each other physically or indirectly. As these subsystems are also very big, therefore, hereafter they are also referred as systems. Let each of the six systems of plant be represented by vertices S_i 's ($i = 1, 2, 3, 4, 5, 6$) and interconnection between two systems (S_i, S_j) is represented by edges c_{ij} 's ($i, j = 1, 2, 3, 4, 5, 6$ and $i \neq j$) connecting the two vertices S_i and S_j . All six systems are connected by flow of air, flue gases, water, steam, heat and work. This flow is shown in Fig. 2 with the help of vertex and edges. This representation is called as System Structure Graph (SSG). This is based upon the functioning of combined cycle power plant as per the following:

1. Air compressor and gas turbine are attached to each other with rigid shaft for continuous power supply to compressor for pressurizing the ambient air. It is represented by the edge c_{31} .
2. Compressed air surging from compressor to the combustion chamber is represented by edge c_{12} . Fuel is injected in the combustion chamber and chemical reaction of fuel with air is at constant pressure. Fuel injection system is considered to be a part of combustion chamber system.
3. Gas turbine blades are cooled by being made hollow so that coolant air, obtained directly from the compressor, can circulate through it. Edge c_{13} represents the air bypassing the combustion chamber.

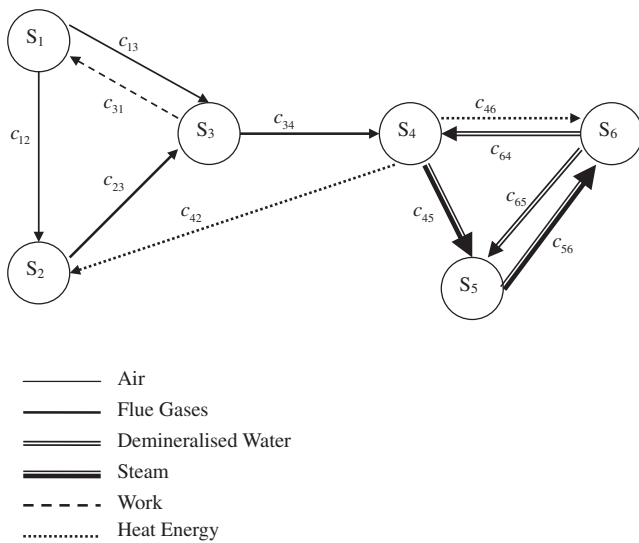


Figure 2 System structural graph of combined cycle power plant.

4. Flue gases surging from combustion chamber to gas turbine is shown by edge c_{23} .
5. HRSG is used for (i) partial heating of compressed air leaving the compressor (c_{42}), (ii) feed water heating (c_{46}), (iii) steam generation at dual or multi-pressure.
6. Flue gases coming out of gas turbine and entering to HRSG system is shown by the edge c_{34} .
7. Superheated steam generated in HRSG and flowing to steam turbine is shown by edge c_{45} .
8. Edge c_{56} represents the flow of steam from steam turbine to the condenser.
9. De-mineralized (DM) water injection to superheated and reheated steam (as an attemperation spray) is represented by the edge c_{65} .
10. DM water supplied to HRSG (as feed water) is represented by edge c_{64} .

The system structure graph shown in Fig. 2 represents the internal structure of the CCPP at system level.

3. GTA for reliability analysis of combined cycle power plant

Reliability has two connotations; probabilistic and deterministic. Probabilistic approach is based upon statistical failure modeling, without research and itemizing causes of failure. Deterministic approach concentrates on understanding how and why a component or system fails, and how it can be designed, repaired and tested to prevent such failure from occurrence or recurrence. In the present analysis, probabilistic approach in conjunction with GTA is applied for combined cycle power plant reliability analysis. GTA consists of following three steps [21]:

- Diagram representation.
- Matrix representation.
- Development of permanent function.

The digraph characterizes the visual representation of the systems and their interdependence. The matrix converts the

digraph into mathematical form and the permanent function is a mathematical model that helps determine the reliability index. It may be noted here that development of permanent function is not merely the determinant of the matrix. It is developed in such a manner that no information regarding the system reliability is lost. For this purpose, a step-by-step methodology is proposed hereafter with the help of two examples.

3.1. Digraph representation of CCPP reliability system

In Fig. 2 it is explained how the air, flue gases, water, steam, heat and work flows from one system to another system. For the GTA it is not necessary to represent these physical interactions by different types of lines. It is required to know whether a system is connected to other systems. If yes then interaction is represented by a line and arrow haggard at the end shows the direction of flow of physical property. A pictogram of systems and their interdependencies in terms of nodes and edges is called digraph. Let nodes (S_i 's) represent systems and edges (c_{ij} 's) symbolize their interactions. S_i indicates the inheritance of systems and c_{ij} indicates degree of dependence of j th system on the i th system. The digraph signifies the proposed CCPP systems and interrelations and represents the system in a simplified manner. Diagram for the SSG (Fig. 2) can be represented as shown in Fig. 3.

Graph theoretic models have adaptability to model any of the RAM (Reliability, Availability and Maintenance) characteristics by associating suitable attributes and interdependencies to the nodes and edges of the SSG [9]. For example, if the node R_i represents the reliability of i th system and r_{ij} represents the reliability of the interconnection between i th and j th systems (nodes) of CCPP; then, systems reliability graph or digraph (SRD) can be obtained from the SSG of a CCPP (Fig. 2 or Fig. 3).

The digraph model (SRD) provides the system structure reliability unequivocally. Reliability of the connection between two systems is considered if the systems are connected either by rigid or imaginary links such as connection between turbine

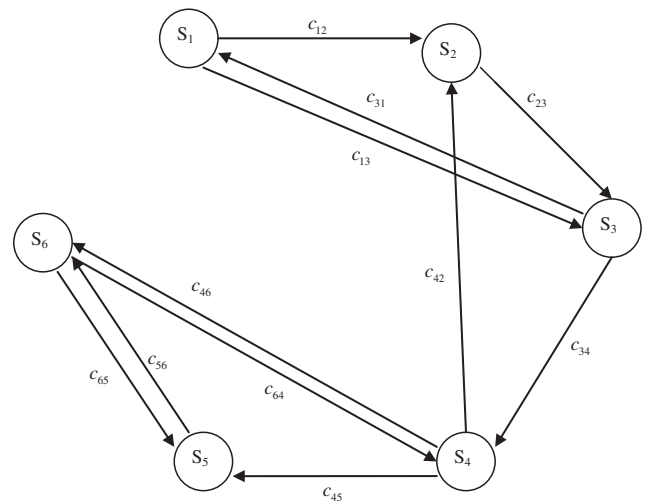


Figure 3 System structure digraph for combined cycle power plant.

and generator rotors through a mechanical shaft or between combustion chamber system and water system of boiler, through flue gases.

3.2. Matrix representation of CCPP system

Although a digraph is very convenient for a visual study, other representations are better for computer processing. A matrix is a convenient and useful way of representing a digraph to a computer. Matrices lend themselves easily to mechanical manipulations. Many results of matrix algebra can be readily applied to study the structural properties of graphs from an algebraic point of view. The starting point in matrix representation is the adjacency matrix.

3.2.1. System reliability adjacency matrix

Consider a case of CCPP system having N systems leading to symmetric adjacency matrix $\{0,1\}$ of order $N \times N$. Let r_{ij} represents the reliability of interconnection between system i and j such that $r_{ij} = 1$, if the reliability of i th system depends on the j th system, (in the digraph, this is represented by an edge (r_{ij}) between node i and j) and is equal to zero, otherwise.

The adjacency matrix A_c for the corresponding digraph (Fig. 4) is as follows:

$$A_c = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} & \begin{matrix} \text{Systems} \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \end{matrix} \quad (1)$$

The off-diagonal elements of this matrix (r_{ij}) represent reliability of interconnection between system i and j . Moreover, this matrix considers only the reliability of connections be-

tween the systems without taking the effect of systems reliability. To consider this effect, another matrix known as "Characteristic System Reliability Matrix" is defined.

3.2.2. Characteristic system reliability matrix

The presence of different systems of the CCPP is realized by defining a characteristic system reliability matrix $B_c = \{RI - A_c\}$. This matrix for system reliability digraph of CCPP (Fig. 4) is expressed as follows:

$$B_c = \{RI - A_c\} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} & \begin{matrix} \text{Systems} \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} R & -1 & -1 & 0 & 0 & 0 \\ 0 & R & -1 & 0 & 0 & 0 \\ -1 & 0 & R & -1 & 0 & 0 \\ 0 & -1 & 0 & R & -1 & -1 \\ 0 & 0 & 0 & 0 & R & -1 \\ 0 & 0 & 0 & -1 & -1 & R \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \end{matrix} \quad (2)$$

where I is the identity matrix and R represents its reliability of systems. Characteristic System Reliability Matrix is analogous to characteristic matrix in the graph theory [2]. Characteristic System Reliability Matrix does not include information about reliability of interdependencies among different systems. The determinant of characteristic system reliability matrix called as characteristic system reliability polynomial, is written as follows:

$$\det\{B_c\} = R^6 + 4R^4 - 3R^3 + 3R^2 - 4R + 1 \quad (3)$$

The characteristic system reliability polynomial which is derived above is invariant of the system and it does not change by altering the labeling of systems. It is a reliability of the system.

In the above matrix, value of R is taken to be same for all the diagonal elements representing that all systems are considered to be identical. Due to this reason reliability polynomial is nonunique. It has been reported in the literature [3] that many graphs may have same characteristic polynomial that is co-spectral graph. In practice, in a CCPP all the six systems do not possess the same reliability. To incorporate distinct systems reliability and reliability of interconnections between them, a matrix called variable characteristic system reliability matrix is proposed.

3.2.3. Variable Characteristic System Reliability Matrix (VCSRM) of CCPP

A variable characteristic system reliability matrix T_a for a combined cycle power plant is defined taking into account reliability of systems and interconnection defined by the system reliability digraph (Fig. 4). This matrix is the combination of two matrices D_a and F_a . Let the off-diagonal elements of a matrix, F_a , representing the reliability of connection between systems is denoted by r_{ij} instead of 1, whenever system i is connected to system j with $i, j = 1, 2, 3, 4, 5, 6$ and 0 otherwise. Diagonal elements of the matrix F_a are 0. Another matrix D_a , is a diagonal matrix with its variable diagonal elements R_i ($i = 1, 2, 3, 4, 5, 6$) representing the reliability of six systems and all the non-diagonal elements are 0. For system reliability digraph of CCPP (Fig. 4) the VCSRM $T_a = [D_a - F_a]$ abbreviated as VCM-r is written as

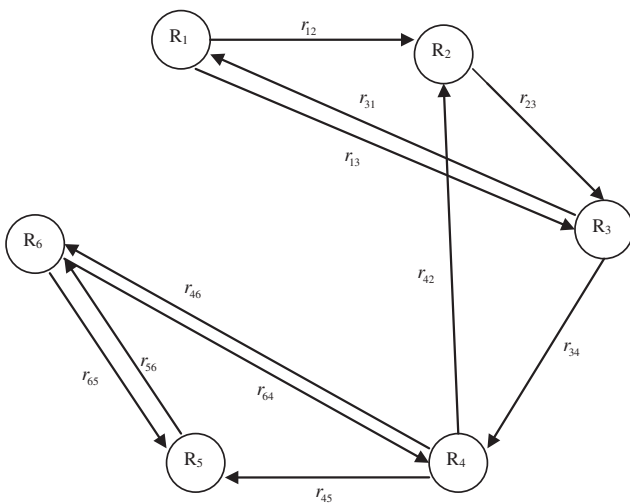


Figure 4 System reliability digraph for combined cycle power plant.

$$T_a = [D_a - F_a] = \begin{array}{c|cccccc|c} & 1 & 2 & 3 & 4 & 5 & 6 & \text{Systems} \\ \hline & R_1 & -r_{12} & -r_{13} & 0 & 0 & 0 & 1 \\ & 0 & R_2 & -r_{23} & 0 & 0 & 0 & 2 \\ & -r_{31} & 0 & R_3 & -r_{34} & 0 & 0 & 3 \\ & 0 & -r_{42} & 0 & R_4 & -r_{45} & -r_{46} & 4 \\ & 0 & 0 & 0 & 0 & R_5 & -r_{56} & 5 \\ & 0 & 0 & 0 & -r_{64} & -r_{65} & R_6 & 6 \end{array} \quad (4)$$

The determinant of VCSRM is called variable characteristic system reliability multinomial, denoted as VCF-r for matrix (4), and is written as follows:

$$\begin{aligned} \text{Per}[T_a] = & [R_1 R_2 R_3 R_4 R_5 R_6 + (r_{13} r_{31})(R_2 R_4 R_5 R_6) \\ & + (r_{46} r_{64})(R_1 R_2 R_3 R_5) + (r_{56} r_{65})(R_1 R_2 R_3 R_4) \\ & - R_1 R_2 R_3 (r_{45} r_{56} r_{64}) - R_1 R_5 R_6 (r_{23} r_{34} r_{42}) \\ & - R_4 R_5 R_6 (r_{12} r_{23} r_{31}) + R_2 R_4 (r_{13} r_{31})(r_{56} r_{65}) \\ & + R_2 R_5 (r_{13} r_{31})(r_{46} r_{64}) - R_1 (r_{56} r_{65})(r_{23} r_{34} r_{42}) \\ & - R_2 (r_{13} r_{31})(r_{45} r_{56} r_{64}) - R_4 (r_{56} r_{65})(r_{12} r_{23} r_{31}) \\ & - R_5 (r_{46} r_{64})(r_{12} r_{23} r_{31}) + (r_{12} r_{23} r_{31})(r_{45} r_{56} r_{64})] \end{aligned} \quad (5)$$

Every term in the $\text{Per}[T_a]$ is representing the part of the systems and interrelations. For example the $R_1 R_2 R_3 R_4 R_5 R_6$ shows that all six systems are linked to each other. Failure of any of the system will trip the plant. But the multinomial (5) is also unsuitable for reliability analysis and Variable permanent system reliability matrix is defined for better interpretation of the results.

3.3. Variable permanent system reliability matrix (VPSRM) for CCPP

The negative sign in Eq. (5) indicates subtraction of reliability information about loops of systems and does not give a true picture of the CCPP reliability. Taking into consideration this fact, a variable permanent system reliability matrix (VPSRM) T_c abbreviated as VPM-r for the combined cycle power plant is written as follows:

$$T_c = [D_c + F_c] = \begin{array}{c|cccccc|c} & 1 & 2 & 3 & 4 & 5 & 6 & \text{Systems} \\ \hline & R_1 & r_{12} & r_{13} & 0 & 0 & 0 & 1 \\ & 0 & R_2 & r_{23} & 0 & 0 & 0 & 2 \\ & r_{31} & 0 & R_3 & r_{34} & 0 & 0 & 3 \\ & 0 & r_{42} & 0 & R_4 & r_{45} & r_{46} & 4 \\ & 0 & 0 & 0 & 0 & R_5 & r_{56} & 5 \\ & 0 & 0 & 0 & r_{64} & r_{65} & R_6 & 6 \end{array} \quad (6)$$

Where the R_i s, r_{ij} s, D_c , and F_c are with same meaning as in the matrix of expression (4).

The permanent of VPSRM is called the variable permanent system reliability function and is abbreviated as VPF-r. The only difference between matrices (4) and (6) is in the signs of the off-diagonal elements. In the VCSRM, expression (4), the off-diagonal elements r_{ij} have negative signs, while these are positive in the VPSRM of expression (6). VPF-r for matrix (6) is written as:

$$\begin{aligned} \text{Per}[T_c] = & [R_1 R_2 R_3 R_4 R_5 R_6 + (r_{13} r_{31})(R_2 R_4 R_5 R_6) \\ & + (r_{46} r_{64})(R_1 R_2 R_3 R_5) + (r_{56} r_{65})(R_1 R_2 R_3 R_4) \\ & + R_1 R_2 R_3 (r_{45} r_{56} r_{64}) + R_1 R_5 R_6 (r_{23} r_{34} r_{42}) \\ & + R_4 R_5 R_6 (r_{12} r_{23} r_{31}) + R_2 R_4 (r_{13} r_{31})(r_{56} r_{65}) \\ & + R_2 R_5 (r_{13} r_{31})(r_{46} r_{64}) + R_1 (r_{56} r_{65})(r_{23} r_{34} r_{42}) \\ & + R_2 (r_{13} r_{31})(r_{45} r_{56} r_{64}) + R_4 (r_{56} r_{65})(r_{12} r_{23} r_{31}) \\ & + R_5 (r_{46} r_{64})(r_{12} r_{23} r_{31}) + (r_{12} r_{23} r_{31})(r_{45} r_{56} r_{64})] \end{aligned} \quad (7)$$

Comparing expressions (5) and (7), that is, the VCF-r and VPF-r, respectively, for the CCPP systems of Fig. 2, it may be noted that all the terms are exactly the same in both expressions. However, they differ in their signs. In VPF-r expression (Eq. (7)), all the terms carry positive signs; while in the VCM-r of expression (5) both positive and negative signs appear in the multinomial. This multinomial (Eq. (7)) uniquely represents the reliability of CCPP system of Fig. 2 and includes all the information regarding various constituents as systems and interactions among them.

A physical meaning is associated with each term of permanent function [3]. Permanent function for Fig. 4 is written as Eq. (7) and graphical representation of different terms is shown in Fig. 5.

A computer program is developed using C++ language for calculating the values of permanent function for square matrix of $N \times N$ matrix.

3.4. Combined cycle power plant real-time reliability index (RTRI_{CCPP})

Concept of real-time reliability index (RTRI) was proposed first time for a steam power plant (SPP) by Mohan et al. [2]. It was defined as the ratio of its reliability under real-time conditions to the reliability under its designed conditions. The reliability of a CCPP decreases regularly with time due to various reasons such as non-availability of some of the systems/subsystems or due to aging effect, etc. Performance of a combined cycle power plant in any case cannot be higher than its designed value. Therefore, for all practical purposes, real-time performance of a CCPP is judged with respect to its designed performance. In view of this, the RTRI for combined cycle power plant is defined as the ratio of real-time reliability, i.e. (Reliability)_{RT} to the designed reliability that is (Reliability)_D. Mathematical expression is as following:

$$\text{RTRI}_{\text{CCPP}} = \frac{(\text{Reliability})_{\text{RT}}}{(\text{Reliability})_{\text{D}}} = \frac{(\text{VPF-r})_{\text{RT}}}{(\text{VPF-r})_{\text{D}}} \quad (8)$$

To calculate this index, the values of R_i and r_{ij} are required to be replaced in Eq. (7). Faisal et al. [18] explained that if data regarding the variables from some previous research or field study are available then it can be used to determine the index. But in case no quantitative values are available and in order to avoid complexity at system or subsystem level, then values for inheritance and interrelation may be taken from Table 1 and 2 respectively. From the literature it has been found that Wani and Gandhi [19] have used data from previous research for selecting the values of the variables while Kulkarni [20] had used a questionnaire to measure each attribute in terms of weightage to arrive at the values of the variables.

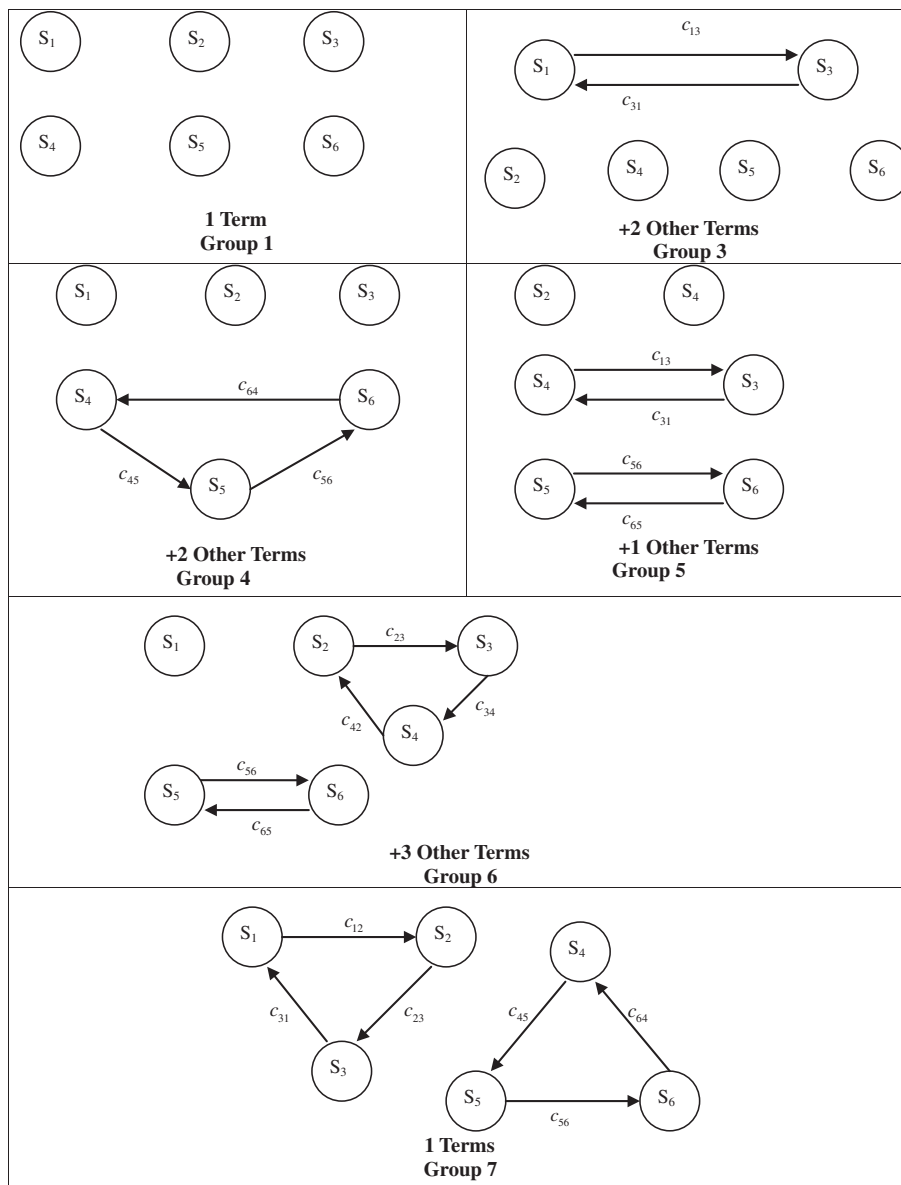


Figure 5 Graphical representation of permanent function (Eq. (7)) of CCPP corresponds to digraph (Fig. 4).

Table 1 Quantification of factors affecting combined cycle power plant reliability.

S. no.	Qualitative measure of parameters affecting combined cycle reliability	Assigned value of parameter (R_i)
1	One failure in 8 h	1
2	One failure in 24 h	2
3	One failure in 80 h	3
4	One failure in 350 h	4
5	One failure in 1000 h	5
6	One failure in 2500 h	6
7	One failure in 5000 h	7
8	One failure in 10,000 h	8
9	One failure in 25,000 h	9

It is worth to notice that one can choose any scale for R_i or r_{ij} [18,19]. The user may opt for an appropriate scale, for example, 0–5, 0–10, 0–50 or 0–100 for R_i 's and r_{ij} 's, but the final

Table 2 Quantification of interdependencies/off-diagonal elements.

S. no.	Qualitative measure of interdependencies	Assigned value of r_{ij}
1	Very strong	5
2	Strong	4
3	Medium	3
4	Weak	2
5	Very weak	1

ranking is not affected as these are relative values. However, lower scale value is desirable to obtain a manageable value of $RTRI_{CCPP}$ and also to reduce partisanship. Index value may differ from plant to plant because every system and interdependency has different values. In this way, different power

plants may be arranged in ascending or descending order, according to their reliability index value.

4. Step-by-step procedure for determining $RTRI_{CCPP}$

A methodology is the key for the evaluation of $RTRI_{CCPP}$ for different combined cycle power plants. In the present work, methodology based on graph theory and matrix method is developed for evaluating current value of combined cycle power plant reliability and it may be compared either with an ideal case or with any other real life operating plant. The main steps of the proposed methodology are as follows:

- *Step 1.* Consider a combined cycle power plant. If it seems to be very large system then divide it into smaller subsystems (e.g., air compressor system, combustion chamber system, gas turbine system, HRSG system, steam turbine system, and water system). Identify the various system categories affecting the CCPP reliability.
- *Step 2.* Develop system structure graph for the reliability of CCPP system based upon the interaction among different subsystems.
- *Step 3.* Convert the system structure graph of CCPP into corresponding system reliability digraph with systems reliability as nodes and edges for the reliability of interconnections.
- *Step 4.* Develop the CCPP system reliability matrix corresponding to the CCPP system reliability digraph. This will be $N \times N$ matrix with diagonal elements of R_i and off-diagonal elements of r_{ij} . The value of inheritance R_i (diagonal elements) for each subsystem is decided by experts or data available in literature. The values of reliability of interactions r_{ij} (off-diagonal elements) are to be determined by the experts or data available in literature.
- *Step 5.* Calculate the permanent function of CCPP system reliability matrix for values of real-time reliability (Reliability) $_{RT}$ and designed reliability (Reliability) $_D$.
- *Step 6.* Calculate the ratio of real-time reliability and designed reliability as in Eq. (8). This is the value of $RTRI_{CCPP}$ which mathematically characterizes the reliability of any combined cycle power plant based on the different systems and their interdependencies.
- *Step 7.* Record the results of this study and document them for future analysis.

5. Illustrative examples

Step-by-step methodology explained in the last section is helpful in estimating the reliability at system level which may be extended to the subsystems level. For the demonstration of proposed methodology two examples are taken into consideration.

5.1. Example 1

Failure of bearing lubricating oil cooler in a combined cycle power plant is taken as first example. For smooth revolution of turbo generator (TG), the bearings are lubricated through the lube oil system. The hot oil from the bearing is cooled through water cooler before feeding back into the lube oil tank.

Suppose four coolers are used in series for this purpose. In the present analysis TG is considered a part of the steam turbine system. Now if one of the cooler is not available then the value of $RTRI_{CCPP}$ for the present case will be computed as follows:

- *Step 1.* Consider the combined cycle power plant shown in Fig. 1.
- *Step 2.* Block diagram for the reliability of CCPP system is shown in Fig. 2.
- *Step 3.* System reliability digraph (SRD) corresponding to the block diagram of CCPP (Fig. 2) is shown in Fig. 4.
- *Step 4.* Under design conditions, it is presumed that all the six systems and components are available at their designed reliabilities during plant operation. Let reliability of these six systems be $(R_i)_d$ ($i = 1, 2, \dots, 6$). Let reliability of interconnections under designed conditions is denoted by $(r_{ij})_d$ ($i, j = 1, 2, \dots, 6$ and $i \neq j$). It is also assumed that all these interconnections are also available during operation at designed reliability. Then the variable permanent system designed reliability matrix for combined cycle power plant under consideration, i.e. $(T_c)_d$ will be corresponding to matrix T_c (Eq. (6)).

$$(T_c)_d = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} & \begin{matrix} Systems \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} R_{1d} & r_{12d} & r_{13d} & 0 & 0 & 0 \\ 0 & R_{2d} & r_{23d} & 0 & 0 & 0 \\ r_{31d} & 0 & R_{3d} & r_{34d} & 0 & 0 \\ 0 & r_{42d} & 0 & R_{4d} & r_{45d} & r_{46d} \\ 0 & 0 & 0 & 0 & R_{5d} & r_{56d} \\ 0 & 0 & 0 & r_{64d} & r_{65d} & R_{6d} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \end{matrix} \quad (9)$$

The matrices T_c and $(T_c)_d$ are similar and number of nodes and interconnections among the nodes are same. If $(R_i)_d = R_i$ and $(r_{ij})_d = r_{ij}$, then the values of $Per(T_c)$ and $Per(T_c)_d$ will be equal. Therefore in this case $RTRI_{CCPP} = 1$. The $Per(T_c)_d$ value gives the measure of designed reliability of combined cycle power plant, i.e. under the conditions when all its systems and subsystems, and the interconnection between them are available at their designed reliability. This condition exists only during the performance guarantee tests, which are conducted at the time of handing over a newly commissioned power plant [2]. Thereafter, the reliability of various systems and subsystems during operation starts falling below their designed values, and is required to be restored back by adopting proper maintenance strategies [6]. On the other hand, the $Per(T_c)_{RT}$ represents reliability function of combined cycle power plant under normal operating conditions or real-time conditions when all the systems and subsystems are available but may not be operating at their designed reliabilities.

In real-time situation, if one of the four coolers are out, real-time reliability gets reduced to three-fourth of its design value and correspondingly the value of R_{5d} will also get reduced to its three-fourth. The real-time value of $(T_c)_{RT}$ in this case will be obtained by replacing in matrix (Eq. (9)), the values of R_{5d} and r_{56d} , r_{64d} and r_{65d} by their three-fourth values: three-fourth R_{5d} , three-fourth r_{56d} , three-fourth r_{64d} , and three-fourth r_{65d} , respectively. Since the reliability, R_{5d} , of steam turbine system is getting reduced; it will correspondingly limit the reliability of the interconnections connected with this

system (Fig. 4), e.g., r_{56d} , r_{64d} and r_{65d} . Then from the matrix, Eq. (9), the value of $(T_c)_{RT}$ is

$$(T_c)_{RT} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ R_{1d} & r_{12d} & r_{13d} & 0 & 0 & 0 \\ 0 & R_{2d} & r_{23d} & 0 & 0 & 0 \\ r_{31d} & 0 & R_{3d} & r_{34d} & 0 & 0 \\ 0 & r_{42d} & 0 & R_{4d} & r_{45d} & r_{46d} \\ 0 & 0 & 0 & 0 & 0.75R_{5d} & 0.75r_{56d} \\ 0 & 0 & 0 & 0.75r_{64d} & 0.75r_{65d} & R_{6d} \end{bmatrix} \quad \text{Systems} \quad (10)$$

- Step 5. Assuming that the reliability of R_{ids} and r_{ijd} is equal to unity then value for permanent function of matrix in Eq. (9), will be equal to 12 and matrix in Eq. (10) will be equal to 9.94.
- Step 6. RTRI is the ratio of real-time reliability, i.e. (Reliability) $_{RT}$ to the designed reliability that is (Reliability) $_D$ and is expressed as

$$RTRI_{CCPP} = \frac{(\text{Reliability})_{RT}}{(\text{Reliability})_D} = \frac{(\text{VPF-r})_{RT}}{(\text{VPF-r})_D}$$

Therefore the RTRI for this case is $= 9.94/12 = 0.828$ of its designed value.

Based on the value of RTRI, the operating staff can adjust the process performance, e.g., reduce the electricity generation so as to match with the real-time reliability value. In case plant is allowed to run above the reliability index, it will lead to inefficient and unsafe operations which may lead to safety hazards or complete shutdown at a later stage.

- Step 7. Record the results of this study and document them for future analysis.

5.2. Example 2

Fouling, erosion and rubbing wear are responsible for physical changes in the compressor blade geometry which causes the performance deterioration. Fouling is the accumulation of deposits on the blade surfaces causing an increase in surface roughness. The accumulation of deposits increases rapidly during the accumulation of operating time/cycles and then levels off to a fairly constant value where the aerodynamic forces prevent any further accumulation. Increased pressure losses due to fouling can be reduced by washing the engine periodically. But frequent engine washing increases engine erosion. Erosion is the removal of material from the blades surface by solid particles colliding with the blades. This material removal causes increased tip clearances and reduced chord lengths. Erosion has been observed to be more severe in the tip region at the rear of the compressor due to centrifugal forces causing the migration of solid particles to the outer diameter [26]. Rubbing wear is the removal of material from the rotor blade tips and knife edge seals due to contact between static and rotating parts. Typically rubbing wear occurs when compressor blade tips rub with the compressor casing. This is usually the result of the engine rotor flexing during heavy operating loads, or in a mismatch of thermal growth between the rotors and casing [27]. The increased loss due to this effect is more a function of engine cycles than total engine operating hours. The rate of increase in clearances is dependent upon

the operating loads imposed on the engine early in its operation. Mass flow and efficiency penalties for fouling, erosion, and wear are calculated with the help of sensors. In general practice the mass flow penalty for erosion and wear is then distributed for each stage across the entire compressor. The efficiency penalty is applied to the compressor section as a whole. A new pressure ratio is adjusted to maintain constant output. Fouling, erosion and rubbing wear affects the reliable availability of compressor and it comes down from the design reliability as mentioned by manufacturer. For a site condition, suppose no data is available regarding the effect of these three factors on compressor performance then an operator with his experience may estimate the intensity of these factors qualitatively based on surrounding and operating conditions. For example, if air is dustier then fouling and erosion are comparatively higher. Further these factors are independent from each other. For the reliability estimation of air compressor at off-design condition, GTA can be applied at subsystem level with the help of the methodology discussed in Section 4.

Under design conditions, it is presumed that all these factors (fouling, erosion, and rubbing wear) are absent. Let these factors are represented by F_i ($i = 1, 2, 3$) and their interdependency is denoted by f_{ij} ($i, j = 1, 2, 3$ and $i \neq j$). Here, the value of F_i is considered to be varying in-between 1 and 0. When any factor F_i is absent, which is the design condition, and it is not affecting the compressor performance then its value is 1. If the presence of factor F_i is so high that compressor stops working then its criticality is highest then it is assigned a value of zero. In other conditions a value in-between 1 and 0 may be assigned based on the observations of site conditions and operating experience. Then cause effect design matrix for these factors is represented as:

$$E_D = \begin{bmatrix} F_{1D} & f_{12D} & f_{13D} \\ f_{21D} & F_{2D} & f_{23D} \\ f_{31D} & f_{32D} & F_{3D} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

As all the factors are independent, therefore, all non-diagonal elements are zero. In real life operating system, at some time operator observes the presence of fouling, erosion and rubbing wear then in the absence of relevant practical data some qualitative value on the scale of 1–0 may be assigned as per the discussion in Section 3.4. Then suppose based on the observation the real-time cause and effect matrix becomes

$$E_{RT} = \begin{bmatrix} F_1 & f_{12} & f_{13} \\ f_{21} & F_2 & f_{23} \\ f_{31} & f_{32} & F_3 \end{bmatrix} = \begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0 & 0 & 0.92 \end{bmatrix} \quad (12)$$

Corresponding to expression (8) RTRI for compressor comes out to be as:

$$RTRI_{Compressor} = \frac{\text{Per}(E_{RT})}{\text{Per}(E_D)} = \frac{0.7866}{1} = 0.7866 \quad (13)$$

Therefore, RTRI for the compressor comes out to be 78.66% of its designed value and it will provide guidance for the quantification of inheritance of compressor in expression (9).

With the help of sensors and control systems, performance of the CCPP may achieve to its design value even if some of its components are deteriorated. But scheduled or unscheduled maintenance has to be done to maintain the reliability up to mark to circumvent complete shutdown. Further reliable working of the sensors and control systems are also dependent

on some other factors such as temperature. For measuring their reliability in real life operating conditions, the procedure adopted in example 2 may be extended. A digraph representing factors and interdependencies can be developed. After converting digraph into matrix form quantification may be done. If practical data is not available then some qualitative values based on experience may be assigned to inheritance and interdependencies.

Two examples discussed in this section, reveals that the proposed methodology can be applied for the development of reliability index in a real life operating power plant.

6. Conclusion

In this paper, graph theoretic approach has been applied to obtain real-time reliability index for a combined cycle power plant. For this purpose, CCPP has been divided into six systems as the CCPP is a very large system. Systems/subsystems affecting the reliable availability of power plant are identified. For successful implementation of graph theoretical methodology, it is required to develop digraph, matrix and permanent function based on reliability of systems and reliability of interconnections. The approach helps to express CCPP reliability in quantitative terms. Using this procedure, an appropriate maintenance strategy for any combined cycle power plant can also be recommended.

The proposed structural approach for the evaluation of real-time reliability index for a CCPP has the following features:

- Reliability assessment of power plant is more accurate with graph theory as quantitative measure of interrelations among different systems is taken care of.
- Graph theoretic model is flexible enough to add on different systems, subsystems of and interaction among them in reliability analysis of a CCPP.
- The methodology is proficient in quantifying the influence of various system, subsystems and parameters on reliability of power plant.
- The value of real-time reliability index is useful for designers in selecting an optimum design in terms of reliability from available alternatives.
- The real-time reliability index enables the plant manager to know the reliable availability of power plant on real-time basis which will help them to take commercial decision on real-time basis.
- Sensitivity analysis may be carried out to identify the critical component or system affecting the power plant reliability.

Practical implementation of the proposed methodology in a systematic manner will help power generation industry to identify, categorize, analyze and evaluate parameters responsible for CCPP reliability. Thus, CCPP reliability index will help an organization to carry out SWOT (strength–weakness–opportunities–threats) of their system and take strategic decision to achieve profitability through productivity.

Similarly, methodology can be developed for obtaining other RAM indices: availability and maintainability; including optimum selection, benchmarking, and sensitivity analysis for combined cycle power plant.

References

- [1] De Souza GFM. Thermal power plant performance analysis. London: Elsevier Butterworth-Heinemann; 2012.
- [2] Mohan M, Gandhi OP, Agrawal VP. Real-time reliability index of a steam power plant: a systems approach. *Proc Inst Mech Eng Part A: J Power Energy* 2008;222:355–69.
- [3] Mohan M, Gandhi OP, Agrawal VP. Systems modeling of a coal based steam power plant. *Proc Inst Mech Eng Part A: J Power Energy* 2003;217:259–77.
- [4] Mohan M, Gandhi OP, Agrawal VP. Real-time efficiency index of a steam power plant: a systems approach. *Proc Inst Mech Eng Part A: J Power Energy* 2006;220:103–31.
- [5] Garg RK, Agrawal VP, Gupta VK. Selection of power plants by evaluation and comparison using graph theoretical methodology. *Electr Power Energy Syst* 2006;28:429–35.
- [6] Mohan M, Gandhi OP, Agrawal VP. Maintenance criticality index of a steam power plant: a graph theoretic approach. *Proc Inst Mech Eng Part A: J Power Energy* 2004;218:619–36.
- [7] Goode KB, Moore J, Roylance BJ. Plant machinery working life prediction method utilising reliability and condition monitoring data. *Proc Inst Mech Eng Part E: J Process Mech Eng* 2000;214:109–22.
- [8] Bradt D. Use of reliability, availability and maintainability techniques to optimise system operation. *Hydrocarbon Process* 1997;76:63–5.
- [9] Gandhi OP, Agrawal VP, Shishodia KS. Reliability analysis and evaluation of systems. *Reliab Eng Syst Safety* 1991;32:283–305.
- [10] Carlier S, Coindoz M, Deneuve L, Garbellini L, Altavilla A. Evaluation of reliability, availability, maintainability and safety requirements for manned space vehicles with extended on-orbit stay time. *Acta Astronautica* 1996;38(2):115–23.
- [11] Ji G, Wu W, Zhang B, Sun H. A renewal-process-based component outage model considering the effects of aging and maintenance. *Electr Power Energy Syst* 2013;44:52–9.
- [12] Eti M, Ogaji S, Probert S. Integrating reliability, availability, maintainability and supportability with risk analysis for improved operation of the AFAM thermal power-station. *Appl Energy* 2007;84:202–21.
- [13] Haghifam MR, Manbachi M. Reliability and availability modelling of combined heat and power (CHP) systems. *Electr Power Energy Syst* 2011;33:385–93.
- [14] Carpaneto E, Chicco G, Mancarella P, Russo A. Cogeneration planning under uncertainty. Part I: Multiple time frame approach. *Appl Energy* 2011;88:1059–67.
- [15] Carpaneto E, Chicco G, Mancarella P, Russo A. Cogeneration planning under uncertainty. Part II: Decision theory-based assessment of planning alternatives. *Appl Energy* 2011;88:1075–83.
- [16] Tang J. Mechanical system reliability analysis using a combination of graph theory and Boolean function. *Reliab Eng Syst Safety* 2001;72:21–30.
- [17] Kehlhofer RH, Warner J, Nielsen H, Bachmann R. Combined cycle gas and steam turbine power plants. Tulsa: PennWell; 1999.
- [18] Faisal MN, Banwet DK, Shankar R. Quantification of risk mitigation environment of supply chains using graph theory and matrix methods. *Euro J Ind Eng* 2007;1(1):22–39.
- [19] Wani MF, Gandhi OP. Development of maintainability index for mechanical systems. *Reliab Eng Syst Safety* 1999;65:259–70.
- [20] Kulkarni S. Graph theory and matrix approach for performance evaluation of TQM in Indian industries. *TQM Mag* 2005;17(6):509–26.
- [21] Dev N, Samsher, Kachhwaha SS, Attri R. GTA-based framework for evaluating the role of design parameters in cogeneration cycle power plant efficiency. *Ain Shams Eng J* 2013;4(2):273–84.
- [22] Deo N. Graph theory with applications to engineering and computer science. New Delhi: Prentice Hall; 2007.

- [23] Raj T, Attri R. Quantifying barriers to implementing Total Quality Management (TQM). *Euro J Ind Eng* 2010;4(3):308–35.
- [24] Dev N, Samsher, Kachhwaha SS. System modeling and analysis of a combined cycle power plant. *Int J Syst Assur Eng Manage* 2012. doi: <http://dx.doi.org/10.1007/s13198-012-0112-y>.
- [25] Koenigsberger F, Tlustý J. *Machine tool structures*. London: Pergamon; 1970.
- [26] Tabakoff W. Compressor erosion and performance deterioration. *AIAA/ASME 4th joint fluid mechanics, plasma dynamics, and lasers conference*, vol. 37. Atlanta (GA): ASME Publication FED; 1986.
- [27] Zaita AV, Buley G, Karlsons G. Performance deterioration modeling in aircraft gas turbine engines. *J Eng Gas Turb Power* 1998;120:344–9.



Nikhil Dev is an assistant professor of mechanical engineering at YMCA University of Science and Technology, Faridabad, India. He obtained his M.E. degree from the Panjab University, Chandigarh in 2005. Presently he is pursuing his Ph.D. from Delhi University, Delhi. His areas of interest include combined cycle power plants, combustion and computational techniques. To his credit, he is having more than thirteen papers published in reputed national and international journals. He is an active member of combustion institute.



Samsher did B.Tech in Mechanical Engineering from HBTI Kanpur and M.Tech. and Ph.D. from IIT Delhi. He Served NTPC for about 5 years, National Power Training Institute about 8 years, NIT Jalandhar for 5 months and presently working in the Department of Mechanical Engineering for about 12 years and presently occupying the post of Professor. Thus, he has total about 25 years of experience. He has published number of research papers in various journals of high repute and presented many papers in the national/international conferences. He has been awarded “Consistently High Performance award” in the year 2000 and also honored by presenting “Scroll of honor” on millennium teacher’s day in 2000. Dr. Samsher also has discharged various administrative duties in the institutions he worked. He is fellow of institution of Engineers (India).



Surendra Singh Kachhwaha completed his B.E. degree in Mechanical Engineering from M.B.M. Engineering College, Jodhpur in 1985. He did his M.Tech. (1988) in Heat Power from Institute of Technology, BHU Varanasi and Ph.D. in Evaporative Cooling from IIT Delhi in 1996. Presently he is working as Professor and Head of Mechanical Engineering Department, School of Technology, Pundit Deendayal Petroleum University since May 2012. He is also performing his duty as Dean, Faculty of Engineering and Technology. He has a teaching experience of more than 23 years in the field of thermal engineering at undergraduate and postgraduate level. He has worked 17 years at Engineering College Kota and a tenure of 6 years at Delhi College of Engineering, Delhi. Dr. Kachhwaha has contributed around 16 technical publications in reputed national and international journals and more than 40 publications in national/international conferences. His research interests include **evaporative cooling, ice slurry generation, Trigeneration, and biodiesel production techniques**. Presently he is guiding five research students pursuing PhD degree. Dr. Kachhwaha has successfully completed various Research and consultancy projects sponsored by government agencies in last one decade. He is a recipient of young scientist award (1998), SERC visiting fellowship (1999), and INSA visiting Fellowship (2003) due to his applied research contribution. He is also a life member of various societies.



Rajesh Attri is an Assistant Professor in the Mechanical Engineering Department at YMCA University of Science & Technology, Faridabad (Haryana), India. He received his B.E. in Mechanical Engineering from MD University, Rohtak, India, in 2005 with Honors and his M.Tech. in Mechanical Engineering (Manufacturing & Automation) from the YMCA Institute of Engineering, Faridabad (Haryana), India, in 2008 with honors. Presently, he is pursuing his Ph.D. from YMCA University of Science & Technology, Faridabad. He has published 20 papers in referred international journals and 10 papers in national & international conferences proceedings and reviewed many papers for international journals. His area of research is Quality management, Production system life cycle and Application of MADM approaches in manufacturing environment.



G- Journal of Environmental Science and Technology

(An International Peer Reviewed Research Journal)

Available online at <http://www.gjestenv.com>

Effect of Indoor Air Pollution on Child Development: A Research Review

Debapriyo Samanta, Aditi Singh, Tarun Joshi and Rajeev Kumar Mishra*

Department of Environmental Engineering, Delhi Technological University, Delhi, INDIA

ARTICLE INFO

Received : 29 Apr 2013

Revised : 21 July 2013

Accepted: 29 July 2013

Key words:

Child development, IQ, PAHs,
Indoor air pollutants.

ABSTRACT

The air quality of our indoor environments majorly affects our health. Children are the most exposed to indoor air pollutants as they spend most of their time at home and indoors, the quality of air becomes a major concern in their course of development. While in rural areas the major indoor pollutants are the emissions from combustion sources, in urban areas the particulates and polycyclic aromatic hydrocarbons (PAHs) have chronic and long-term effects. The aim of this study is to review the available literature dealing with various effects of child development. The study showed the relationship between prenatal exposure to airborne PAHs and Child IQ. Children were monitored from in utero to 5 years of age, with determination of prenatal PAH exposure through personal air monitoring for the mothers during pregnancy. It was found that high PAHs level were inversely associated with IQ. The other study comprehended lung development with exposure to nitrogen dioxide, acid aerosols, particulate matter and elemental carbon. Lung function tests were performed annually for 8 years among the children and a decrease in Forced Expiratory Volume (FEV) was observed. The results of the study indicated that current ambient levels of air pollution have chronic and adverse effects on lung development in children leading to significant deficits in FEV as these children reach adulthood. In rural India, the risks associated with indoor air pollution are found to be extremely high. The study revealed that most of the rural households were exposed to acute upper and lower infections, the reason being the use of traditional fuels and poorly ventilated cooking areas. The study focuses on creating awareness towards the adverse effects of indoor air pollution specifically on children health, with reference to the existing studies.

1) INTRODUCTION

The effects of air pollution on health have been reported in many research studies over the past 30 years. These effects majorly include mortality, respiratory and cardiovascular diseases and hospitalizations, changes in lung function and asthma attacks. Air quality has significant effect on children, however, previous environmental regulation has not focused on the effects caused by air pollution exclusively in children. And despite the extent of this mounting problem, the health impacts on children due to exposure to indoor air pollutants have yet to become a central focus of research especially in developing countries. According to studies, prenatal exposure to airborne polycyclic aromatic hydrocarbons (PAHs) is inversely related to a child's intelligence [1]. There have also been studies which link urban air pollution with asthma risk and behavioural problems in children. New findings also link exposures to chemicals found in homes with childhood respiratory problems.

1.1 Indoor Air Pollutants

In the developed countries where energy efficiency improvements sometime make houses relatively airtight and

raising pollutant levels, indoor air pollution has been a serious concern. But now with the construction of more tightly sealed buildings, reduced ventilation, use of synthetic materials for building and furnishing the issue has been taken into account in the sub-continent as well. In India, the greatest threat from indoor pollution is to rural areas, where some 0.8 billion people continue to rely on traditional fuels such as firewood and charcoal for cooking and heating (WHO. 2013) [2]. Concentrations of indoor pollutants in households that burn traditional fuels are alarming. Burning such fuels produces large amount of smoke and other air pollutants in the confined space of the home, resulting in high exposure. In 1992, the World Bank designated indoor air pollution in the developing countries as one of the four most critical global environmental problems [3]. Daily averages of pollutant level emitted indoors often exceed current WHO guidelines and acceptable levels. Although many hundreds of separate chemical agents have

* Corresponding Author: Dr. Rajeev Kumar Mishra

Email address : rajeevmishraiitr@gmail.com

been identified in the smoke from bio-fuels, the four most serious pollutants are particulates, carbon monoxide, polycyclic organic matter, and formaldehyde. In this study our focus remains on the exposure to piperonyl butoxide (PBO), polycyclic aromatic hydrocarbons (PAHs) and phthalates.

1.2 Sources and their exposure

Indoor air pollution can begin within the building or can be drawn in from outdoors. **Volatile organic compounds** mainly originate from solvents and chemicals. The main indoor sources are perfumes, hair sprays, furniture polish, glues, air fresheners, moth repellents, wood preservatives and many other products used in the house. **Tobacco smoke** generates a wide range of harmful chemicals and is known to cause cancer. It is very well known that passive smoking causes a wide range of problems to the passive smoker, the person who is in the same room with a smoker and is not himself/herself a smoker, ranging from burning eyes, nose and throat irritation to cancer, bronchitis, severe asthma and a decrease in lung function. In several research papers it is showed that maternal smoking may play a significant role in childhood obesity. Data from the US Collaborative Perinatal Project, a study of 35,000 children born between 1959 and 1964, show that children of smokers had an increased risk of becoming overweight before the age of 8 compared with the offspring of non-smokers [4]. The link between maternal smoking and obesity was stronger in girls than in boys. If **pesticides** are used carefully and the manufacturers, instructions followed carefully they do not cause too much harm to the indoor air. But piperonyl butoxide (PBO) which is used to bolster the effects of pyrethroid pesticides, the most common pesticides used for professional pest control has been found to increase the risk of non-infectious cough among children exposed to these chemicals in the womb [5]. **Biological pollutants** include pollen from plants, mite, hair from pets, fungi, parasites and some bacteria. Most of them are allergens and can cause asthma, hay fever, and other allergic diseases. **Formaldehyde** is a gas that comes mainly from carpets, particle boards, and insulation foam. It causes irritation to the eyes and nose and may even cause allergies in some people. **Asbestos** is a major concern because it is suspected to cause cancer. **Radon** is a gas which is emitted naturally by the soil. It is due to modern houses having poor ventilation that it is confined inside the house causing harm to the dwellers. In a study from CCCEH (*Columbia Center for Children's Environmental Health*), researchers reported that children exposed to diethyl phthalate (DEP) and butylbenzyl phthalate (BBzP) have a greater risk of asthma-related airway inflammation [6]. Phthalates are used widely in consumer products, including plastics, vinyl flooring, and personal care products. Phthalate exposure can occur through ingestion, inhalation, and absorption through the skin.

1.3 Effects of Indoor Air Pollutants on Children Health

Nearly 2 million people a year die prematurely from illness attributable to indoor air pollution due to solid fuel use. Among these deaths, 44% are due to pneumonia, 54% from chronic obstructive pulmonary disease (COPD), and 2% from lung cancer. Nearly half of deaths among children under five years old from acute lower respiratory infections (ALRI) are due to particulate matter inhaled from indoor air pollution from household solid fuels [2]. But in urban areas due to the absence of direct smoke from indoor coal burning the problems can be subtle and do not always produce easily

recognized impacts on health. In urban states, the effects are majorly constrained to chronic problems. Two NIEHS-funded studies from the Columbia Center for Children's Environmental Health (CCCEH) at the Columbia University Mailman School of Public Health have showed that children exposed to piperonyl butoxide (PBO) in the womb had an increased risk of non-infectious cough that is they are more likely to have a cough, unrelated to a cold or the flu, when they were between 5 and 6 years old. The findings provide evidence that children's respiratory systems are susceptible to damage from contaminants during the prenatal period. Another research team performed a prospective epidemiologic study on 1,759 children from 12 communities in Southern California. The communities had a wide-range of exposures to air pollutants including particulate matter, acid aerosols, ozone, and nitrogen dioxide [7]. Lung function tests were performed annually for 8 years. Over the 8-year period, decreases in a measurement of lung function known as forced expiratory volume (FEV) were observed. Exposure to these pollutants was associated with statistically and, more importantly, clinically significant deficits in FEV. For example, the risk of low FEV was almost 5 times higher at the highest level of particulate matter exposure than at the lowest level. In another study a significant association between prenatal exposure to polycyclic aromatic hydrocarbons (PAHs) and attention deficit and symptoms of anxiety and depression among children ages 6-7 was found in the CCCEH longitudinal cohort study [8].

2) RESULTS AND DISCUSSION

2.1 Prenatal Airborne Polycyclic Aromatic Hydrocarbon Exposure and Child IQ at Age 5 Years

American Academy of Pediatrics (April 2013) determined a relationship between Polycyclic Aromatic Hydrocarbon exposures and a strong interdependence between exposure and poor health of the child IQ [1]. The Polycyclic aromatic hydrocarbons (PAHs) are released to air during incomplete combustion and/or pyrolysis of fossil fuel, tobacco, and other organic material. Although exposure is ubiquitous, high-risk groups both for disproportionate exposure to air pollution and for adverse health and developmental outcomes is represented by urban minority populations. In fact, 100% of the mothers in the Columbia Center for Children's Environmental Health (CCCEH) cohort had detectable levels of PAHs in prenatal personal air samples, and 40% reported environmental tobacco smoke (ETS) exposure during pregnancy. Among the 249 children 5 years of age, prenatal PAH exposure levels ranged from 0.49 ng/m³ to 34.48 ng/m³. A total of 140 (56.2%) of the 249 children were classified as having high PAH exposure (2.26 ng/m³), with 2.26 ng/m³ being the median for the entire cohort.

As shown in **Table 1**, the inverse associations between high/low PAH exposure and full-scale and verbal IQ scores remained significant after adjustment was done for covariates. The association with performance IQ was inverse but it was not significant. The associations between logarithmically transformed, continuous, PAH levels as the independent variable and IQ also were significant for full-scale IQ and verbal IQ. As discussed above, results from this cohort indicated that exposure to PAH air pollutants during pregnancy is a risk factor for developmental delay at age 3, which was also identified with the Bayley Scales of Infant

Development. The present analysis shows continued effects of prenatal PAH exposure on child IQ at age 5. After adjustment for potential confounders, full-scale and verbal IQ scores of the high- and low-exposure groups differed by 4.31 points and

term effects of air pollution on children's respiratory health [9]. The data for the study which was on pulmonary function were obtained by trained field technicians, who travelled to study schools annually from the spring of 1993 through the

Table 1: Associations Between Prenatal PAH Exposure and Children's Full-Scale IQ, Verbal IQ and Performance IQ

	Full Scale IQ		Verbal IQ		Performance IQ	
	β	P	β	P	β	P
PAH exposure (high /low)	-4.307	.007	-4.668	.003	-2.369	.170
ETS exposure	1.736	.289	1.803	.265	1.609	.368
Gender	3.632	.021	4.695	.002	2.354	.168
Maternal Education	5.233	.002	4.510	.007	4.969	.007
Ethnicity	1.440	.373	6.74	.000	.121	.453

Source: Perera. 2012

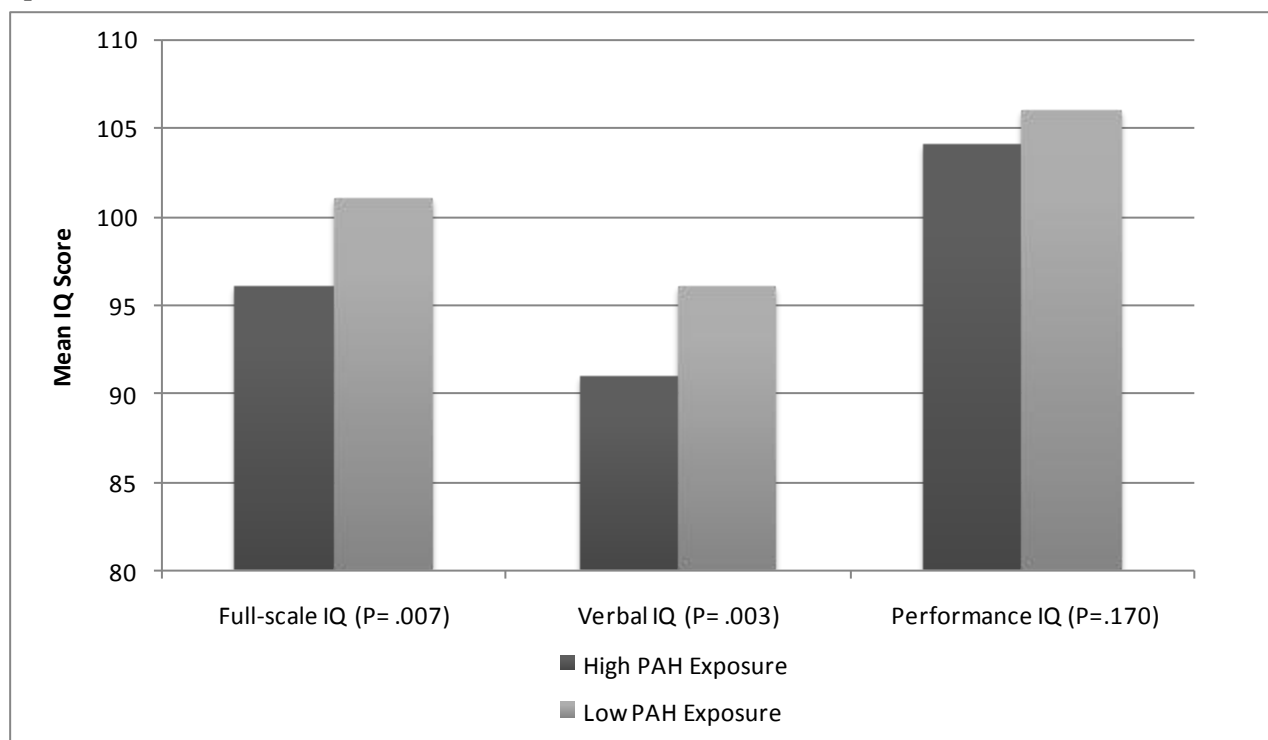
4.67 points, respectively. The present findings are of concern because verbal and full-scale IQ scores measured with the WPPSI (Wechsler Preschool and Primary Scale of Intelligence) during the preschool period were shown to be predictive of subsequent elementary school performance in a range of populations.

2.2 The Effect of Air Pollution on Lung Development from 10 to 18 Years of Age

There is substantial evidence to show that air pollution has

spring of 2001 to perform maximal-effort spirometric testing of the children. The three main measures of pulmonary function that are analyzed in the study are: forced vital capacity (FVC), forced expiratory volume in the first second (FEV), and maximal mid-expiratory flow rate (MMEF). Questions administered at the time of annual pulmonary-function testing were used to update information on asthma status, personal smoking status, and exposure to environmental tobacco smoke. In addition to examining the

Figure 1: Differences in full-scale, verbal and performance IQ scores associated with high levels of prenatal PAH exposure.



Source: Perera. 2012

chronic, adverse effects on pulmonary development in children. In 1993, the Children's Health Study recruited 1759 children (average age, 10 years) from elementary schools in 12 southern California communities to investigate the long-

growth in lung function over the eight-year period, study was done on the FEV measurements obtained in 746 subjects to determine whether exposure to air pollution was associated with clinically significant deficits in attained FEV. A low

FEV was defined as an attained FEV below 80 percent of the predicted value which is a common criterion used in clinical settings to identify persons who are at increased risk for adverse respiratory conditions. Among the girls, the average FEV increased from 1988 ml at the age of 10 years to 3332 ml at the age of 18 years, yielding an average growth in FEV of 1344 ml over the eight-year period.

The corresponding averages in boys were 2082 ml and 4464 ml, yielding an average growth in FEV of 2382 ml over the

The study resulted in providing robust evidence that lung development, as measured by the growth in FVC, FEV, and MMEF from the ages of 10 to 18 years, is reduced in children exposed to higher levels of ambient air pollution and is the major result of this study. The strongest associations were observed between FEV and a correlated set of pollutants, specifically nitrogen dioxide, acid vapour, and elemental carbon. The effects which these pollutants have on FEV were similar in boys and girls and remained significant among

Table 2: Mean Levels of Growth in Pulmonary Function during the Eight-Year Study Period, from 1993 to 2001.

Pulmonary-Function Measure	Girls			Boys		
	Age of 10 yr	Age of 18 yr	Average 8-yr growth	Age of 10 yr	Age of 18 yr	Average 8-yr growth
FVC (mL)	2262	3790	1528	2427	5202	2775
FEV (mL)	1988	3332	1344	2082	4464	2382
MMEF (mL/sec)	2311	3739	1428	2287	4709	2422

Source: Gauderman, 351 (11): 1-11.

eight-year period. Similar patterns of growth over the eight-year period were observed for FVC and MMEF (**Table 2**).

The differences which were estimated in the growth of FEV, FVC, and MMEF during the eight-year period with respect to all pollutants are summarized in **Table 3**. Deficits in the growth of FEV and FVC were observed for all pollutants, and deficits in the growth of MMEF were observed for all but ozone, with several combinations of outcome variables and

children with no history of asthma and among those with no history of smoking, suggesting that most children are susceptible to the chronic respiratory effects of breathing polluted air. The effects of air pollution on the growth in lung function was observed and it was found that its magnitude during this age interval was similar to those that have been reported for exposure to maternal smoking and smaller than those reported for the effects of personal smoking.

Table 3: Difference in Average Growth in Lung Function over the Eight-Year Study Period from the Least to the Most Polluted Community.

Pollutant	FVC		FEV		MMEF	
	Difference (95% CI) mL	P Value	Difference (95% CI) mL	P Value	Difference (95% CI) mL	P Value
O ₃						
10 am – 6 pm	-50.6	0.37	-22.8	0.62	85.6	0.40
1-Hour Maximal level	-70.3	0.20	-44.5	0.32	45.7	0.65
NO ₂	-95.0	0.05	-101.4	0.005	-211.0	0.02
Acid vapour	-105.2	0.03	-105.8	0.004	-165.0	0.07
PM ₁₀	-60.2	0.33	-82.1	0.08	-154.2	0.16
PM _{2.5}	-60.1	0.24	-79.7	0.04	-168.9	0.06
Elemental Carbon	-77.7	0.08	-87.9	0.007	-165.5	0.04
Organic Carbon	-58.6	0.37	-86.2	0.08	-151.2	0.19

Source: Gauderman, 351 (11): 1-11.

pollutants attaining statistical significance.

2.3 Associated diseases with indoor air pollution in the rural households in India

Field surveys revealed that most of the rural households (almost 97 per cent) used traditional fuels (fuel wood, dung, crop residues, coal, kerosene) along with traditional stoves for cooking food [10]. Consequently, women are regularly exposed to high levels of pollutants emitted from cooking fuels during cooking. Though the specific effects on Indian

identification of strategies for reducing levels of air pollutants is warranted especially in the Indian subcontinent.

REFERENCES

1. Perera, F. P., Li, Z., Whyatt, R., Hoepner, L., Wang, S.,

Table 4: Indoor air pollution associated diseases (in percentages) in the rural households in Aligarh, India

Occurrence of associated diseases	Percentage of households affected
1. Acute upper respiratory infection	70.84
2. Acute lower respiratory infection	48.29
3. Low birth weight	47.15
4. Chronic obstructive pulmonary diseases	46.70
5. Pre natal mortality	15.49
6. Eye irritation and cataract	14.12
7. Asthma	12.98
8. Pulmonary tuberculosis	10.93

Source: Abha Lakshmi Singh. 2012.

children have not been researched yet still many cases of low birth weight and pre-natal mortality have been recorded, which directly links the effect on children with air pollution caused by traditional fuels and stoves (*chulhas*) in rural India. An examination of **Table 4** shows the indoor air pollution associated diseases (in percentages) in the rural households. In a survey conducted in Aligarh in 2010, nearly all the households reported exposure to heat for more than 2 hours per day and to smoke for more than 1 hour per day, as most of them (97 per cent) use biomass fuels/traditional stoves for cooking purpose. When these biomass fuels are burnt in inefficient *chulhas* they emit lot of smoke. More than half of the households (65 per cent) reported of living in *kutchha*/semi-*pucca* houses which absorb the emitted pollutants from the combustion of biomass fuels for longer duration. Of the total sample, 71 per cent reported of acute upper respiratory infection (AURI), 48 per cent of acute lower respiratory infection (ALRI), 47 per cent of low birth weight, 47 per cent of chronic obstructive pulmonary diseases (COPD), 15 per cent of prenatal mortality, 14 per cent of eye irritation and cataract, 13 per cent of asthma and 11 per cent of pulmonary tuberculosis.

3. CONCLUSION

On the basis of various research papers reviewed it can be concluded that environmental PAHs at levels encountered in the ambient air can affect child IQ scores adversely. And since IQ is an important predictor of consequent academic performance of a child the study of PAHs is a potential concern and further experiments must be carried on to confirm the results. Also, a reduced IQ may affect the overall mental health of the individual. Our observation of associations between air pollution and all three measures of lung function — FVC, FEV, and MMEF — suggest that there are certainly some unknown mechanisms which lead to reduced lung development when exposed to pollutants. People at higher risk are the rural children who are exposed to smoke for long hours because of the lack of good housing and infrastructure facilities. Given the magnitude of the observed effects and the importance of lung function as a determinant of morbidity and mortality during adulthood, continued emphasis on the

- Camann, D. and Rauh, V. 2009. Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. *Pediatrics*, 124(2), e195-202.
2. WHO. 2013. *Indoor Air Pollution and Health*. Online April 10, 2013 from: <http://www.who.int/medicentre/factsheets/fs292/en/>.
3. TERI. 2007. *Indoor Air Pollution*. Available from: <http://edugreen.teri.res.in/explore/air/indoor.htm>. [Accessed: 10/4/2013].
4. Chen, A., Pennell, M. L., Klebanoff, M. A., Rogan, W. J. and Longnecker, M. P. 2006. Maternal smoking during pregnancy in relation to child overweight: Follow-up to age 8 years. *International Journal of Epidemiology*, 35(1), 121-130.
5. Liu, B., Jung, K. H., Horton, M. K., Camann, D. E., Liu, X., Reardon, A. M., Perzanowski, M. S., Zhang, H., Perera, F. P., Whyatt, R. M. and Miller, R. L. 2012. Prenatal exposure to pesticide ingredient piperonyl butoxide and childhood cough in an urban cohort. *Environment International*, 48, 156-161.
6. Just, A. C., Whyatt, R. M., Miller, R. L., Rundle, A. G., Chen, Q., Calafat, A. M., Divjan, A., Rosa, M. J., Zhang, H., Perera, F. P., Goldstein, I. F. and Perzanowski, M. S. 2012. Children's urinary phthalate metabolites and fractional exhaled nitric oxide in an urban cohort. *American Journal of Respiratory and Critical Care Medicine*, 186(9), 830-7; doi:10.1164/rccm.201203-0398OC. Online August 23, 2012.
7. Cornell, A. G., Chillrud, S. N., Mellins, R. B., Acosta, L. M., Miller, R. L., Quinn, J. W., Yan, B., Divjan, A., Olmedo, O. E., Lopez-Pintado, S., Kinney, P. L., Perera, F. P., Jacobson, J. S., Goldstein, I. F., Rundle, A. G. and Perzanowski, M. S. 2012. Domestic airborne black carbon and exhaled nitric oxide in children in NYC. *Journal of Exposure Science and Environmental Epidemiology*, 22(3), 258-66; doi: 10.1038/jes.2012.3 Online February 29, 2012.
8. Perera, F. P., Tang, D., Wang, S., Vishnevetsky, J., Zhang, B., Diaz, D., Camann, D. and Rauh, V. 2012. Prenatal Polycyclic Aromatic Hydrocarbon (PAH) Exposure and Child Behavior at age 6-7. *Environ Health*

Perspect, 120(6), 921-6; doi: 10.1289/ehp.1104315
Online March 22, 2012.

9. Gauderman, W. J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., Margolis, H., Bates, D., and Peters, J. 2004. The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine*, 351(11), 1-11.
10. Singh, A. L. and Jamal, S. 2012. A study of risk factors associated with indoor air pollution in the low income households in Aligarh city, India. *E3 Journal of Environmental Research and Management*, 3(1), 001-008.

Ensemble based Active Annotation for Biomedical Named Entity Recognition

Mridula Verma

Department of Computer Science Engineering
Delhi Technological University
Delhi, India
verma.mridula@gmail.com,

Utpal Sikdar, Sriparna Saha, Asif Ekbal

Department of Computer Science and Engineering
Indian Institute of Technology Patna
Patna, India
{utpal.sikdar,sriparna, asif }@iitp.ac.in

Abstract—Active Learning is an important prospect of machine learning for information extraction to deal with the problems of high cost of collecting labeled examples. It makes more efficient use of the learner's time by asking them to label only instances that are most useful for the trainer. We propose a novel method for solving this problem and show that it favorably results in the increased performance. Our proposed framework is based on an ensemble approach, where Decision Tree and Memory-based Learner are used as the base learners. The proposed approach is applied for solving the problem of named entity recognition (NER) in biomedical domain. Results show that the proposed technique indeed improves the performance of the system significantly.

Keywords—Name Entity Recognition, Decision Tree, Memory-based Learning, Ensembled Classifier, Biomedical Domain

I. INTRODUCTION

Information Extraction is one of the most popular field in which various machine learning techniques have been applied. One difficulty with these techniques is the high cost of collecting labeled examples. Active learning [1] is, nowadays, a popular research area due to its many-fold potential benefits. It can lead to drastic reductions in the amount of annotation that is necessary for training a highly accurate statistical classifier. It can make more efficient use of the learner's time by asking them to label only instances that are most useful for the trainer.

Named Entity Recognition (NER) has important role in many Natural Language Processing (NLP) application areas such as information extraction, information retrieval, machine translation and automatic summarization etc. The applications related to NER in Biomedical Domain, like mentions of proteins, DNA, RNA etc., have started to emerge only very recently. As part of the biomedical domain, there are some existing works that cover a few techniques [2,3,4,17,18].

One of the major challenges for NER in biomedical domain is the unavailability of annotated corpus. There are many existing techniques for the news-wire domains, however it was realized that there are not much work published in

biomedical domains. The major difficulties associated with the identification and classifications of biomedical NEs are discussed in [2]. It is both time consuming and cost sensitive to prepare the large amount of data required to achieve good performance for NER. In contrast we can acquire a good amount of unlabeled documents with relatively lower costs. Thus it would be effective if we can come up with an alternative technique that will be capable enough to efficiently select the unlabeled data to be added to the initial training data. This way we can create the good amount of training data automatically.

In this paper we propose a novel active learning technique based on the concept of classifier ensemble, where Decision Tree (DT) [5] and Memory Based Learner (MBL) [6] are used as the underlying base classification methods. Based on the *confidence interval* between the two most probable classes for each token of the development data from both Decision Tree and Memory Based Learner, we select most uncertain samples from the unlabeled data to be added to the initial training data. The outputs of DT and MBL are combined together using a weighted voting approach.

The proposed approach is evaluated for GENIA version 3.02 corpus of the GENIA project. A very rich feature set developed in [2,17] that includes variety of features based on orthography, local contextual information and global contexts is used in the current paper. Evaluation results show that our proposed approach in general performs well for different datasets. We also observe that our proposed ensemble based active learning method performs better compared to each of the baselines developed using either of the base classifiers, i.e. DT or MBL. It also performs superior compared to a random sampling based active annotation technique.

Early works in the field of biomedical Name Entity Extraction are mainly of two classes i.e. rule-based and machine learning based approaches. Rule based approaches [7, 8] are not mainly used for datasets with inherent complex structures, due to the significant difficulty in designing the systems. A very good knowledge of domain is required, so to obtain a high performance in these models is very difficult.

Including this, they also suffer from the problem of adaptability to new domains as well as new named-entity types. In contrast, machine learning based approaches are easy to adapt new named-entities and relatively less expensive to maintain. Various supervised approaches in NE extraction in biomedical texts are discussed in [9,10,11]. A way of comparing the existing biomedical NE extraction systems is provided by tagged GENIA corpus[12].

One of the significant characteristic of our system is that feature identification and selection tasks are mostly done in absence of any domain knowledge or resources [17]. The approach proposed in [13] was applied for two popular resource-constrained Indian languages, namely Hindi and Bengali. In contrast, here, we evaluate the approach for the biomedical NE extraction, which is more challenging. A diverse set of effective features that are useful to achieve good performance is identified and implemented [17]. The proposed approach is evaluated on the benchmark datasets of JNLPBA [14]. Evaluation results show the precision, recall and F-measure values of 74.02%, 78.28%, 75.9%, respectively.

The remainder of this paper is organized as follows. Section II briefly describes about the supervised classifiers that we used. Section III describes the proposed active learning technique. Section IV sheds light on the set of features that we have identified and implemented for biomedical NE extraction. Section V reports the dataset, evaluation scheme and results. We conclude our work in section VI and discuss some areas of developing this idea further.

II. PROPOSED ACTIVE LEARNING TECHNIQUE

In this section we describe the two base classifiers, namely Decision Tree and Memory Based Learner.

A. Decision Tree Classifier

Decision Tree Classifiers (DTC's) has been successfully applied for solving different pattern recognition problems like remotely sensed multisource data classification, medical diagnosis, speech, and character recognition, to mention a few. Major characteristics of DTC are the capability to use different feature subsets and decision rules at different stages of classification and the capability of tradeoffs between classification accuracy and time/space efficiency. In the current paper decision tree algorithm C4.5 is used as the underlying classification tool.

In order to apply DT for solving NER problems, C4.5 requires that text data to be given in a feature-vector format. Here each named-entity (NE) phrase is consisting of one or more words, plus some external information, i.e., words in the close neighborhood of the NE phrase.

B. Memory Based Learner

Memory-based learning is another machine learning technique which uses K-nearest neighbor principle. Here examples can be re-used directly in processing natural language problems. Training examples are stored in a memory without any modification. During the testing process, the most similar examples from the training data are searched, and new examples are classified based on the classes of these training data [6].

The training phase of MBL consists of storing all the training instances in memory. In testing phase, in order to classify a query, at first it has to be compared with all the training instances. This means distance between query point and all the training instances have to be calculated. Then we search for k nearest training instances. Based on the majority of classes of these k instances final class is assigned to the test query [6]. We have used IGTREE algorithm in this system.

III. PROPOSED ACTIVE LEARNING TECHNIQUE

In this section we describe our proposed active learning technique to select the informative samples from the unlabeled data. It is based on an ensemble approach, where two supervised classifiers, namely Decision Tree and Memory Based Learner are used.

Active annotation is one of the many advanced applications of active learning which can help in significant reductions in the amount of annotation that is necessary for training a highly accurate statistical classifier. Traditionally, Random Sampling approach is used, in which, unlabeled data is selected for labeling at random. In contrast, in active learning, the most useful data for the classifier are carefully selected. In a typical active learning setup, a classifier is trained on a small sample of the data (usually selected randomly), known as the seed examples. The classifier is subsequently applied to a pool of unlabeled data with the purpose of selecting additional examples that the classifier views as informative. The selected data is annotated and the cycle is repeated, allowing the learner to quickly refine the decision boundary between the classes. The key question in this approach is how to determine the samples that will be most useful to the classifier.

For each of the base classifiers, DT and MBL, a feature vector consisting of the features described in the following section is extracted for each word in the annotated data. Now, we have a training data in the form (w_i, Ti) , where, w_i is the i -th word and its feature vector and Ti is its output class. We consider the feature vector consisting of all the features of the current token and varied the contexts within $w_{i-2}^{j+2} = w_{i-2} \dots w_{i+2}$. Based on some selection criterion, sentences are chosen from the development set and added to the initial training set in such a way that the performance on the test set improves.

Our technique is based on the combined decisions of both DT and MBL. C4.5 algorithm for decision tree performs classification by constructing an N-ary tree that optimally separates data into various categories; whereas IGTree algorithm of MBL also converts the training set to a tree structure. Gain ratio values of the features are utilized to determine tree ordering. After training the base classifiers i.e. generating the trees, we first evaluated them. Then, we computed confidence value of each token for each output class. As both algorithms produce two different kinds of probabilistic scores, we first normalize all the confidence values within the range [0, 1]. We consider these as the actual confidence scores of outputs.

Our proposed selection criterion is based on the differences between the confidence values of the most probable two classes for a token, the hypothesis being that items for which this difference is smaller are those of which the classifier is less certain. A threshold on the confidence interval is defined, and for each base classifier we generate a set of uncertain samples. These sets contain the selected sentence identifiers along with the confidence intervals for which they are included into the respective sets. Thereafter we combine the decisions of base classifiers and generate a new set of uncertain samples by taking the unions of these two sets. The union is taken in such a way that the common sentence is assumed to have the confidence interval, equal to the minimum of two values assigned to that particular sentence in two sets. Finally, we select 10 most uncertain sentences from the development data. Thus, in each iteration of the algorithm, we actually add 10 most informative sentences to the training set. We run the algorithm for the maximum 10 iterations. In some cases, the performance starts to decrease even at the earlier step of the algorithm. In order to account this fact we stop the algorithm's iteration, and retrieve last iteration's training data as the final one. The main steps of the proposed active annotation are shown in Figure 1.

IV. NAMED ENTITY FEATURES

We used the following features [2,17] for constructing ensemble-based learner. These features are easy to derive and don't require deep domain knowledge and external resources for their generation. Due to the used of variety of features, the individual classifiers acquire high accuracies.

1) **Context words:** These are the words occurring within the context window $w_{i-3}^{i+3} = w_{i-3} \dots w_{i+3}$, $w_{i-2}^{i+2} = w_{i-2} \dots w_{i+2}$ and $w_{i-1}^{i+1} = w_{i-1} \dots w_{i+1}$, where w_i is the current word. This feature is considered with the observation that surrounding words carry effective information for the identification of NEs.

2) **Word prefix and suffix:** These are the word prefix and suffix character sequences of length up to n . The sequences are stripped from the leftmost (prefix) and rightmost (suffix)

positions of the words. We set the feature values to 'undefined' if either the length of w_i is less than or equal to $n-1$, w_i is a punctuation symbol or if it contains any special symbol or digit. We experiment with $n=3$ (i.e., 6 features) and 4 (i.e., 8 features) both.

3) **Word Length:** We define a binary valued feature that fires if the length of w_i is greater than a pre-defined threshold. Here, the threshold value is set to 5. This feature captures the fact that short words are likely not to be NEs.

4) **Infrequent word:** A list is compiled from the training data by considering the words that appear less frequently than a predetermined threshold. The threshold value depends on the size of the dataset. Here, we consider the words having less than 10 occurrences in the training data. Now, a feature is defined that fires if w_i occurs in the compiled list. This is based on the observation that more frequently occurring words are rarely the NEs.

5) **Part-of-Speech (PoS) information:** PoS information is a critical feature for entity extraction. In this work, we use PoS information of the current and/or the surrounding token(s) as the features. This information is obtained using GENIA tagger V2.0.2[12], which is a freely available well-known system used to extract PoS information from the biomedical texts. The accuracy of the GENIA tagger is 98.26%.

6) **Chunk information :** We use GENIA tagger V2.0.2 to get the chunk information. Chunk information (or, shallow parsing features) provide useful evidences about the boundaries of biomedical entities. In the current work, we use chunk information of the current and/or the surrounding token(s). This is an important feature for MBL.

7) **Unknown token feature:** This is a binary valued feature that checks whether the current token was seen or not in the training corpus. In the training phase, this feature is set randomly.

8) **Word normalization:** We define two different types of features for word normalization. The first type of feature attempts to reduce a word to its stem or root form. This helps to handle the words containing plural forms, verb inflections, hyphen, and alphanumeric letters. The second type of feature indicates how a target word is orthographically constructed. Word shapes refer to the mapping of each word to their equivalence classes. Here each capitalized character of the word is replaced by 'A', small characters are replaced by 'a' and all consecutive digits are replaced by '0'. For example, 'IL' is normalized to 'AA', 'IL-2' is normalized to 'AA-0' and 'IL-88' is also normalized to 'AA-0'.

9) **Word class feature:** Certain kinds of entities, which belong to the same class, are similar to each other. The word class feature is defined as follows: For a given token, capital letters, small letters, numbers and non-English characters are converted to "A", "a", "O" and "-.", respectively. Thereafter,

the consecutive same characters are squeezed into one character. This feature will group similar names into the same NE class.

10) **Informative words:** In general, biomedical NEs are too long and they contain many common words that are actually not NEs. In this feature we have selected the most important effective words for the identification of NEs.

11) **Head nouns:** Head noun is the major noun or noun phrase of a NE that describes its function or the property. In this work, we use only the unigram and bigram head nouns like receptor, protein, binding protein etc. For domain independence, we extract these head nouns from the training data only.

12) **Content words in surrounding contexts:** This feature is semantically motivated and exploits global context information. This is based on the content words in the surrounding context. We consider all unigrams in contexts $w_{i-3}^{i+3} = w_{i-3} \dots w_{i+3}$ of w_i (crossing sentence boundaries) for the entire training data. We convert tokens to lower case, remove stopwords, numbers, punctuation and special symbols. We define a feature vector of length 10 using the 10 most frequent content words. Given a classification instance, the feature corresponding to token t is set to 1 if and only if the context w_{i-3}^{i+3} of w_i contains t . Evaluation results show that this feature is very effective to improve the performance by a great margin. We used the PoS information of test instances (extracted from the GENIA tagger) to compute this feature for GENIA.

13) **Orthographic features:** We define a number of orthographic features depending upon the contents of the wordforms. Several binary features are defined which use capitalization and digit information. These features are: initial capital, all capital, capital in inner, initial capital then mix, only digit, digit with special character, initial digit then alphabetic, digit in inner. The presence of some special characters like ('', '-', '.', ',') etc.) is very much helpful to detect NEs, especially in biomedical domain. For example, many biomedical NEs have '-' (hyphen) in their construction. Some of these special characters are also important to detect boundaries of NEs. We also use the features that check the presence of ATGC sequence and stop words. The complete list of orthographic features is shown in [2,17].

14) **Verb trigger:** These are the special type of verbs (e.g., binds, participates etc.) that occur preceding to NEs and provide useful information about the NE classes. However, for the sake of domain independence, we do not use a predefined list of trigger words. Based on their frequencies of occurrences, these trigger words are automatically extracted from the training data. A feature is then defined that fires iff the current word appears in the list of trigger words.

V. DATASET, EXPERIMENT AND DISCUSSION

In this section, we present the description of performance metrics, dataset, experimental setup, and report the detailed evaluation results of our proposed approach.

A. Performance Measure

Classifiers are measured in terms of Recall, Precision and F-measure values. Precision is the ratio of the number of correctly found NE chunks (more than one token) to number of found NE chunks. The value of F-measure,

which is the weighted harmonic mean of recall and precision, is calculated as below,

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

where, $\beta = 1$. We used JNLPBA 2004 shared task evaluation script [12] to measure recall, precision and F-measure.

B. Dataset and Experimental Setup

The dataset was extracted from the GENIA version 3.02 corpus of the GENIA project. The training data used in the task came from the GENIA version 3.02 corpus.

The decision-tree induction algorithm C4.5 [3] is used here. For the experimental purpose, different levels of tree pruning are carried out, resulting in decision trees of various sizes. 10-fold cross-validation is performed at each size in order to gain an unbiased estimate of the performance of the system on unseen data. Here data set is divided into ten equally-sized subsets and at a particular run nine sets are considered together as a training set for developing the decision tree and tenth set is considered as a test set for evaluation. The final result is the average over ten runs.

In IGTREE used with the memory-based learner, we had the tree building step, in which we get a lossless version of the training examples and the tree searching step, in which the tree is searched top-down. IGTREE is regarded as the fastest algorithm of MBL due to its small size and capability to search top-down, committing to any matches made and stopping search when a mismatch occurs. In order to measure the importance of a feature, gain ratio weight is used in determining the classification of instances. These reasons forced us to choose the IGTREE algorithm for MBL. All experiments were conducted with number of nearest neighbors, $k=3$. Gain ratio was used as the feature weight due to its effectiveness. All experiments were done using the Overlap metric and the majority voting technique. The experiments were executed using the Tilburg Memory Based Learner TiMBL[6].

C. Result Analysis

In our experiment, we first generated the decision tree using C4.5 algorithm and trained the IGTree using features as described in section IV. We conducted a number of experiments with the various context sizes within the context window of w_{i-2}, \dots, w_{i+2} , and various subsets of the features. We observed the best performance with the context of w_{i-1}, w_i, w_{i+1} , and thus only report its results. Results are shown in Tables 1 and 2.

In Table 1, we present a comparison of results obtained with traditional learning technique with random selection and the proposed active learning technique with threshold values 0.2 and 0.3 both. Here, in each iteration of the algorithm, 10 most effective (informative) sentences are removed from the development set and added to the training set. The overall F-measure value with threshold 0.2 is 73.71% (with

Step 1: Train the base classifiers with the initial training data and evaluate with the gold standard test data.
Step 2: Train the base classifiers with the initial training data and evaluate with the development data.
Step 3: Calculate the confidence value of each token for each output class.
Step 4: Normalize the confidence scores within the range of [0, 1].
Step 5: Compute the confidence interval (CI) between the two most probable classes for each token of the development data. This is computed on the outputs of both DT and MBL.
Step 6: From each token of the development data, perform the following operations:
Step 6.1: if CI is below the threshold value (set to 0.2) then add the NE token along with its sentence identifier and CI in a set of effective sentences, selected for active annotation.
Step 6.2: Create two different sets, (<i>Set DT</i> and <i>Set MBL</i>) for two classifiers.
Step 7: Combine two sets into one, named as EA in such a way that if the sentence identifiers are same, then for that sentence $CI_{new} = \min(CI_{DT}, CI_{MBL})$. All the dissimilar sentences are added as they are.
Step 8: Sort EA in ascending order of CI_{new} .
Step 9: Select the top most 10 sentences, and remove these from the development data.
Step 10: Add the sentences to the training set. This generates new training set. Retrain the DT and MBL classifiers and evaluate with the test set.
Step 11: Repeat steps 3-10 for some iteration (10 in our case).

Fig. 1 Main steps of the proposed ensemble technique for active annotation

TABLE 1. EVALUATION RESULTS OF ACTIVE LEARNING WITH (A) THRESHOLD = 0.2 (B) THRESHOLD = 0.3 (C) RANDOM SELECTION. HERE R=RECALL, P=PRECISION AND FM=F-MEASURE (ALL VALUES ARE IN %)

Iteration No.	Threshold=0.2			Threshold=0.3			Random		
	R	P	FM	R	P	FM	R	P	FM
1	76.8	69.6	73.02	77.4	69.8	73.4	75.9	68.84	72.2
2	76.97	69.71	73.16	77.42	69.9	73.47	75.93	68.86	72.2
3	76.98	69.72	73.17	77.43	69.9	73.47	76.01	68.9	72.28
4	77.14	69.72	73.24	77.43	69.95	73.5	76.01	68.91	72.29
5	77.15	69.83	73.31	77.43	69.99	73.52	76.03	68.92	72.3
6	77.2	69.9	73.37	77.51	70.02	73.57	76.04	68.94	72.32
7	77.22	70.0	73.43	77.58	70.4	73.81	76.09	68.97	72.35
8	77.22	70.4	73.65	77.6	70.43	73.84	76.1	69.02	72.39

9	77.23	70.41	73.66	77.7	70.44	73.89	76.11	69.03	72.39
10	77.28	70.45	73.71	78.0	70.6	74.12	76.13	69.04	72.41

corresponding recall and precision values 77.28% and 70.45%, respectively) and with threshold 0.3 is 74.12% (with corresponding recall and precision values of 78.0% and 70.6%, respectively) .

We also develop a baseline model, where in each iteration, 10 sentences are randomly chosen from the development set and added to training set. Results of this baseline show the recall, precision and F-measure values of 76.13%, 69.04% and 72.41%, respectively.

Table 2 shows the overall F-measure values for DT, MBL and the ensemble model. We have also applied the active learning technique proposed in [2] for DT and MBL. In each iteration, 10 most confusing sentences are selected for

addition in the training set by using the algorithm proposed in [2]. The algorithm is executed for 10 iterations along with DT and MBL. We conducted active learning experiments with the thresholds of 0.1, 0.2 and 0.3. However, we show the results only with 0.3 threshold values as it yielded better performance. The ensemble model yields the overall F-measure value of 74.12% with threshold 0.3. This is clearly superior to the individual baseline models. DT and MBL based classifiers along with active learning technique attain the final F-measure values of 69.7% and 73.8%, respectively. Thus ensemble technique attains improvements of 4.42% and 0.32% over DT and MBL based approaches, respectively.

TABLE 2. EVALUATION RESULTS OF ACTIVE LEARNING WITH THRESHOLD = 0.3

Iteration	DT	MBL	Combined
1	68.3	73.1	73.4
2	68.55	73.3	73.47
3	68.73	73.41	73.47
4	68.88	73.49	73.5
5	68.97	73.5	73.52
6	69.2	73.52	73.57
7	69.43	73.55	73.81
8	69.61	73.6	73.84
9	69.64	73.62	73.89
10	69.7	73.8	74.12

VI. CONCLUSION AND FUTURE SCOPE

In this paper we have proposed an ensemble based active annotation technique that could be helpful for many applications where there is a scarcity in the amount of available labeled data, and preparation of these data is both time consuming and cost-intensive. We develop an ensemble based model by combining two supervised classifiers, namely DT and MBL for solving the problem of NER. The proposed system is evaluated for the GENIA dataset of biomedical domain. Evaluation results show satisfactory performance.

In future, we would like to add some more features based on external resources like gene/protein dictionary. Use of genetic algorithms and multiobjective optimization (MOO) for feature selection in biomedical entity extraction will be an interesting experiment to be carried out.

VII. REFERENCES

- [1] B. Settles. Active learning literature survey. In "Computer Sciences Technical Report" 1648 University of Wisconsin-Madison, 2009.
- [2] S. Saha, A. Ekbal, M. Verma, U. K. Sikdar, M. Poesio, "Active learning technique for biomedical named entity extraction", ICACCI, pp. 835-841, 2012.
- [3] Haochang Wang, Tiejun Zhao, Hongye Tan, and Shu Zhang, "BIOMEDICAL NAMED ENTITY RECOGNITION BASED ON CLASSIFIERS ENSEMBLE", International Journal of Computer Science and Applications, Vol. 5, No. 2, pp. 1-11, 2008.
- [4] Ki-Joong Lee, Young-Sook Hwang, Seonho Kim, Hae-Chang Rim, "Biomedical named entity recognition using two-phase model based on SVMs", Journal of Biomedical Informatics, Volume 37, Issue 6, December 2004, Pages 436-447
- [5] J. Han and M. Kamber, "Data Mining: Concepts and Techniques". Morgan Kaufmann, 2000.
- [6] Daelemans, W. et al. (2001) TiMBL: Tilburg Memory Based Learner Reference Guide. Version 4.2. Technical Report ILK 02-01, Computational Linguistics Tilburg University
- [7] Tsuruoka, Y. and J. Tsujii. Boosting Precision and Recall of Dictionary-Based Protein Name Recognition. In: Proceedings of NLP in Biomedicine, ACL 2003. 2003. Sapporo, Japan. p. 41-48.
- [8] Hanisch, D., et al. 2003. Playing biology's name game: Identifying protein names in scientific text. Proc. Pacific Symposium on Biocomputing (PSB), 403-411.
- [9] Lee, K., Y. Hwang, and H. Rim. Two-Phase Biomedical NE Recognition based on SVMs, In Proceedings of NLP in Biomedicine, ACL 2003. 2003. Sapporo, Japan. p. 33-40.
- [10] Kazama, J., T. Makino, Y. Ohta, and J. Tsujii. Tuning support vector machines for biomedical named entity recognition. In: Proceedings of Workshop on NLP in the Biomedical Domain, ACL, 2002. 2002. Philadelphia. pp.1-8.
- [11] Shen, D., J. Zhang, G. Zhou, J. Su, and C. Tan. Effective Adaptation of Hidden Markov Modelbased Named Entity Recognizer for Biomedical Domain. In: Proceedings of NLP in Biomedicine, ACL 2003. 2003. Sapporo, Japan. p. 49-56.
- [12] <http://www.nactem.ac.uk/tsujii/GENIA/ERTask/report.html>
- [13] A. Ekbal, S. Saha and D. Singh: Ensemble based Active Annotation for Named Entity Recognition. 2012 Third International Conference on Emerging Applications of Information Technology (EAIT 2012), November 30-December 01, 2012, Kolkata.
- [14] J. Kim, T. Ohta, Y. Tsuruoka, Y. Tateisi, and N. Collier. Introduction to the bio-entity task at jnlpba. In N. Collier, P. Ruch, and A. Nazarenko, editors, Proc. Workshop on Natural Language Processing in Biomedicine and its Applications, pages 70-76, 2004.
- [15] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
- [16] Paliouras, G., V. Karkaletsis, G. Petasis and C. Spyropoulos, 2000. Learning decision trees for named-entity recognition and classification. Proceedings of the 14th European Conference on Artificial Intelligence (ECAI), Aug. 20-25, Berlin, Germany.
- [17] A. Ekbal, S. Saha, 2013. Stacked ensemble coupled with feature selection for biomedical entity extraction. Knowl.-Based Syst. 46: 22-32.
- [18] M. Krauthammer, G. Nenadic, 2004. Term identification in the biomedical literature. Journal of Biomedical Informatics. 37(6), pages 512-526.

Evaluating the Ambiguity of Fuzzy Clustering with Fuzzy Entropy

Seba Susan*, Saarthak Sankalp, Prakhar Porwal, Prasanna Kumar, Simrat Singh

Department of Information Technology, Delhi Technological University, Bawana Road, Delhi, India - 110042

* Email id: seba_406@yahoo.in

ABSTRACT

In this paper, we strive to interpret the significance of the solution for the categorisation problem by the fuzzy c-means clustering algorithm in terms of the fuzzy membership values obtained by clustering. We use the concept of fuzzy entropy for interpreting the results. A threshold value for fuzzy entropy is computed in our work for labelling sample data as ambiguous or of an uncertain class. Results on images from the Chinese Academy of Sciences- Institute of Automation (CASIA) iris database with features extracted as per Libor Masek's iris segmentation model confirm our argument.

Keywords: Fuzzy c-Means Clustering, Fuzzy Entropy, Fuzzy Memberships, Ambiguity of Fuzzy Results, Libor Masek's Iris Segmentation Model

1. INTRODUCTION

The fuzzy c-means clustering (FCM) algorithm, a popular unsupervised data clustering algorithm proposed by James Bezdek^[1], is commonly found affected by outlier and inlier problems. Several fuzzy clustering schemes were later introduced to overcome this defect of the conventional fuzzy c-means such as possibilistic fuzzy c-means^[2], which perform well in the presence of noise. The De Luca and Termini definition of fuzzy entropy^[3] is an ideal measure of the ambiguity or uncertainty surrounding the fuzzy memberships, with crisp membership values indicating low uncertainty. Usually, the defuzzification of fuzzy membership values for a categorisation problem is carried out by defining a threshold or searching for the maximum membership, as in the case of the fuzzy co-clustering algorithm for images proposed recently^[4]. The uncertainty or ambiguity of a fuzzy set, measured by the fuzzy entropy, is an indicative factor for fuzzy decision making^[5]. The fuzzy entropy was used recently for feature selection^[6,7] where the fuzzy memberships were derived from a distance based similarity measure. In this paper, we formulate a technique to gauge the efficiency of FCM by checking the resulting fuzzy memberships for their fuzziness. The outlier or suspiciously corrupt sample can be detected by a higher value of fuzzy entropy. We use the Shannon entropy to formulate the De Luca and Termini fuzzy entropy as it is proved to be a good identifier of the maximum entropy in chaotic conditions while the recent version of De Luca and Termini fuzzy entropy based on the non-extensive entropy^[8] is a good identifier of minimum fuzzy entropy in structured environments^[7]. The paper is divided into the following sections: section II contains a brief review about the fuzzy clustering and fuzzy entropy basics. Section III contains the results and Section IV draws the conclusions.

2. A REVIEW OF FUZZY CLUSTERING AND FUZZY ENTROPY AS A MEASURE OF AMBIGUITY

The typical categorisation problem involves assigning a class label to the sample data under test. Any of the well-known distance measures can be used for this purpose; for instance, many recent works advocate the use of Euclidean distance^[9], Mahalanobis distance^[10], Bhattacharya distance^[11], Minkowski distance^[12], etc.

The minimum distance between training and test feature vectors decides the class of the test sample. These distance measures, however, do not indicate the degree of belongingness of the sample data to a particular class and this is where fuzzy logic steps in. Fuzzy logic assigns a fuzzy membership value in the range 0 to 1 to the sample data that indicates the degree of its belongingness to a particular class. The class is assigned mostly depending on the maximum membership among all classes. Such a categorisation is justified only when the maximum membership value is clearly in a majority as compared with the membership to other classes. Fuzzy memberships range between 0 and 1 and a membership of 0.5 and above indicates strong membership to a class. However, the number of memberships greater than 0.5 may be more than one, and may not necessarily indicate the right answer. Also, the category memberships may lie very close to each other and simply taking the maximum value may not indicate the true class. The boundary line regarding the clear majority is ambiguous, thereby opening an interesting field of speculation in this direction. One solution is to use the unsupervised FCM algorithm^[1] which minimises a distance-based optimisation function given by the equation:

$$J_{FCM} = \sum_{c=1}^C \sum_{i=1}^N u_{ci} Dist(x_i, p_c) \quad (1)$$

subject to the constraint,

$$\sum_{c=1}^C u_{ci} = 1, \quad u_{ci} \in [0, 1], \quad \forall i = 1, \dots, N \quad (2)$$

Where, the symbols represent:

C, N : The number of clusters and data points, respectively

u_{ci} : Fuzzy membership function

$Dist(x_i, p_c)$: The dissimilarity term equal to the square of the Euclidean distance between pixel x_i and cluster centre p_c .

The constraint in Equation (2) ensures that the sum of all fuzzy memberships is 1; therefore, ideally, a clear majority membership would now be one that is greater than 0.5 for any one class as it automatically points to the fact that the sum of all the other memberships is less than 0.5. The iterative fuzzy c-means algorithm is composed of the following steps:

1. Initialise $U=[u_{ci}]$ matrix, $U^{(0)}$
2. At k^{th} iteration step: calculate the centroid vectors $P^{(k)}=[p_c]$ with $U^{(k)}$ by the following formula

$$p_c = \frac{\sum_{i=1}^N u_{ci}^m x_i}{\sum_{i=1}^N u_{ci}^m} \quad (3)$$

3. Update $U^{(k)}$ with $U^{(k+1)}$ given by the formula

$$u_{ci} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - p_c\|}{\|x_i - p_k\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

4. If $\|U^{(k)} - U^{(k+1)}\| < \varepsilon$ then STOP, otherwise return to step 2.

In our experiments, $m=2$ by popular choice, the error threshold ε in Step 4 is 0.01 and the number of clusters or categories C is equal to the number of individuals whose iris data are stored in the database. We need a measure to find how well the fuzzy memberships indicate the category and the fuzzy entropy is apt for this purpose. The fuzzy entropy is a concave function that is zero for extreme values of memberships of 0 and 1. The peak of this bell shaped curve is when all memberships equal to 0.5, i.e. the data belong equally to all classes.

The expression for fuzzy entropy given a set of fuzzy memberships $\{\mu\}$, as defined by De Luca and Termini [3] and based on the Shannon entropy functional [13], is given by the equation:

$$H = -\sum_{\mu} [\mu \log \mu + (1 - \mu) \log(1 - \mu)] \quad (5)$$

In the context of clustering results of FCM algorithm, the expression for fuzzy entropy for sample data i can be written as:

$$H_i = -\sum_c [u_{ci} \log u_{ci} + (1 - u_{ci}) \log(1 - u_{ci})], i=1, 2, \dots, N \quad (6)$$

It is to be noted that the constraint in Equation (2) holds for Equation (6) also. This implies that if a single membership is greater than 0.5 then the sum of all other memberships amounts to less than 0.5.

The essential conditions for a proper categorisation by FCM algorithm are as follows:

- 1) A membership greater than 0.5 for any one class.
- 2) The other class memberships should tend towards 0 in such a way that the sum of their memberships is lower than 0.5 (depends on the membership in Equation (1) since sum of all the memberships is equal to one).

The same is illustrated in an example below.

Suppose there are three classes of data C_1 , C_2 and C_3 ; therefore, $C=3$ in Equation (4). For an unambiguous clustering result, a data sample should belong to one of these three classes, say C_1 , with a fuzzy membership value greater than 0.5. We study the worst case scenario for an unambiguous clustering result. Substituting $C=3$ and $u_{ci} = 0.5$ for C_1 class ($c=1$) in Equation (4) yields the following equation for the sample data i :

$$H_i = -0.5 \log 0.5 - 0.5 \log 0.5 - \sum_{c=2}^3 [u_{ci} \log u_{ci} + (1 - u_{ci}) \log(1 - u_{ci})] \quad (7)$$

$$H_i = 0.6931 - \sum_{c=2}^3 [u_{ci} \log u_{ci} + (1 - u_{ci}) \log(1 - u_{ci})] \quad (8)$$

The second term in Equation (8) is maximum when $u_{ci}=0.25$ for both classes, leading to a confusion scenario as $\sum_{c=2}^3 u_{ci} = 0.5$ for the second term in Equation (8).

Equation (8) then reduces to:

$$H_i = 0.6931 - 2 \times (0.25 \log 0.25 + 0.75 \log 0.75) = 1.817 \quad (9)$$

Any value exceeding 1.817 for a three-class classification would lead to poor categorisation, while lesser values indicate efficient categorisation of the sample data i under test.

The generalisation of Equation (8) for any number of categories/classes C would be:

$$H_i = 0.6931 - (C-1) \times \left[\left(\frac{0.5}{C-1} \right) \log \left(\frac{0.5}{C-1} \right) + \left(1 - \frac{0.5}{C-1} \right) \log \left(1 - \frac{0.5}{C-1} \right) \right] \quad (10)$$

Substitution of $C=3$ would yield the result in Equation (9).

The fuzzy entropy values can be used to predict an outlier or noisy sample data which is equidistant from all clusters and tends to give equal category memberships. The highest values of fuzzy entropy are obtained in

this case. The maximum fuzzy entropy is obtained when the all memberships are equal to $\frac{1}{C}$, where C =number of clusters and it is then given by the equation:

$$H_i = -(C) \times \left[\left(\frac{1}{C} \right) \log \left(\frac{1}{C} \right) + \left(1 - \frac{1}{C} \right) \log \left(1 - \frac{1}{C} \right) \right] \quad (11)$$

For the three-class problem, the maximum fuzzy entropy is:

$$H_i = -(3) \times \left[\left(\frac{1}{3} \right) \log \left(\frac{1}{3} \right) + \left(1 - \frac{1}{3} \right) \log \left(1 - \frac{1}{3} \right) \right] = 1.9095 \quad (12)$$

Therefore, all data samples i , for which $1.817 < H_i \leq 1.9095$, are considered to be ambiguous data, not leading to clear categorisation. On the other hand, if H_i is less than or equal to 1.817, then its distance from 1.817 indicates the efficiency of the clustering process, the lower the value the better the categorisation.

3. EXPERIMENTAL RESULTS

We use Libor Masek's^[14] iris detection and segmentation technique, used popularly for segmentation and feature extraction representation of the iris image. Iris segmentation and feature extraction is of immense importance in biometric and security-based research areas^[15-19]. Three categories of iris images from the state-of-the-art CASIA^[20] database are considered for our experiment, with two images belonging to a single person/category, as shown in Figure 1. The iris is first segmented (Figure 2(a)) using circular Hough transform and the 2D Polar array feature matrix is constructed (Figure 2(b)) using convolution with 1D Gabor wavelet filter bank and subsequent phase quantising of result. Each row in the 2D polar array corresponds to a circular ring on the iris region. This two-dimensional feature matrix is converted to one dimension by concatenating all rows, which is used as the input for fuzzy clustering. The FCM algorithm is used to cluster the features derived from the six iris images shown in Figure 1. The parameters of the FCM algorithm are: number of clusters, $C=3$, $m=2$ and $\epsilon=0.01$, as discussed before. The result of fuzzy clustering is a set of three membership functions for each sample data. The argument of the maximum membership normally decides the class of the sample. As seen in Table 1, two samples each are categorized together

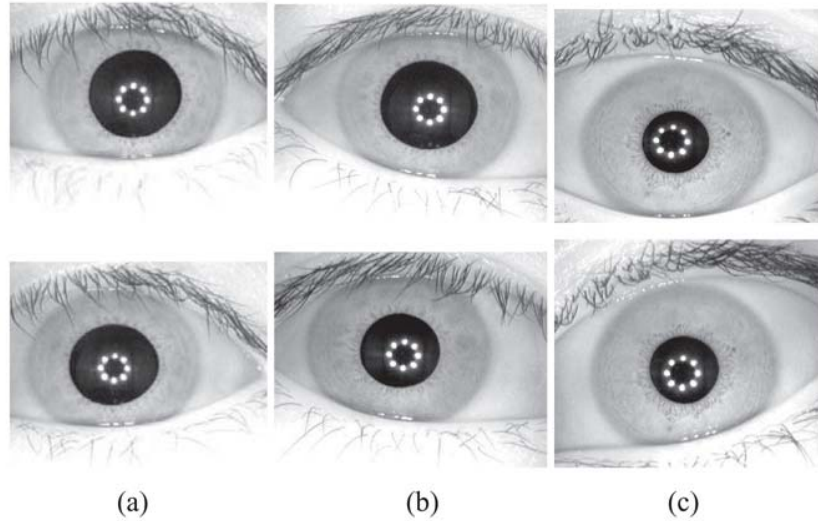


Figure 1: The six iris images from the CASIA database belonging to three persons with two samples per person; images are labelled 1 to 6 from left to right and top to bottom.

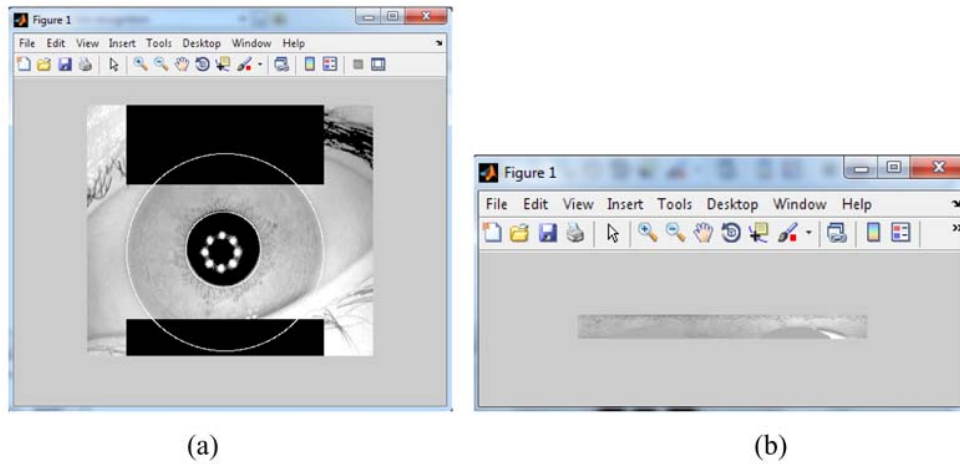


Figure 2: Iris segmentation and feature extraction from a single iris image by Libor Masek's model.

Table 1: The fuzzy memberships (maximum values highlighted in bold) and respective fuzzy entropy for each of the six iris images, obtained by fuzzy c-means clustering of the iris features

	u_{ci} (Image 1) ($i=1$)	u_{ci} (Image 2) ($i=2$)	u_{ci} (Image 3) ($i=3$)	u_{ci} (Image 4) ($i=4$)	u_{ci} (Image 5) ($i=5$)	u_{ci} (Image 6) ($i=6$)
Cluster 1 ($c=1$)	0.2494	0.2414	0.1918	0.1974	0.5297	0.4100
Cluster 2 ($c=2$)	0.5695	0.6053	0.1619	0.1700	0.2547	0.3260
Cluster 3 ($c=3$)	0.1811	0.1533	0.6463	0.6326	0.2156	0.2640
Fuzzy entropy for each data, H_i	1.718	1.652	1.581	1.610	1.780	1.885

forming three clusters correctly. However, sample 6 (shown in grey in Table 1 and observed to be visually distinct in Fig. 1) is ambiguous as its three fuzzy memberships are close to each other and none are clearly above 0.5. The fuzzy entropy is computed for each of the six test images and the results together with the fuzzy memberships are listed in Table 1. It is observed that, a low value of entropy (≤ 1.817) indicates a strong sample with ‘crisp’ values of membership inclined towards 0 or 1, while a high value of fuzzy entropy (> 1.817) indicate an outlier with memberships near to 0.5 for each class. In Table 1, it is observed that only iris image 6 has fuzzy entropy equal to 1.885 that is greater than the limit of 1.817, while image 3 turns out to be strongest sample with a low fuzzy entropy of 1.581. We therefore consider all results except that of Sample 6 as valid clustering results.

4. CONCLUSIONS

In this paper, we strive to solve the ambiguity regarding fuzzy memberships obtained by clustering with the fuzzy entropy measure with regards to the categorisation problem. The upper bound for fuzzy entropy is found to be 1.817 for an efficient three-class categorisation problem. The clustering results on features extracted from human iris images show that a high value of fuzzy entropy, above 1.817, for a three-class problem is sufficient justification to discard the sample data under test as being indecisive or corrupted by noise. The generalization of the upper bound of fuzzy entropy for any number of clusters or classes is also given.

REFERENCES

1. Bezdek, J.C. *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York, 1981.
2. Tjhi, W.C., and Chen, L. *Possibilistic fuzzy co-clustering of large document collections*, *Pattern Recognition*, Vol. 40, pp. 3452-3466, 2007.
3. Luca, A. De, and Termini, S. A definition of non-probabilistic entropy in the setting of fuzzy sets theory, *Information and Control*, Vol. 20, pp. 301-312, 1972.
4. Hanmandlu, M., Verma, O.P., Susan, S., and Madasu, V.K. Color segmentation by fuzzy co-clustering of chrominance color features. *Neurocomputing*, 2013, Available online at <http://dx.doi.org/10.1016/j.neucom.2012.09.043> [Accessed on 20 July 2013]
5. Klir, G.J., Folger, T.A., *Fuzzy Sets, Uncertainty and Information*, Pearson Education Inc., New Jersey, 1988
6. Luukka, P., Feature Selection Using Fuzzy Entropy Measures with Similarity Classifier, *Expert Systems with Applications*, 38, pp. 4600-4607, 2011
7. Susan, S., and Hanmandlu, M. A novel fuzzy entropy based on the non-extensive entropy and its application for feature selection. *FUZZ-IEEE*, Hyderabad, July 2013.
8. Susan, S., and Hanmandlu, M. A non-extensive entropy feature and its application to texture classification. *Neurocomputing*, 2013, Available online at <http://dx.doi.org/10.1016/j.neucom.2012.08.059>. [Accessed on 20 July 2013]
9. Lee, L.H., Wan C. H., Rajkumar, R., Isa, D . An enhanced support vector machine classification framework by using Euclidean distance function for text document categorization. *Applied Intelligence*, Vol. 37, Issue 1, pp. 80-99, 2012. .
10. Banerjee, M., Bandyopadhyay, S., and Pal, S.K. A clustering approach to image retrieval using range based query and Mahalanobis distance. *Rough Sets and Intelligent Systems-Professor Zdzisław Pawlak in Memoriam*. Springer, Berlin, Heidelberg, Vol. 43, 79-91, 2013.

11. Perronnin, F. *Bags of visual context-dependent words for generic visual categorization*. U.S. Patent No. 8,165,410. 24 April 2012.
12. Arruda, G. F. D., Costa, L. D. F., & Rodrigues, F. A. A complex networks approach for data clustering. *Physica A: Statistical Mechanics and its Applications*, vol. 391, Issue 23, pp. 6174-6183, 2012
13. Shannon, C.E., A Mathematical theory of communication, *The Bell System Technical Journal*, vol. 27, pp. 379-423, July 1948
14. Masek, L., *Recognition of human iris patterns for biometric identification*. Diss. Master's thesis, University of Western Australia, 2003 [Accessed on 20 July 2013] <http://people.csse.uwa.edu.au/~pk/studentprojects/libor/LiborMasekThesis.pdf>
15. Sanderson, S., and Erbetta, J. Authentication for secure environment based on iris scanning technology. *IEE Colloquium on Visual Biometrics*, London, page 8, January 2000.
16. Daugman, J. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1), 21-30, 2004.
17. Wolff, E. *Anatomy of the eye and orbit*. 7th edition. H. K. Lewis & Co. LTD, London, 1948.
18. Wildes, R. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, Vol. 85, No. 9, 1348-1363, 1997.
19. Boles, W., and Boashash, B. A human identification technique using images of the iris and wavelet transform. *IEEE Transactions on Signal Processing*, Vol. 46, No. 4, 1185-1188, 1998.
20. Chinese Academy of Sciences – Institute of Automation. Database of 756 grayscale eye images, Version 1.0. Available at <http://www.cbsr.ia.ac.cn/english/Databases.asp>, 2003. [Accessed on 20 July 2013]

Grey Relational Effort Analysis Technique using Regression Methods for Software Estimation

Nagpal Geeta¹, Uddin Moin², and Kaur Arvinder³

¹Department of Computer Science and Engineering, National Institute of Technology, India

²Faculty of Engineering and Technology, Delhi Technological University, India

³Department of Information Technology, University School of Information Technology, India

Abstract: *Software project planning and estimation is the most important confront for software developers and researchers. It incorporates estimating the size of the software project to be produced, estimating the effort required, developing initial project schedules, and ultimately, estimating on the whole cost of the project. Numerous empirical explorations have been performed on the existing methods, but they lack convergence in choosing the best prediction methodology. Analogy based estimation is still one of the most extensively used method in industry which is based on finding effort from similar projects from the project repository. Two alternative approaches using analogy for estimation have been proposed in this study. Firstly, a precise and comprehensible predictive model based on the integration of Grey Relational Analysis (GRA) and regression has been discussed. Second approach deals with the uncertainty in the software projects, and how fuzzy set theory in fusion with grey relational analysis can minimize this uncertainty. Empirical results attained are remarkable indicating that the methodologies have a great potential and can be used as a candidate approaches for software effort estimation. The results obtained using both the methods are subjected to rigorous statistical testing using Wilcoxon signed rank test.*

Keywords: *Software Estimations, estimation by analogy, fuzzy clustering, robust regression, GRA.*

Received April 9, 2012; accepted March 19, 2013

1. Introduction

Software project development is a creative process where each person's efficiency is different. It is difficult to plan and estimate at the beginning as most software projects have deficient information and vague associations amongst effort drivers and the required effort. Software developers and researchers are using different techniques and are more concerned about accurately predicting the effort of the software product being developed. Estimation by analogy (EbA) appears to be well suited to effort estimation, especially when the software product is poorly understood. It is concerned with finding solutions for a new problem based on known solutions from a set of similar projects. The motivation behind estimation by analogy is that the information retrieved from comparable software projects can help the management to improve the planning process, get superior in accurate bidding and for risk analysis and also similar projects tend to have similar costs. It uses a distance measure, in order to determine the similarity between two projects. EbA methods range from machine learning methods, regression techniques, and Grey Relational Analysis (GRA) a technique of Grey System Theory (GST), soft computing methods to a combination of these. In this research work grey relational analysis, a technique of GST has been applied as a similarity metrics between

two projects in integration to regression and fuzzy logic.

GST a recently developed system engineering theory first established by Deng [5] draws out valuable information by generating and developing the partially known information. It has been applied in different areas of Image Processing [14], mobile communication [29], machine vision inspection [15], decision making [18], stock price prediction [35], and system control [9]. The accomplishment of GST motivated us to examine its application in software effort estimation. In this study, GRA a technique of GST utilizes the concept of absolute point-to-point distance between cases [30]. In the first method, the focus is effort prediction using k nearest projects from total of n projects to the reference project based on their Grey Relational Grade's (GRG), and then regressing the effort of those k projects. In the second method, the grey relational coefficient uses fuzzy c-means algorithm to calculate the distance between two projects. Researchers have used various techniques from time to time for efficiently generating software effort estimates. Mukhopadhyay *et al.* [19] developed ESTOR, a CBR tool to estimate project effort. The metrics used by ESTOR are function point components and inputs to the intermediate COCOMO model. Shepperd *et al.* [27], expresses EbA in an automated environment known as ANaloGy software tool (ANGEL) that supports the collection, storage and

identification of the most analogous projects from the repository in order to estimate the cost and effort. It uses Euclidean distance as the distance measure to reduce the amount of computation involved. The research was carried out on six different datasets and it has outperformed the traditional algorithmic methods. Shepperd and Schofield [28] also validated nine different industrial datasets and concluded that in all cases analogy outperforms algorithmic models based on Step wise regression. Song *et al.* [30], proposed Grey Relational Analysis based on software Project Effort Prediction (GRACE). Huang *et al.* [13] made software effort estimation based on similarity distances. They applied genetic algorithm to analogy based software effort estimation models. It is used to derive linear model from the similarity distances between pairs of projects for adjusting the reused effort. Li *et al.* [17], describes a new flexible method called AQUA which combines the key features from two known analogy based estimation techniques: case based reasoning (CBR) and collaborative filtering (CF). The results have demonstrated better accuracy and broader applicability by combining techniques of CBR and CF with existing analogy-based effort estimation methods. Hsu and Huang [10] use weighted grey relational analysis for software development and have proposed six weighted methods to be integrated into GRA. Jorgenson and Shepperd [16] made a very systematic review, they considered 304 studies describing research on Software Cost Estimations. Mittas and Angelis [23] compared cost prediction models by re sampling techniques. They proposed the effect of iterated bagging on EbA and validated it using artificial and real data sets. Mittas *et al.* [22] also improved analogy based cost estimation by re sampling method. Mittas and Angelis [24] combined regression and estimation by analogy in a semi-parametric model for software cost estimation. The results were improved by the utilization of this semi parametric model. Azzeh *et al.* [2, 3] have used EbA based on the integration of fuzzy set theory with GRA. In order to improve the performance of analogy based estimation at early stages of software development, two new methods based on integration of GRA with regression and fuzzy have been proposed. In the first method, GRA has been applied in order to generate the grey relational grades for the objective projects with respect to the reference project. The projects are ranked and selected on the basis of grades, effort is estimated by regressing the effort on the other independent variables of k projects. The value of k varies with each reference project. In the second method, the grey relational coefficient of GRA uses fuzzy c-means algorithm to generate the distance between two projects. Both the methods have shown comparable results and also considerable improvement over GRACE [30], GRACE⁺ [31] and FGRA [2].

The remainder of the paper is organized as follows. In section 2, a brief review of GRA, Fuzzy logic and various regression techniques is provided. Section 3, describes the two proposed software effort prediction mechanisms using *GREAT RM* and *Fuzzy GRA*, Section 4, gives a brief discussion about the validation and evaluation criteria the datasets used and the empirical results obtained from evaluation of the proposed methodologies. Section 5, presents the conclusion and directions for future work.

2. Modeling Methods

2.1. Grey System Theory

GST works on unascertained systems with partially known and partially unknown information. Systems with completely unknown information are black systems. Systems with complete information available are called white systems. The term “Grey” lies between “Black” and “White” and it indicates that the information is partially available. GRA is one of the several aspects of the GST.

2.1.1. Grey Relational Analysis

GRA is comparatively a novel technique in Software Estimations for analyzing the relationships that exists between two series. The magnetism of GRA to software effort estimation shoots from its suppleness to model complex nonlinear relationship between effort and cost drivers [30]. The basic concepts of GRA is the Factor space and Grey Relational space.

- *Factor Space*: Let $p(X)$ be a theme characterized by a factor set X , and Q be an influence relation, $\{p(X); Q\}$ is a factor space. The factor space $\{p(X); Q\}$ have the following properties: Existence of key factors, Number of factors is limited and countable, Factor independence, Factor expansibility.
- *Comparable series*: Suppose $x_i = \{x_i(1), x_i(2), \dots, x_i(m)\}$, where $i = 0, 1, 2, \dots, n \in N; m \in N$, is a data series. This series is said to be comparable if, and only if it is Dimensionless, Scaled and Polarized.
- *Grey Relational Space*: If all the series in a factor space $\{p(X); Q\}$ are comparable, the factor space is a grey relational space which is denoted as $\{p(X); \Gamma\}$. In a grey relational space $\{p(X); \Gamma\}$, X is a collection of data series $x_i (i = 0, 1, \dots, n)$, in which $x_i = \{x_i(1), x_i(2), \dots, x_i(k)\}$, is the series; and $k = 1, 2, \dots, m$, are the factors. Γ , which is the Grey Relational Map set and based on geometrical mathematics, has the four properties: Normality, Symmetry, Entirety, and Proximity.

2.1.2. Grey Relational Analysis by Deng's Method

GRA is used to quantify all the influences of various factors and the relationship among data series that is a collection of measurements [5, 6, 7].

- **Data Processing:** Data Processing reduces the randomization and increase the regularity of data by using Upper-bound effectiveness (i.e., larger-the-better)[8]

$$x_i^*(k) = \frac{x_i(k) - \min_i x_i(k)}{\max_i x_i(k) - \min_i x_i(k)} \quad (1)$$

where $i=1,2,\dots,m$ and $k=1,2,\dots,n$. where $x_i(k)$ is the value of the k_{th} attribute in the i_{th} series; $x_i(k)$ is the normalized value of the k_{th} attribute in the i_{th} series; $\max_i x_i(k)$ and $\min_i x_i(k)$ are the maximum and minimum of the k_{th} attribute in all series;

- **Difference Series:** GRA uses the grey relational coefficient γ to describe the trend relationship between an objective series and a reference series at a given point in a system.

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \zeta \Delta_{\max}}{\Delta_{0,i}(k) + \zeta \Delta_{\max}} \quad (2)$$

where $\Delta_{0,i}(k) = |x_0(k) - x_i(k)|$ is the difference of the absolute value between $x_0(k)$ and $x_i(k)$; $\Delta_{\min} = \min_j \min_k |x_0(k) - x_j(k)|$ is the smallest value of $\Delta_{0,j} \forall j \in \{1, 2, \dots, n\}$; $\Delta_{\max} = \max_j \max_k |x_0(k) - x_j(k)|$ is the largest value of $\Delta_{0,j} \forall j \in \{1, 2, \dots, n\}$; and ζ is the distinguishing coefficient, $\zeta \in (0, 1]$. The ζ value will change the magnitude of $\gamma(x_0(k), x_i(k))$. In this study the value of ζ has been taken as 0.5 [2].

- **Grey Relational Grade:** GRG is used to find overall similarity degree between reference tuple x_0 and comparative tuple x_i . When the value of GRG approaches 1, the two tuples are “more closely similar”. When GRG approaches a value 0, the two tuples are “more dissimilar”. The GRG $\Gamma(x_0, x_i)$ between an objective series x_i and the reference series x_0 was defined by Deng as follows:

$$\Gamma(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n \gamma(x_0(k), x_i(k)) \quad (3)$$

2.2. Fuzzy Set Theory

Fuzzy set theory was introduced by Zadeh [37] and provides mathematical capabilities to handle ambiguous or vague concepts of human perception for complex systems problems, where it is extremely difficult to build the system models mathematically. It presents a structure to associate fuzzy sets to linguistic values. Every set is symbolized by Triangle, Trapezoidal, Gaussian, Sigmoid etc. and assigns a membership value between 0 and 1 for each point in the universe of disclosure. The membership value represents how much a point belongs to the fuzzy set [36]. Fuzzy models are generally used for simulation, identification of system behavior as well as for the prediction and control purposes.

2.3. Regression Techniques

Regression analysis is a statistical technique for modelling and analysis of variables. It is used to study the relationship that exists between dependent variable and one or more independent variables.

2.3.1. Ordinary Least Square Regression

It is the most popular and widely applied technique to build software cost estimation models. According to principle of least squares the ‘best fitting’ line is the line which minimizes the deviations of the observed data away from the line. It is referred to as multiple linear regression and is given by:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k} + \varepsilon_i \quad (4)$$

where, Y_i is a dependent variable where as x_1, x_2, \dots, x_k are k independent variables. β_0 is the y intercept, β_1, β_2 are the slope of y, ε_i is the error term. The corresponding prediction equation is given as

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \dots + \hat{\beta}_k x_{i,k} \quad (5)$$

In this equation $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ are the least square coefficients and \hat{y}_i is estimated response for i^{th} term.

Thus, the response estimated from the regression line minimizes the sum of squared distances between the regression line and the observed response. The least square method tries to minimize $\sum e_i^2$.

2.3.2. Robust Regression and M Estimation

It is a type of regression technique which prevails over limitations of OLS. OLS estimates are extremely non-robust to outliers. Outliers should be either treated or detached from analysis as they can inefficiently influence the whole process of fitting. Robust Regression (RR) [1, 32, 42] is an application of Iteratively Reweighted Least Squares (IRLS) regression in which the weights are iteratively set by taking the residual terms of the previous iteration into consideration. The least square method tries to minimize $\sum e_i^2$, which is unstable in case there are outliers present in the data, whereas the RR M Estimators tries to minimize the effect of these outliers by minimizing $\sum w_i^2 e_i^2$ in each iteration. The steps involved in the Iteratively Reweighted Least Square (ILRS) [20] are:

- **Step 1:** In the first iteration, each observation is allocated equal weight and the coefficients of the model are estimated using ordinary least squares (OLS).
- **Step 2:** In the second step, after the OLS, residuals are used to find weights. The observation with larger residual is assigned lower weight.

- *Step 3:* In the third iteration, the new model parameters and the residuals are recomputed using weighted least squares (WLS).
- *Step 4:* In step 4, new weights as per step 2 are found and the procedure continues until the values of the parameter estimates converge within a specified tolerance. The tuning constants play a significant role in the performance of estimators as they establish the shape and cutoff points of weighting functions. The objective function, weight function and tuning constant [11, 12, 21] for all the estimators is shown in Table 1.

Table 1. Estimators with weight functions.

Estimator	Objective function $\rho_H(e)$	Weight function $w_H(e)$	Tuning Constant
Bisquare	$\begin{cases} \frac{k^2}{6} \left[1 - \left(\frac{e}{k} \right)^2 \right]^3 & \text{for } e \leq k \\ \frac{k^2}{6} & \text{for } e > k \end{cases}$	$\begin{cases} \left[1 - \left(\frac{e}{k} \right)^2 \right]^2 & \text{for } e \leq k \\ 0 & \text{for } e > k \end{cases}$	4.685
Fair	$k^2 \left(\frac{ e }{k} - \log \left(1 + \frac{ e }{k} \right) \right)$	$\left(1 + \frac{ e }{k} \right)^{-1}$	1.400
Huber	$\begin{cases} \frac{1}{2} e^2 & \text{for } e \leq k \\ k e - \frac{1}{2} k^2 & \text{for } e > k \end{cases}$	$\begin{cases} 1 & \text{for } e \leq k \\ \frac{k}{ e } & \text{for } e > k \end{cases}$	1.345

2.3.3. Stepwise Regression

It is method for adding and removing terms based on their statistical importance. The forward approach starts with no variables in the model, trying out the variables one by one and including them if they are 'statistically important'. The selection has been used for estimating the effort of reference project from various similar projects. At each step, a predictor is entered based on partial F -tests or t test. The procedure continues till more variables can be justifiably entered. The first variable that is put in the stepwise model is the variable having the smallest t -test P -value (below $\alpha_E = 0.05$). The level of significance (α) is taken to be 5%.

3. Proposed Methodologies

3.1. Modeling Grey Relational Effort Analysis Technique with Regression Methods (GREAT_RM)

GREAT_RM focuses on project selection based on GRA and effort prediction by regression. In the GRA based studies so far, effort is estimated by generating similar projects to the target project and then estimating effort from the k most similar projects. This methodology, uses GRA for generating similar projects but effort is generated by applying regression on best k projects most similar to the target project. The value of k varies with each reference project. The structural framework of GREAT_RM is shown in Figure 1.

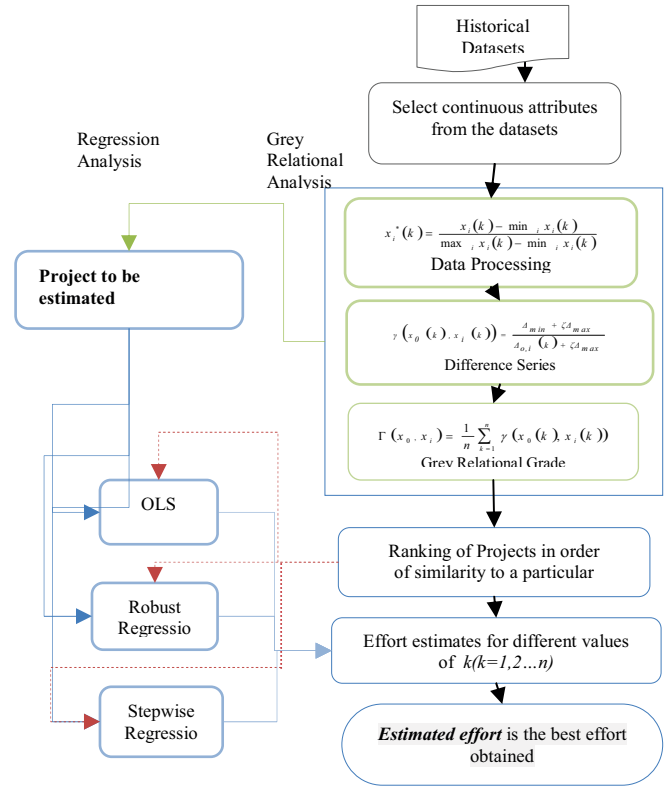


Figure 1. Structural Framework of GREAT_RM technique.

The basic steps of the methodology are:

- *Step 1.* Select continuous attributes from the dataset.
- *Step 2. Data series construction:* The data set consists of series $x_0 = \{x_1(1), x_1(2), \dots, x_1(m)\}$, $x_1 = \{x_2(1), x_2(2), \dots, x_2(m)\}$, $x_2 = \{x_3(1), x_3(2), \dots, x_3(m)\}$ and $x_n = \{x_n(1), x_n(2), \dots, x_n(m)\}$, x_0 is the reference series whose effort is to be estimated based on the objective series x_1, x_2, \dots, x_n .
- *Step 3. Data Preparation:* The numerical features are normalized in a specified range so that each feature has same weight on effort.
- *Step 4. Ranking k closest projects:* This aims at retrieving software projects from the dataset that exhibit large similarity with project under investigation. The distance between two tuples at k^{th} feature, is calculated by the formula as shown in Equation (6).

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \zeta \Delta_{\max}}{\Delta_{oj}(k) + \zeta \Delta_{\max}} \quad (6)$$

All Grey Relational Grades between the reference project o and i^{th} comparative project, $\Gamma(x_o, x_i)$ are calculated according to Equation (6). The range of Γ is from 0 to 1 in each case. For more similarity between projects the value of Γ approaches one and for zero it means that two projects are completely dissimilar, The projects that have the higher value on GRG gets the greatest opportunity to contribute in the final estimate.

- *Step 5. Effort Prediction by GRA:* Effort for GRA is the simple aggregation of k most influential projects [25].

$$\hat{\varepsilon} = \sum_{i=1}^k w_i * \varepsilon_i \quad (7)$$

where weight w_i is given by,

$$w_i = \frac{\tau(x_0, x_i)}{\sum_{j=1}^k \tau(x_0, x_j)} \quad (8)$$

ε_i = effort of i^{th} most influential project .

- **Step 6. Effort Prediction by Regression:** In this step, effort estimate for a given project is calculated by applying various regression techniques on only k most similar projects obtained from Step 2.

3.2. Modeling using Fuzzy Grey Relational Analysis for Software Effort Estimation (Fuzzy GRA)

The framework of Fuzzy GRA for individual projects is shown in Figure 2. The steps involved in the process are as explained below:

- **Step 1. Select continuous attributes from the dataset.**
- **Step 2. Data series construction:** [explained in step 2, section 3.1]
- **Step 3. Data Preparation:** Normalization of data and list wise deletion of missing values are performed in the data preparation step.
- **Step 4. Case retrieval:** In this step, those projects are retrieved from the dataset that exhibit large similarity with the reference project. $\Delta_{0,i}(k)$ between two tuples is calculated using the Fuzzy distance, the data is fuzzified using FCM before calculating the distance (explained in Clustering Data). After generating the grey relational coefficient between the tuples, the grey relational grades between the reference project o and i^{th} comparative project are calculated, the $\Gamma(x_0, x_i)$ values are calculated for each i according to Equation (3). The projects are ranked on the basis of their grades and k projects with highest grades are selected. The value of k is different with each reference project.

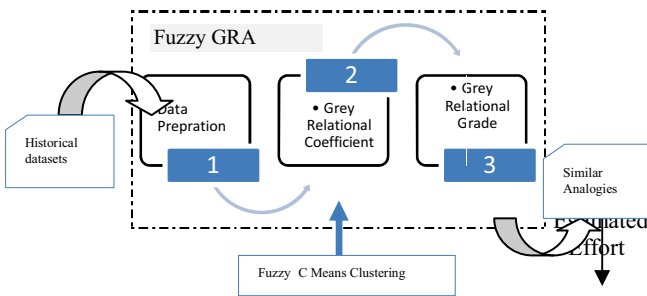


Figure 2. Structural framework for Fuzzy GRA software effort estimation.

3.2.1. Clustering Data

To the data values obtained after normalization, we apply Fuzzy C means clustering, for generating the

$\Delta_{0,i}(k)$ of the grey relational coefficient. Instead of generating the Δ as per equation 2, we make significant modification to $\Delta_{0,i}(k)$ by incorporating the fuzzy distance between two numeric feature values. The purpose behind using fuzzy is that it tries to minimize the uncertainty associated with similarity measurement.

Fuzzy C means algorithm enables to group closest projects together in the same cluster enabling better and efficient case retrieval. It clusters the data set into n clusters with every data point in the dataset belonging to every cluster to a certain degree. *genfis3* a function of MATLAB, generates a FIS using fuzzy c-means (FCM) clustering by extracting a set of rules that models the data behavior. This algorithm is based on the minimizing the objective function that signifies the distance from any given data point to a cluster center weighted by that data point's membership grade.

$$fismat = genfis3(Xin, Xout, type, cluster_n) \quad (9)$$

The function requires separate sets of input and output data as input arguments. It generates a FIS structure and allows specifying the number of clusters (*cluster_n*) to be generated by FCM. The number of clusters determines the number of rules and membership functions in the generated FIS, *cluster_n* is assigned either an integer or 'auto'. In this study *cluster_n* is 'auto', the function uses the subtractive clustering algorithm with a radii of 0.5 and the minimum and maximum values of X_{in} and X_{out} as *xBounds* to find the number of clusters.

The subtractive clustering method assumes each data point as a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. The steps involved are:

- Selects the data point with the highest potential to be the first cluster center
- Removes all data points in the vicinity of the first cluster center (as determined by radii), in order to determine the next data cluster and its center location
- Iterates on this process until all of the data is within radii of a cluster center

4. Experimental methods

4.1. Validation and Evaluation Criteria

The datasets available have less than 100 observations, therefore instead of using hold out or 10 fold validation, Leave one Out Cross Validation (*LOOCV*) has been applied. In *LOOCV*, for each iteration one project is left out once as test data and used entirely to assess the performance of the data set that is trained on the remaining projects. In order to measure the accuracy of the software estimation, we have used four most popularly used evaluation criteria.

Table 2. Evaluation Criteria's.

Performance Measures	Formula
MRE (Magnitude of Relative Error)	$MRE = \frac{ actual_i - estimated_i }{actual_i}$
MMRE (Mean MRE)	$MMRE = \frac{1}{N} \sum_{x=1}^N MRE_x$
MdMRE (Median of MRE's)	$MdMRE = median(MRE_x)$
MER (Magnitude of Relative Error relative to the estimate)	$MER = \frac{ actual_i - estimated_i }{estimated_i}$
MMER (Mean MER)	$MMER = \frac{1}{N} \sum_{x=1}^N MER_x$
Pred(l)	$k/n*100$

The boxplots of the absolute residuals $|actual_i - estimated_i|$ gives good indication of the distribution of residuals and can explain summary statistics such as *MMRE* and *Pred(25)*. The Box plot shows the five number summary i.e the median as the central tendency of the distributions, the Inter quartile range (IQR) and the min –max values. It also shows the outliers of the individual distributions. The length is the spread of the distribution. The box represents 50% of the observations in the distribution. A small box is a peaked distribution, whereas a long box is flattened distribution.

4.2. Data Sources

In order to evaluate the models based upon the proposed methodologies, five well established datasets from the Promise repository [26] have been used for validating our models. The descriptive statistics of the data sets are shown in Table 3. given below. Though these dataset are old, still they are extensively being used to assess the comparative accuracy of the new technique.

Table 3. Descriptive Statistics of the datasets.

	Dataset	Cases	Features	Effort Mean	Effort Standard Deviation
1	Finnish	38	8	7678.29	7135.28(hours)
2	Desharnais	77	9	4834	4188(hours)
3	Cocomo-81	63	17	683.526	1821.51133 (hours)
4	Albrecht	24	8	21875	28417(hours)
5	Kemerer	15	5	219.25	263(man hours)

5. Experimental Results

The experimental results of the proposed models are shown in the table 9.

5.1. Comparison for Finnish dataset

The best results have been achieved with Finnish dataset, with *MMRE* =10.03% and *Pred(25)* = 89.47%.

with *Fair robust estimator* integrated with GRA. Using *FuzzyGRA*, also equally good results were obtained with *MMRE*=21.66 and *Pred(25)* accuracy=81.57%.

5.2. Comparison for Albrecht Dataset

As shown in Table 4. GRACE model that was developed by Song *et al.*[30], obtained *MMRE* =60.25% and *Pred(25)* = 52.63%. In the second study by Song *et al.*[31] obtained *MMRE* =26.1% and *Pred(25)* = 50%. In the study carried out by Shepperd and Schofield [27] between regression and analogy estimation models on Albrecht dataset, regression model resulted in *MMRE*=90% and *Pred*=33%, while analogy obtained *MMRE*=62% and *Pred*=33%. The FGRA model by Azzeh [2] obtained *MMRE* = 51.1% and *Pred(25)* =48.6%. The GREAT_RM model appears significantly better than those of Shepperd results, GRACE results and FGRA results with *MMRE* = 24.16% and *Pred(25)* = 70.83%. Best results were obtained using *FuzzyGRA* demonstrating that fuzzy performs the best while handling uncertainties.

Table 4. Comparison over Albrecht Dataset.

ALBRECHT				
	MMRE	Md MRE	Pred(25)	MMER
Fuzzy GRA (individual projects)	45.01	3.96	75.00	16.33
GREAT_RM	24.16	10	70.83	22.57
GRACE+	26.1	24.2	50	-
FGRA	51.1	48	28.6	60.4
GRACE	60.25	21.35	52.63	-

5.2. Comparison for COCOMO-81 Dataset

The *GREAT_RM* results have outperformed the other techniques in case of COCOMO-81 dataset. Song *et al.* [30] applied the GRACE model to the COCOMO dataset and they obtained *MMRE*=76.09%, and *Pred(25)* = 20.63%. Another study by Song *et al.*[31] obtained *MMRE* =49.8% and *Pred(25)* = 29%.FGRA [2] obtained *MMRE*=23.2% and *Pred(25)* = 66.7% . *GREAT_RM* results obtained *MMRE*=21.04% and *Pred(25)* = 76.19% which is an improvement over GRACE FGRA and GRACE⁺.

Table 5. Comparison over cocomo_81 dataset.

Cocomo 81				
	MMRE	Md MRE	Pred(25)	MMER
GREAT_RM	21.04	9.42	76.19	48.71
GRACE+	49.8	55.2	29	-
FGRA	23.2	14.8	66.7	25.6
GRACE	76.09	60.52	20.63	-

5.3. Comparison for Desharnais Dataset

The Desharnais dataset has been widely used to test software estimation models. Shepperd and Schofield [27] employed analogy estimation on 77 available projects after removing four projects that have missing

values using Angel tool. They obtained MMRE=64% and Pred(25) = 36%. In similar way, Song *et al.* [30] developed GRACE software estimation model based on grey relational analysis, they obtained MMRE=49.83%, Pred(25) = 30%. FGRA obtained MMRE=30.6% and Pred(25) = 64.7% . Another study by Song *et al.* [31], GRACE⁺ obtained MMRE=41.4%, Pred(25) = 45.3%. Thus *GREAT_RM* achieves the best MMRE = 19.48 % and Pred(25) = 89.61%. Equally good results were obtained using *FuzzyGRA* with MMRE=24.6% and Pred(25)=71.42%

Table 6. Comparison over Desharnais.

Desharnais				
	MMRE	MdMRE	Pred(25)	MMER
Fuzzy GRA (individual projects)	24.69	7.40	71.42	23.59
GREAT_RM	16.78	7.84	74.02	31.63
GRACE+	41.4	29.2	45.3	-
FGRA	30.6	17.5	64.7	34.4
GRACE	49.83	33.93	30	-

5.3.1. Comparison for Kemerer Dataset

For the Kemerer dataset, the fair robust estimator gives better results when used with GRA. *FuzzyGRA* also produced reasonable results with MMRE =29.63% and Pred(25) = 60.00%. GRACE obtained MMRE =58.83% and Pred(25) = 26.67%. Another study GRACE⁺ by Song *et al.*(2011) obtained MMRE =19.6% and Pred(25) = 78.6%. The FGRA model by Azzeh [2]obtained MMRE = 36.2% and Pred(25) = 52.9%. The *GREAT_RM* model appears significantly better than those to GRACE results and FGRA results, however less significant to GRACE⁺.

Table 7. Comparison over kemerer dataset.

Kemerer				
	MMRE	Md MRE	Pred(25)	MMER
Fuzzy GRA (individual projects)	41.19	8.58	66.67	33.21
GREAT_RM	29.63	16.56	60	27.93
GRACE+	19.6	13.8	78.6	-
FGRA	36.2	33.2	52.9	34.3
GRACE	58.83	46.94	26.67	-

The boxplots of the residuals obtained suggest that:

- The medians are very close to zero, signifying that the estimates were biased towards the minimum value where they have tighter spread.
- The median and range of absolute residuals of both methods are small, that shows that at least half of the predictions are accurate.
- The boxes of GRA, GRA+OLS, GRA+RR and Fuzzy GRA overlays the lower tail for all datasets which also presents accurate predictions.
- The results of GRA+SWR were not very accurate, this can be observed from the boxplots also.

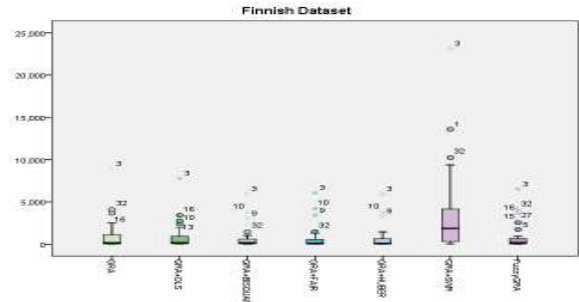


Figure 3. Box plot of absolute residuals for finnish dataset.

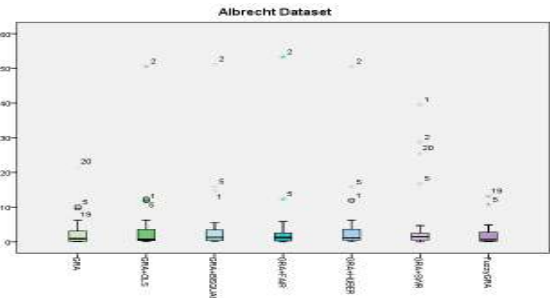


Figure 4. Box plot of absolute residuals for albrecht dataset.

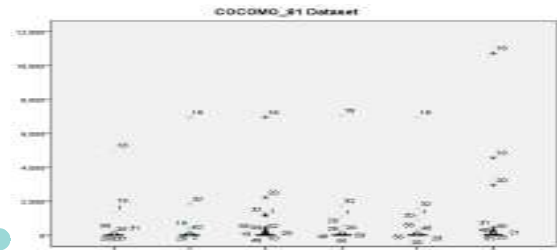


Figure 5. Box plot of absolute residuals for cocomo-81 dataset.

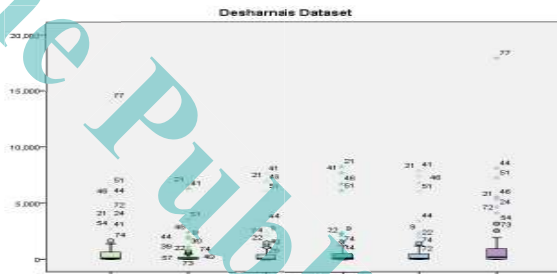


Figure 6. Box plot of absolute residuals for desharnais dataset.

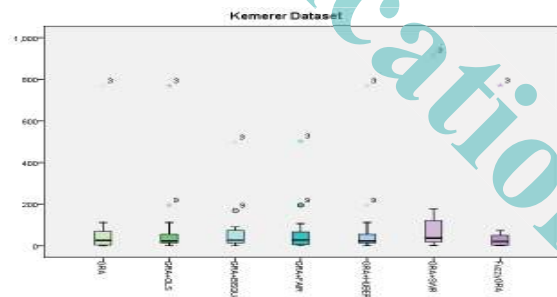


Figure7. Box plot of absolute residuals for kemerer dataset.

The results achieved by *GREAT_RM* and *FuzzyGRA* are subjected to statistical testing using Wilcoxon signed rank test by setting the test value to 0 (shown in Table 8.). The level of significance is taken to be 0.05. If the resulting p-value is less than 0.05, then a statistically significant difference exists between the

sample median and test value but in case it is greater than 0.05 than it can accepted that no statistical difference exists between sample median and test value.

- Most of the results of the residuals obtained in most of the cases were not significantly different from the test value and hence the null hypotheses were

accepted in the case of GRA, GRA with OLS, GRA with RR and Fuzzy GRA for all datasets.

- The results obtained with integration of GRA and SWR for Desharnais, Kemerer and COCOMO-81 dataset were however significantly different and cannot be accepted.

Table 8. Wilcoxon Signed Rank test.

	GRA - Actual	(GRA+OLS) - Actual	(GRA+Bisquare) - Actual	(GRA+Fair) - Actual	(GRA+Huber) - Actual	(GRA+SWR) - Actual	FuzzyGRA - Actual
Z (Finnish)	-.138 ^a	-.964 ^a	-.457 ^a	-.732 ^a	-.703 ^a	-1.922 ^b	.000 ^c
Asymp. Sig. (2-tailed)	0.89	0.335	.648	.464	.482	0.055	1.000
Z (Albrecht)	-.729 ^a	-.171 ^a	-.457 ^b	-.629 ^a	-.514 ^b	-.743 ^a	-.800 ^a
Asymp. Sig. (2-tailed)	0.466	0.864	.648	.530	0.607	0.458	.424
Z (COCOMO 81)	-.698 ^a	-1.595 ^a	-1.171 ^a	-1.458 ^a	-1.184 ^b	-2.615 ^b	
Asymp. Sig. (2-tailed)	0.485	0.111	.242	.145	.236	.009	
Z (Desharnais)	-1.234 ^a	-.784 ^a	-.724 ^b	-.145 ^b	-.307 ^b	-4.359 ^b	-.129 ^b
Asymp. Sig. (2-tailed)	0.217	0.433	.469	.885	.759	0	.897
Z (Kemerer)	-.369 ^a	-.568 ^b	.000 ^a	.000 ^c	-.568 ^b	-2.385 ^b	-.483 ^a
Asymp. Sig. (2-tailed)	0.712	0.57	1	1.000	.570	0.017	.629

a. Based on negative ranks. b. Based on positive ranks. c. Wilcoxon Signed Ranks Test

Table 9. Outcomes of the two methodologies (GRA, GRA+OLS, GRA+ Bisquare, GRA+ Fair, GRA+ Huber, GRA+SWR, Fuzzy GRA).

DATA SET	GRA	GRA + OLS	GRA + Bisquare	GRA+ Fair	GRA+ Huber	GRA + Stepwise	Fuzzy GRA
Finnish							
MMRE	11.37	11.88	11.41	10.3	10.76	58.61	21.66
Median (MRE)	2.67	2.19	2.3	1.93	2.07	27.93	3.30
MMER	12	12.88	7.81	9.81	10.52	40.99	11.59
Pred(25)	76.32	89.47	89.47	89.47	92.1	47.37	81.57
Albrecht							
MMRE	46.35	29.83	27.09	26.15	24.16	32.64	45.01
Median (MRE)	4.89	7.47	11.31	7.72	10	12.39	3.96
MMER	17.15	21.72	24.43	21.62	22.57	24.87	16.33
Pred(25)	70.83	70.83	70.83	66.67	70.83	70.83	75.00
COCOMO -81							
MMRE	30	32.35	37.37	24.66	28.52	21.04	--
Median (MRE)	7.07	5.32	5.59	5.45	3.86	9.42	--
MMER	26.86	15.71	19.91	19.72	21.42	48.71	--
Pred(25)	68.25	76.19	71.43	73.02	74.6	76.19	--
Desharnais							
MMRE	34.9	18.19	25.44	28.07	28.21	16.78	24.69
Median (MRE)	5.07	1.51	8.36	8.44	7.31	7.84	7.40
MMER	22.12	8.74	18.97	19.42	19.23	31.63	23.59
Pred(25)	68.83	90.9	79.22	75.32	79.22	74.02	71.42
Kemerer							
MMRE	46.67	35.18	32.68	29.63	35.12	38.65	41.19
Median (MRE)	11.46	21.92	28.37	16.56	29.1	31.35	8.58
MMER	35.58	41.04	29.88	27.93	30.61	78.8	33.21
Pred(25)	60	53.33	46.67	60	46.67	40	66.67

6. Conclusions and Future Scope

The empirical evaluations have revealed that the GREAT_RM and Fuzzy GRA techniques can certainly improve the estimation process. The models can be used for early stage estimation where the data is uncertain. The methodology presents a significant improvement over GRACE [30] and FGRA [2]. The results obtained are superior over our previous results wherein the value of k was fixed for each reference project using GREAT_RM [25].

Thus, the results obtained are inspiring and urge us to endeavour different methodologies for producing enhanced estimates. The proposed methodologies can

further be applied on some other large datasets with different validation criteria's and also feature selection methodology can be applied in both methods to enhance prediction. Further, GREAT_RM methodology can be carried out with other Robust Regression techniques like S-Estimators, Least Trimmed Squares or MM Estimations etc.

Acknowledgment

The authors would like to thank Promise Repository for providing such valuable datasets for carrying out research work.

References

- [1] Andrews D., Bickel P., Hampel F., Huber P., Rogers W., and Tukey J., *Robust Estimates of Location: Survey and Advances*, vol. 173, Princeton, NJ: Princeton University Press, 1972.
- [2] Azzeh M., Neagu D. and Cowling P., "Fuzzy Grey Relational Analysis for Software Effort Estimation", *Journal of Empirical software Engineering*, vol. 15, no. 1, 2010. [doi:10.1007/s10664-009-9113-0]
- [3] Azzeh M., Neagu D. and Cowling P., "Analogy-based Software Effort Estimation using Fuzzy Numbers", *Journal of Systems and Software*, vol. 84, no.2, pp:270-284, 2011. [doi: 10.1016/j.jss.2010.09.028]
- [4] Boehm B. and Valerdi R., "Achievements and Challenges in Software Resources Estimation", CSSE Tech Report, 2005 [doi:10.1145/1414004.1414006]
- [5] Deng J., "Control problems of Grey System". *System and Control Letters*, vol.1, pp:288–94, 1982.
- [6] Deng J., "Introduction to Grey System theory", *The Journal of Grey System*, vol.1, pp:1-24, 1989^a
- [7] Deng J., "Grey information space", *The Journal of Grey System*, vol.1, no.1, pp: 103–117, 1989^b
- [8] Hsia K. and Wu J., "A Study on the Data Preprocessing in Grey Relation Analysis". *Journal of Chinese Grey System Association*, vol.1, no.1, pp:47-53, 1998
- [9] Hsu C. and Huang C., "Improving Effort Estimation Accuracy by Weighted Grey relational Analysis During Software development". 14th Asia-Pacific Software Engineering Conference, pp:534-541, 2007
- [10] Huang S., Chiu N. and Chen L., "Integration of the Grey Relational Analysis with Genetic Algorithm for Software Effort Estimation," *European Journal of Operational and Research* vol. 188, pp. 898-909, 2007. [doi:10.1145/1540438.1540440]
- [11] Huang S. and Huang C., "Control of an Inverted Pendulum using Grey Prediction Model". *IEEE Transactions on Industry Applications*, vol.36, no.2, pp:452–458, 2000.
- [12] Huber P., "Robust Estimation of a Location Parameter". *Annals of Mathematical Statistics*, vol.35, pp:73–101, 1964
- [13] Huber P., "Robust regression: Asymptotics, Conjectures and Monte Carlo", *The Annals of Statistics*, vol.1, pp: 799–821, 1981
- [14] Jiang B., Tasi S. and Wang C., "Machine Vision-based Gray Relational Theory applied to IC marking inspection". *IEEE Transactions on Semiconductor Manufacturing*, vol.15, no.4, pp:531–539, 2002.
- [15] Jou J., Chen P. and Sun J., "The Gray Prediction Search Algorithm for Block Motion Estimation". *IEEE Transactions on Circuits and Systems for Video Technology*, vol.9, no.6, pp:843–848, 1999.
- [16] Jorgenson J. and Shepperd M., "A Systematic Review of Software Development Cost Estimation Studies". *IEEE Transactions on Software Engineering*, vol.33, no.1, pp:33-53, 2007 [doi:10.1109/TSE.2007.3]
- [17] Li G., Ruhe J., Emran A. Al-and Richter M., "A Flexible method for software Effort Estimation by Analogy", *Empirical Software Engineering*, vol.12, pp:65–106, 2007 [doi:10.1007/s10664-006-7552-4]
- [18] Luo R., Chen T., and Su K., "Target Tracking using a Hierarchical Grey-Fuzzy motion Decision making method", *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol.31, no.3, pp:179–186, 2001.
- [19] MATLAB® Documentation, <http://www.mathworks.com/help/techdoc/>
- [20] Mbamalu G. and Hawary M., "Load Forecasting via Suboptimal Seasonal Autoregressive models and Iteratively Reweighted Least Squares Estimation" *IEEE Transactions on Power Systems*, vol.8, no.1, pp:343-347, 1993
- [21] Mittas N. and Angelis L. "Comparing Cost Prediction models by Resampling Techniques", *Journal of Systems and Software*, vol.81, pp:616–632, 2008 [doi:10.1016/j.infsof.2010.07.006]
- [22] Mittas N., Athanasiades M. and Angelis L. "Improving Analogy-based Software Cost Estimation by a Resampling Method," *Journal of Information and software technology*, vol. 50 , no. 3 , pp. 221-230, 2008. [doi:10.1016/j.infsof.2007.04.001]
- [23] Mittas, N. and Angelis, L., "LSEbA: Least Squares Regression and Estimation by Analogy in a Semi-Parametric Model for Software Cost Estimation". *Empirical Software Engineering*, vol.15, no.5, pp :523-555, 2010 [doi:10.1007/s10664-010-9128-6]
- [24] Mukhopadhyay T., Vicinanza S. and Prietula M., *Examining the Feasibility of a Case-based reasoning model for Software Effort Estimation*, MIS Quarterly, vol. 16, no. 2, pp:155–171, 1992
- [25] Nagpal G., Uddin M. and Kaur A., "A Hybrid Technique using Grey Relational Analysis and Regression for Software Effort Estimation using Feature Selection", *International Journal of Soft Computing and Engineering (IJSCE)*, vol.1, no.6, 2012
- [26] PROMISE Repository of Empirical Software Engineering data. <http://promisedata.org/repository>.
- [27] Shepperd M., Schofield C. and Kitchenham B., "Effort Estimation using Analogy", *In*

Proceedings of the 18th International Conference on Software Engineering, pp:170-178,1996.

- [28] Shepperd M. and Schofield C., "Estimating Software Project Effort Using Analogies", *IEEE Transaction on Software Engineering* vol.23,pp:736-743,1997
- [29] Song Q., Shepperd M., Mair C., "Using Grey Relational Analysis to Predict Software Effort with Small Data Sets", *Proceedings of the 11th International Symposium on Software Metrics (METRICS'05)*, pp.35-45,2005
- [30] Song Q. and Shepperd M., "Predicting software project effort: A grey relational analysis based method", *Expert Syst. Appl.* vol.38,no.6,pp: 7302-7316, 2011 [doi:10.1016/j.eswa.2010.12.005]
- [31] Su S., Su Y. and Huang J., "Grey-based power control for DS-CDMA cellular mobile systems". *IEEE Transactions on Vehicular Technology*, vol.49, no.6, pp2081-2088, 2000.
- [32] Verardi V. and Croux C., "Robust regression in Stata", *Stata Journal, StataCorp LP*, vol.9,no.3,pp: 439-453,2009
- [33] Wang Y., "On-demand forecasting of stock prices using a real-time predictor," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp.1033-1037, 2003.
- [34] Wu J., You M. and Wen K., "A Modified grey relational analysis," *The Journal of Grey System*, vol.11, no. 3, pp. 287-292, 1999.
- [35] Wu J. and Chen C., "An alternative form for grey relational grades," *The Journal of Grey System*, vol. 11, no. 1, pp.7-12, 1999.
- [36] Yan L., "Modeling Fuzzy Data with Fuzzy Data Types in Fuzzy Database and XML Models," vol.10, no. 6, pp.4847-6, 2011.
- [37] Zadeh L., "Toward a Theory of Fuzzy Information Granulation and its centrality in Human Reasoning and Fuzzy logic," *Journal Fuzzy Sets and Systems*, vol. 90, pp.111-127, 1997.



Geeta Nagpal pursuing Ph D programme in the Department of Computer Science and Engineering, National Institute of Technology, Jalandhar, INDIA. She completed her Master's degree in Computer Science from Punjab Agricultural University, Ludhiana. She is presently working as Associate Professor in the Department of Computer Science and Engineering at NIT, Jalandhar. Her research interests are Software Engineering, Databases and Data mining.



Moin Uddin Pro Vice Chancellor, DTU, Delhi, India. He obtained his B.Sc. Engineering and M.Sc. Engineering (Electrical) from AMU, Aligarh in 1972 and 1978 respectively. He obtained his Ph. D degree from University of Roorkee, Roorkee in 1994. Before joining as the Pro Vice Chancellor of Delhi Technological University, he was the Director of NIT, Jalandhar for five years. He has worked as Head Electrical Engineering Department and Dean Faculty of Engineering and Technology at Jamia Millia Islamia (Central University) New Delhi. He supervised 14 Ph. D thesis and more than 30 M.Tech dissertations.



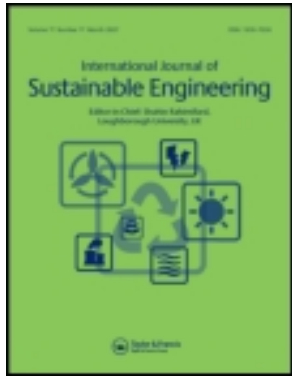
Arvinder Kaur Associate Professor, University School of IT, GGSIPU, Delhi, India. She completed her Master's degree in Computer Science from Thapar Institute of Engineering and Technology and Ph D from Guru Gobind Singh Indraprastha University, Delhi. Prior to joining the school, she worked with Dr. B. R. Ambedkar Regional Engineering College, Jalandhar and Thapar Institute of Engineering and Technology. Her research interests include Software Engineering, Object-Oriented Software Engineering, Software Metrics, Microprocessors, Operating Systems, Artificial Intelligence, and Computer networks. She is a lifetime member of ISTE and CSI. Her paper titled, "Analysis of object oriented Metrics" was published as a chapter in the book *Innovations in Software Measurement* (Shaker-Verlag, Aachen 2005).

This article was downloaded by: [Delhi Technological University]

On: 03 November 2013, At: 21:07

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Sustainable Engineering

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tsue20>

IJSE reviewers September 2013 reviewer acknowledgement

Published online: 29 Oct 2013.

To cite this article: (2013) IJSE reviewers September 2013 reviewer acknowledgement, International Journal of Sustainable Engineering, 6:4, 372-374, DOI: [10.1080/19397038.2013.848685](https://doi.org/10.1080/19397038.2013.848685)

To link to this article: <http://dx.doi.org/10.1080/19397038.2013.848685>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

IJSE reviewers September 2013 reviewer acknowledgement

The Editors are indebted to the following for their valued assistance as referees:

Name	Institution
Narasimha Rao A V	NIT Warangal
Ali Abdulkhani	University of Tehran
Nazirah Abidin	Science University of Malaysia
Saroj Acharya	Iter, Soa University
Ali Albahi	King Abdulaziz University
Jalal Ashayeri	Tilburg University
Taha Ashour	Geotechnical Institute
Jan Baeyens	University of Warwick
Baiju baiju velappan	NIT Calicut
Nagaraj Banapurmath	B.V.B. College of Engineering and Technology
Rangan Banerjee	Indian Institute of Technology Bombay
Theresa Barker	University of Washington
A Baskaran	
Soumitra Basu	Stevens Institute of Technology
Gero Benckiser	Justus Liebig Univeriry
Mohamed Benghanem	Faculty of Sciences, taibah University
Nikolaos Bilalis	Technical University of Crete
Jacqueline Bloemhof	Wageningen University
Andrea Bonfiglio	University of Genoa
E Bottani	University of Parma
Wojciech Budzianowski	Wrocław University of Technology
Justina Catarino	INETI
Allen Clegg	Loughborough University
Alan Davies	Cardiff University
Jorge de Brito	Technical University of Lisbon
Rommert Dekker	
S R Devadasan	PSG College of Technology
Raf Dewil	Katholieke Universiteit Leuven
H. Dharmadhikari	Maharashtra Institute of Technology
George Dodos	National Technical University of Athens
Tom Dyer	University of Dundee
Orady Elsayed	University of Michigan
Napoleon Enteria	Tohoku University
Tulay Esin	Gebze Institute of Technology
Xianfeng Fan	University of Edinburgh
Adel Francis	Cmrdi
Francis Francis	Notre Dame University-Louaize
Antonio Frattari	University of Trento
Jacqueline Glass	Loughborough University
maria grazia gnoni	University of salento
BingXin Gong	Not affiliated with any institute
Kasthurirangan Gopalakrishnan	Iowa State University
Alfred Gordon	National Institute of Technology, Surathkal
Corey Griffin	Portland State University
Patricia Guarnieri	UFPE
Tuncay Gunhan	Ege University
Anand Gurumurthy	Indian Institute of Management, Kozhikode, India
F Halicioglu	Dokuz Eylul University
Sophie Hallstedt	Blekinge Inst of Technology
Can Hasimoglu	

Hanbey Hazar	Firat University
Ehsan Houshfar	Norwegian University of Science and Technology (NTNU)
M Iqbal	Memorial University of Newfoundland
T Ishiyama	Kyoto University
Ibrahim Iskin	Portland State University
Mallikarjuna J	IIT, Madras
Shashi Jain	Technocrats Institute of Technology & Science
Siddarth Jain	NUS
Patrick James	University of Southampton
Matty Janssen	École Polytechnique de Montréal
Zhigang Jiang	Wuhan University of Science and Technology
Anil Jindal	Birla Institute of Technology and Science
S Jindal	College of Technology and Engineering
Vilas Kalamkar	
Stamatios Kalligeros	Hellenic Naval Academy
Hakan Karakaya	Batman University
Rahat Khan	UPES Dehradun UK India
Kamal Khatri	Mody Institute of technology and Science, lakshmangarh
Ahmet Kolip	University of Sakarya
A Kumar	Dhanalaksmi College of
Naveen Kumar	Delhi Technological University
Ola Lædre	NTNU
Aisyah Larasati	
Barbara Linke	University of California, Davis
Mr Todd Litman	Victoria Transport Policy Institute
Win Jet Luo	National CHIN-YI University of Technology, Taiwan
Fanhua Ma	Tsinghua University
Arif Malik	SQU
Rajib Mallick	WPI
Ezio Mancaruso	Istituto Motori - CNR
Abedalraouf Mayyas	Arizona State University
S Moura	University of California, Berkeley
Ramesh Babu Nallamothu	Adama University
Paolo Nava	Secondo Mona
Hoon Ng	The University of Nottingham Malaysia Campus
Soo-Young No	Chungbuk National University
Sang Noh	Sungkyunkwan University
H Ong	University of Malaysia
Emeka Osaji	University of Nottingham
Michael Overcash	Wichita State University
Mumin Ozcan	Gediz University
Ilhan Ozturk	Kocaeli University
Nugroho Pambudi	Semarang State University
Hitesh Panchal	GTU
Siba Panigrahi	C.V. Raman College of Engineering
N Panwar	College of Technology & Engineering
Byung-Sik Park	Korean Institute Energy Research
Tony Parry	University of Nottingham
Pablo Pascual-Muñoz	University of Cantabria
Alexandre Pelegrini	Universidade Federal do Paraná - UFPR
Rosca Radu	University of Agricultural Sciences
Mujib Rahman	Nottingham Trent University
K Ramamurthy	IIT Madras
RM Chandima Ratnayake	University of Stavanger
Matthieu Rauch	Ircyn UMR CNRS 6597 - Ecole Centrale Nantes
Krishna Reddi	Syracuse University
I Regupathi	
Anthony Reynolds	Dublin Institute of Technology
W Rivera	Universidad Nacional Autonoma de Mexico
Sameh Saad	Sheffield Hallam University
Atul Sagade	New Satara college of Engineering & Management
M Sakei	Shibaura Institute of Technology
A Samuila	

Dileep Satappa More	Indian Institute of Management Culcutta
Supakit Sayasoonthorn	Kasetert University, Bangkok, Thailand
S Shaahid	King Fahd University of Petroleum and Minerals
Guodong Shao	National Institute of Standards and Technology
Brajendra Sharma	University of Illinois-Urbana Champaign
Nitin Shrivastava	University Institute of Technology, Rajiv Gandhi Proudhyogiki Vishwavidyalaya
Chandrakasan Solaimuthu	Paavi Engineering College
Erik Sundin	Linköping University
guluwadi Suresh	Nehru College of Engineering and Research Centre
Mochammad Syamsiro	Tokyo Institute of Technology
Zsuzsa Szalay	Budapest University of Technology and Economics
Fernando Torgal	University of Minho
Deb Tripura	National Institute of Technology Agartala (India)
A Tsolakis	University of Birmingham
Yasushi Umeda	Osaka University
S Vinodh	National Institute of Technology
Lihui Wang	University of Skövde
Renee Wever	Delft University of Technology
Ke Xing	University of South Australia
Yangang Xing	Welsh School of Architecture, Cardiff University, Wales, UK
Qingang Xiong	Iowa State University
Avadhesh Yadav	NIT Kurukshetra
V S Yaliwal	S.D.M. College of Engineering and Technology
Juan Yi-Kai	National Taiwan University of Science & Technology
Chris Yuan	University of Wisconsin, Milwaukee
Hong-Chao Zhang	Texas Tech University

Research Article

Magnetotransport Behaviour of Nanocrystalline $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ ($0.40 \leq x \leq 0.60$)

Neelam Maikhuri,^{1,2} Anurag Gaur,¹ Vasudha Aggarwal,³ Umesh Gaur,⁴ and H. K. Singh³

¹ Department of Physics, National Institute of Technology, Kurukshetra 136119, India

² Department of Applied Physics, Delhi Technological University, Delhi 110042, India

³ National Physical Laboratory, Dr. K.S. Krishnan Marg, New Delhi 110012, India

⁴ Centre of Nanotechnology, Indian Institute of Technology, Roorkee 247667, India

Correspondence should be addressed to Anurag Gaur; anuragdph@gmail.com

Received 16 July 2013; Accepted 1 September 2013

Academic Editors: D. M. Chipara and C. Carbonaro

Copyright © 2013 Neelam Maikhuri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The nanocrystalline samples of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ (PSMO) ($x = 0.40, 0.50, 0.55$, and 0.60) were synthesized by wet-chemical sol-gel route. Structural, magnetic, and magnetotransport properties have been studied systematically. It is found that $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ samples with Sr content $x = 0.40$ and 0.50 show paramagnetic to ferromagnetic (PM-FM) transition at $T_C \sim 308$ K with no trace of FM-AFM transition within the temperature range of 77 – 350 K. However, interestingly a second transition is observed at $T \sim 273$ and 255 K, respectively, for samples $x = 0.55$ and 0.60 correspond to an A-AFM magnetic structure. This indicates that samples $x = 0.40$ and 0.50 are ferromagnetic below T_C , while other samples ($x = 0.55$ and 0.60) have a mixed phase consisting of FM and A-type AFM phases. Resistivity versus temperature (ρ - T) curve shows that the resistivity of all the samples is much larger than the single crystals of corresponding compositions due to large contribution of grain boundaries in the present nanocrystalline samples. Moreover, the decrease in metallic component at higher Sr concentration is also evidenced by the successive reduction in magnetoresistance (MR) with increasing Sr content from $x = 0.40$ to 0.60 .

1. Introduction

Several experimental and theoretical studies have focused on the exploration of grain size effect on the structural, magnetic, and electrical transport properties of alkaline-earth doped rare earth perovskite manganites chemically represented by $\text{RE}_{1-x}\text{AE}_x\text{MnO}_3$ (RE = rare earth cation and AE = alkaline earth cation) because of their unusual magnetic and electronic properties like colossal magnetoresistance (CMR), charge ordering, orbital ordering, and phase separation [1–10]. These studies focus on and clearly highlight the significance of broken Mn–O–Mn exchange bonds at the grain surface and their likely impact on the magnetic and electrical transport properties. Since size reduction leads to increased contribution from the surface regions, the broken Mn–O–Mn bonds are indeed expected to have a definite impact on magnetoelectrical properties in manganites. However, in view of the fact that manganites exhibit a strong competition and

correlation between various structural and electronic degrees of freedom even more intriguing and complex phenomena are expected. It has been shown that in nanomanganites, size reduction below ~ 100 nm renders the charge and orbitally ordered (CO–OO) ground state with unstable antiferromagnetic (AFM) spin order, giving rise to a ferromagnetic (FM) ground state [3, 11]. Size induced transition from the AFM/CO to the weak ferromagnetic (WFM) state was observed in both nanowires [3] and nanoparticles [4]. It has been shown by Lu et al. [11] that destabilization of the AFM–CO state and formation of an FM order can result in an enhancement of magnetization by two orders of magnitude. The WFM ground state in nanomanganites resulting from the destabilization of AFM ground state has been regarded as a direct consequence of size reduction because when the size is small enough (e.g., 20 nm), the effect of surface spin disordering would become more evident. However, WFM induced by the destabilization of the AFM order has been also

reported in single crystals [12, 13] as well as epitaxial thin films [14]. This suggests that the evolution of WFM out of the AFM-CO state cannot be a consequence of material downsizing to nanometric scale alone and that some additional effects such as orbital disordering may also be equally important [13–17].

The effect of material downsizing can have more dramatic effect in the vicinity of compositions that possess strong magnetic phase coexistence and hence show bicritical/multicritical points. Among the manganites, $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ (PSMO) has larger bandwidth than $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$ (NSMO) and exhibits PM-FM at $T_C \sim 320$ K for $x \sim 0.30$ – 0.40 [15, 17, 18]. In single crystals, epitaxial thin films, and large grain polycrystalline bulk, insulator to metal (I-M) transition is also observed to be simultaneous with the PM-FM transition. At higher Sr concentrations ($x > 0.50$), it transforms into an A-type AFM metal and when x exceeds 0.75 the magnetic structure becomes C-type AFM insulator. Therefore, it is interesting to study the behavior of this compound in this critical range, $x = 0.40$ to 0.60 , for its nanocrystalline particles.

In view of the above, we have tried to understand the magnetic and magnetotransport behaviour of nanocrystalline $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ in this critical range, $x = 0.40$ to 0.60 . The $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x = 0.40, 0.50, 0.55$, and 0.60) nanoparticles are prepared by wet-chemical sol-gel route and their magnetic and magnetotransport properties for their nanometric size (~ 40 nm) grains are studied.

2. Experimental Procedure

The wet-chemical sol-gel route has been adopted to synthesize $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ (with $x = 0.40, 0.50, 0.55$, and 0.60) nanosized particles at a significantly lower sintering temperature as compared to conventional solid-state reaction method. In this technique, the aqueous solution of high purity $\text{Pr}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$, $\text{Sr}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, and $\text{Mn}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ has been taken in the desired stoichiometric proportions. An equal amount of ethylene glycol has been added to this solution with continuous stirring. This solution is then heated on a hot plate at a temperature of ~ 100 – 140°C till a dry thick brown colour sol is formed. At this temperature ethylene glycol polymerizes into polyethylene glycol, which disperses the cations homogeneously forming a cation polymer network. The polymerized ethylene glycol assists in forming a close network of cations from the precursor solution and helps the reaction in enabling the phase formation at low temperatures as compared to that in bulk synthesis via solid state route. The gel forms a resin and the high viscosity of the resin prevents different cations from segregating and ensures a high level of homogeneity. This has been further decomposed in an oven at a temperature of $\sim 250^\circ\text{C}$ to get a polymer precursor in the form of a black resin-like material. This material was then finely ground into powder. The same procedure was used to synthesize all the samples and then pellets were made for each sample. A pellet from each sample was finally sintered at $\sim 900^\circ\text{C}$ for about 12 hrs. All the synthesized samples have been subjected to phase identification and structural characterizations using

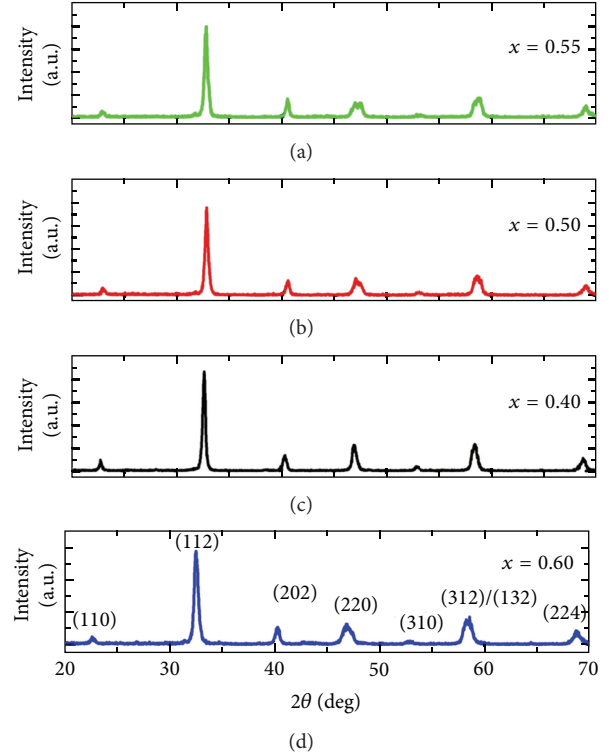


FIGURE 1: XRD patterns of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ samples with $x = 0.40, 0.50, 0.55$, and 0.60 .

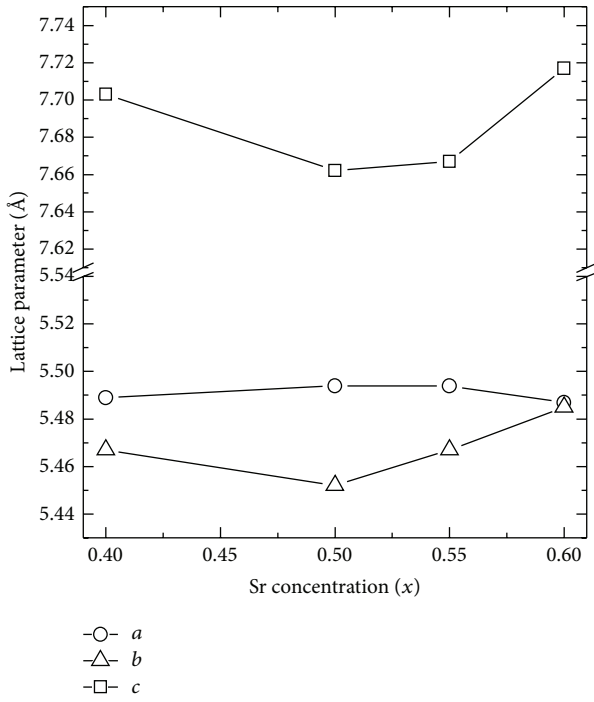
a powder X-ray diffractometer [XRD, Philips PW1710] using $\text{CuK}\alpha$ radiation at room temperature and microstructural characterization by the scanning electron microscopic technique [SEM, Philips XL20]. The magnetotransport measurements have been performed by standard dc four-probe technique in the temperature range of 300 – 77 K at an applied magnetic field of 3 kG. The magnetic characterizations have been carried out by temperature dependent ac susceptibility measurements in the temperature range of 77 – 350 K.

3. Results and Discussion

3.1. Structural and Microstructural Characterization. The powder X-ray diffraction (XRD) patterns of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ (PSMO) ($x = 0.40, 0.50, 0.55$, and 0.60) samples are shown in Figure 1. All the samples are single phase polycrystalline and possess orthorhombic structure (space group Pbnm). The degree of crystallinity remains almost unaffected by the value of x (Sr content). In all the samples, the most intense diffraction maxima corresponds to the (112) plane followed by (312)/(132), (220), and so forth. The lattice parameters of all the PSMO samples were evaluated from the XRD data and are listed in Table 1. For clarity, the variation of the lattice parameters and unit cell volume with Sr content (x) is plotted in Figures 2 and 3, respectively. From the XRD data of Table 1 and Figure 2, it is clear that as the Sr content increases, the in-plane lattice parameters, namely, a and b , come closer, and the structure becomes nearly tetragonal. The out-of-plane lattice parameter c first decreases and then again

TABLE 1: Variation of CSs, lattice parameters, and unit cell volumes of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ samples with different Sr concentrations.

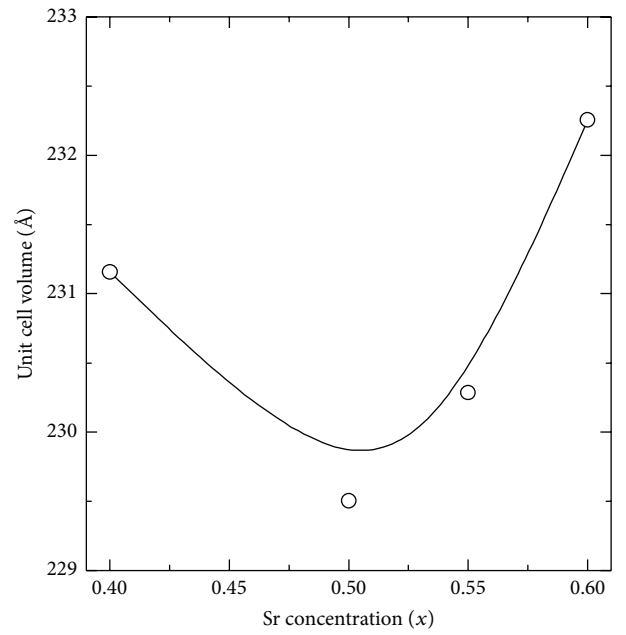
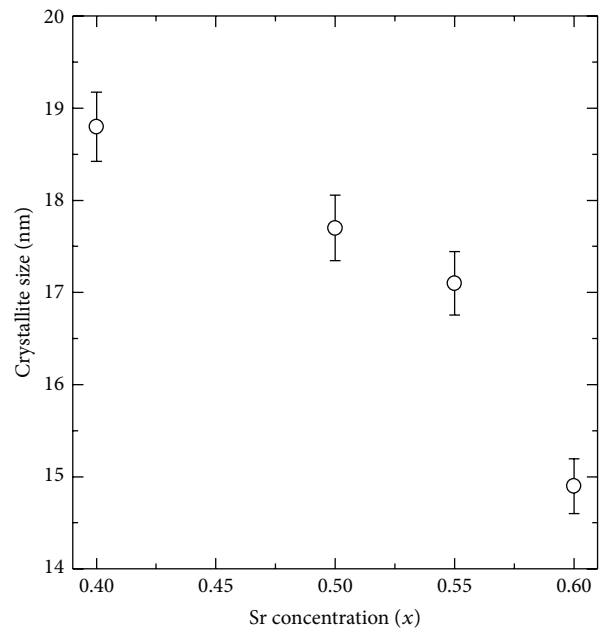
Sr concentration (x)	a (Å)	b (Å)	c (Å)	Volume of unit cell (Å) ³	Crystalline size (CS) (nm)
0.40	5.489	5.467	7.703	229.293	19
0.50	5.494	5.452	7.662	229.502	18
0.55	5.494	5.467	7.667	230.283	17
0.60	5.487	5.485	7.672	230.898	15

FIGURE 2: Variation of lattice parameters of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ with different Sr concentration (x).

increases slightly. This variation in the lattice parameters is in agreement with the previously reported results [15, 17]. However, in the present case, the in-plane lattice parameters are slightly larger than the previously reported values. This could be due to the nanocrystalline nature of the present samples where strain is expected to be more dominant than in microcrystalline samples.

The average crystallite size (CS) of the samples is obtained by the X-ray line width using Scherer formula, $\text{CS} = k\lambda/\beta \cos \theta$, where $k = 0.89$ is the shape factor, λ is the wavelength of X-rays, β is the actual FWHM due to CS only, and θ is the angle of diffraction. The average CS of samples having Sr content $x = 0.40, 0.50, 0.55$, and 0.60 is found to be approximately equal to ~ 19 nm, 18 nm, 17 nm, and 15 nm, respectively. This suggests that Sr concentration (x) also affects the crystallite size. The variation of CS is plotted in Figure 4. The CSs, lattice parameters, and unit cell volumes obtained for the different samples are listed in Table 1.

The surface microstructure of the samples, as revealed by scanning electron microscopy (SEM), was found to consist of nanometric grains. The average grain size is found to be ~ 40 nm in all the samples. This grain size is larger than

FIGURE 3: Variation of unit cell volume of PSMO samples with different Sr concentration (x).FIGURE 4: Variation of crystallite size of PSMO samples with different Sr concentration (x).

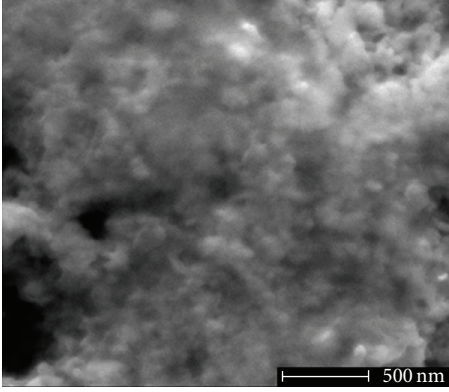


FIGURE 5: SEM picture showing the morphology of a fractured portion of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x = 0.40$).

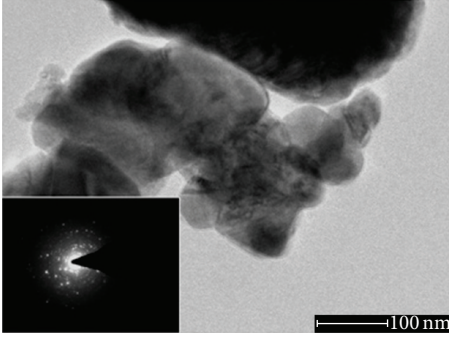


FIGURE 6: Representative TEM picture showing the crystallite size of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x = 0.40$). The inset shows the corresponding selected area diffraction pattern.

the crystallite size calculated from the XRD data. This difference is due to the fact that grains are composed of several crystallites, probably due to the internal stress or defects in the structure. A representative typical micrograph showing the surface morphology of the fractured portion of $x = 0.40$ sample is shown in Figure 5. The crystallite size was also determined by TEM. All the samples were observed to consist of nanocrystalline crystallites of average size ~ 40 nm and in majority of cases these nanocrystals were found to be present in the form of clusters. The TEM investigations also revealed that the crystallite size was not uniform and in some areas crystals as large as 60–70 nm were also observed. However, the density of such crystals was relatively small. A representative TEM micrograph unraveling local area microstructure and the corresponding selected area electron diffraction pattern showing the presence of nanocrystals and nanocrystalline cluster are shown in Figure 6.

3.2. Magnetic Characterization. The magnetic phase characterization was carried out by the temperature dependent AC susceptibility (χ - T) which was measured using the lock-in technique. The variation of AC susceptibility with temperature is shown in Figure 7 (left panels). The paramagnetic to ferromagnetic (PM-FM) phase transition temperature (T_C)

was determined from the first-order temperature derivative ($d\chi/dT$) of the χ - T data, which is plotted in the right panel of Figure 7. The T_C has been defined as the temperature corresponding to the peak in the $d\chi/dT$ curve. The sample with Sr content $x = 0.40$ shows onset of transition around 315 K and a sharp PM-FM transition at $T_C = 308$ K (value corresponding to the peak in the first-order derivative of χ). This value is nearly equal to the values reported for single crystals of similar composition [15] and similar to the bulk samples. There is no trace of FM-AFM transition within the temperature range of 77–350 K. The half doped sample with Sr content $x = 0.50$ also shows transition from PM-FM state at around 308 K, but, in this case, the transition width has slightly larger value, as shown by relatively broader peak in the $d\chi/dT$. On the lower temperature side, the susceptibility decreases around $T \sim 225$ K. This could be due to appearance of A-type AFM ordering. In samples with Sr content $x = 0.55$ and 0.60, the FM transition temperature is found to decrease slightly to $T_C \sim 300$ K and the transition width also increases. But the most interesting observation is the occurrence of a second transition for the samples with Sr content $x = 0.55$ and 0.60 that generally corresponds to an A-AFM magnetic structure. The PM-FM transition is followed by a kink and change in the slope of the χ - T curve (marked by double arrow in Figure 7). The smaller peak, which occurs at $T \sim 273$ and 255 K, respectively, for sample $x = 0.55$ and 0.60, in the $d\chi/dT$ curve corresponds to this transition. This is due to the presence of a second magnetic phase that has a lower magnetic moment, such as the A-AFM phase that consists of two-dimensional ferromagnetic sheets coupled in an antiferromagnetic manner. In samples having Sr concentration $x = 0.55$ and 0.60, the presence of the AFM phase is also evidenced by sharp decrease in the susceptibility in the lower temperature regime. Thus, the present results show that at $x = 0.40$ and 0.50 samples are ferromagnetic below T_C , while other samples ($x = 0.55$ and 0.60) have a mixed phase consisting of FM and A-type AFM phases. In nanomanganites with an AFM ground state, the superexchange interaction drives AFM, which is diluted by the surface disorder [6, 16, 19, 20]. This induces a reorganization of the disordered surface spins and has been explained in terms of the core-shell model [19, 20].

3.3. Electrical Transport Characterization. Electrical transport characterization of all samples was performed by resistance measurements in zero magnetic field as well as a DC magnetic field $H = 3$ kOe by four-probe technique. The resistivity was calculated from the formula $R = \rho L/A$, where R is the resistance, ρ is the resistivity, L is the distance between the voltage probes, and A is the cross-sectional area of the sample. Percentage magnetoresistance/magnetoresistivity (MR) was calculated by $\text{MR} = (\rho_0 - \rho_H) \times 100/\rho_0$, where ρ_0 is the resistivity at zero magnetic field and ρ_H is the resistivity at $H = 3$ kOe. The temperature dependence of resistivity (in the temperature range of 4.2–350 K) and MR (in the temperature range of 77–300 K) is plotted in Figure 8. The resistivity of all the samples is much larger (\sim few Ω -cm) than the single crystals of corresponding composition [15, 18] which is of the order m Ω -cm or even smaller. This is due

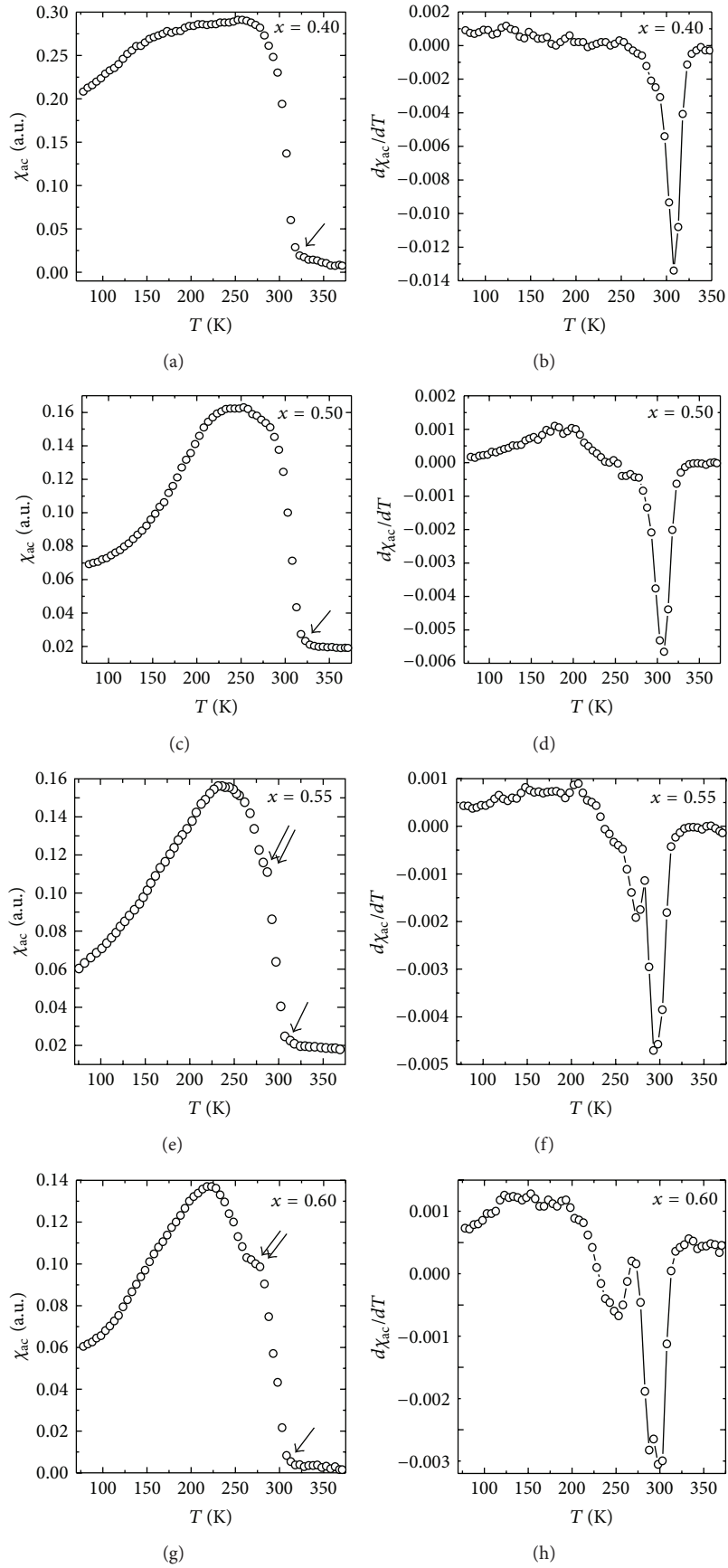


FIGURE 7: Temperature dependent AC susceptibility (left panel) and corresponding temperature derivatives (right panel) of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ samples with $x = 0.40, 0.50, 0.55,$ and 0.60 .

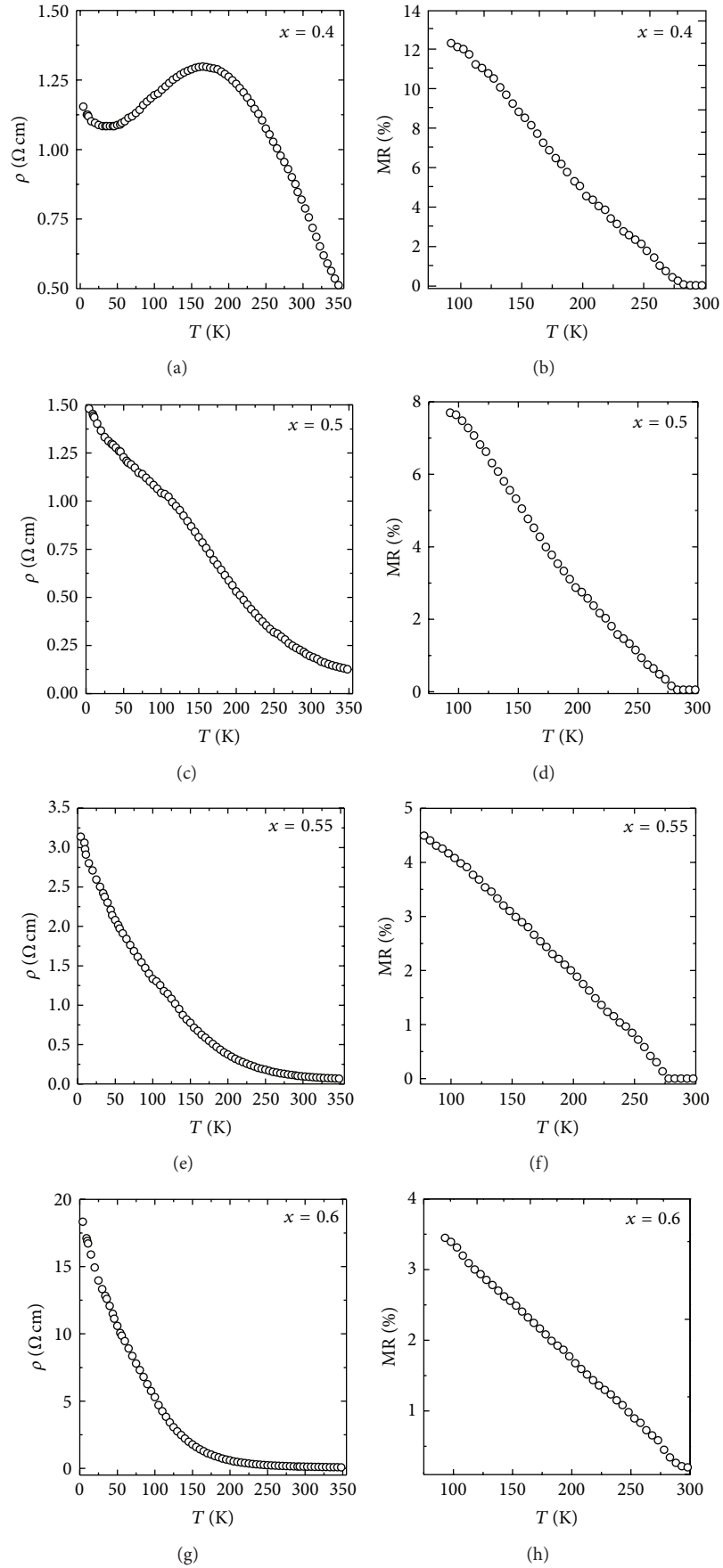


FIGURE 8: The temperature dependent resistivity (left) and MR (right) plots of $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ samples with $x = 0.40, 0.50, 0.55, \text{ and } 0.60$.

to large contribution of the grain boundaries in the present nanocrystalline samples. Because of the increased surface area which leads to increased grain boundary disorder as discussed in a previous section, the carrier scattering is strongly enhanced in the GB region. As seen in the ρ - T curves plotted in Figure 8, for the sample $x = 0.40$ the resistivity first increases on lowering the temperature up to 165 K and then shows an insulator to metal (I-M) transition, which is much lower than the PM-FM transition temperature T_C . In case of $x = 0.50$ sample, the resistivity first increases on lowering the temperature and then shows a hump- or plateau-like region. However, no I-M transition is seen in this case and, in fact below the humped region, the resistivity shows a very strong enhancement. Nearly similar trend is showing samples having higher Sr concentration ($x = 0.55$ and 0.60). Disappearance of the I-M transition at $x = 0.50$ or higher is indicative of the fact that in these samples metallic component is reduced. In the composition range $x = 0.50$ – 0.60 , the electrical characteristic is determined by competing metallic and insulating phases. The metallic phases are generally contributed by FM and A-AFM phases, while the insulating characteristics are mainly due to the presence of grain boundaries. The absence of I-M transition shows that the contribution of grain boundaries is rather dominant in these samples ($x = 0.50$ – 0.60) and one possible scenario is that the induced FM phase is not metallic but insulating or has relatively lower conductivity. This can be attributed to the nanocrystalline nature of the sample. Because of the small grain/crystallite size the electrical transport is dominated by the contribution from the grain boundaries. The grain boundary contribution envelopes the I-M transition, which occurs in the vicinity of T_C in single crystalline materials.

The decrease in the metallic component at higher Sr concentration is also evidenced by the successive reduction in the magnetoresistance (MR) as Sr content increases from $x = 0.40$ to 0.60 . As seen in the right panel of Figure 8, sample $x = 0.40$ shows the highest MR $\sim 12\%$ at 77 K and $H = 3$ kOe. The MR values are measured to be ~ 8 , 4.5 , and 3.5% for $x = 0.50$, 0.55 , and 0.60 , respectively. This gradual and systematic decrease in MR magnitude confirms the conjecture that induced FM phase in the present nanocrystalline samples may not be metallic.

4. Conclusions

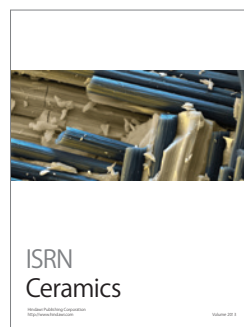
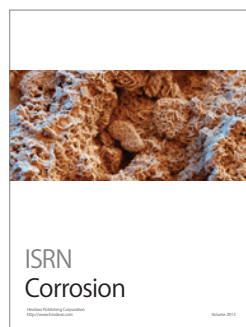
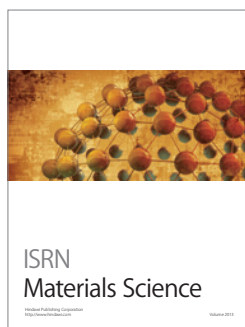
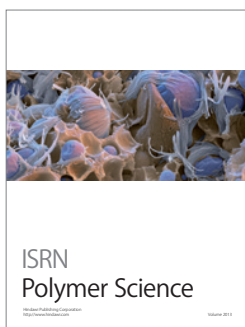
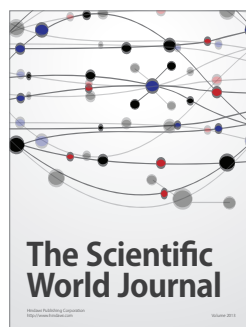
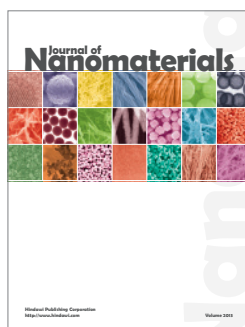
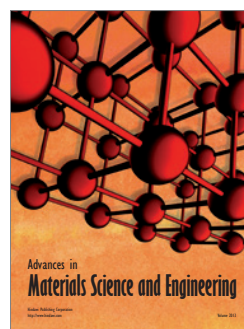
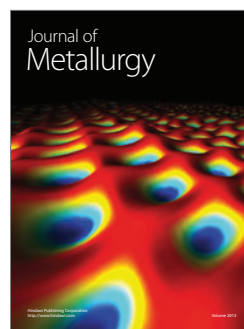
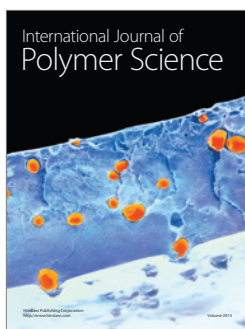
We have synthesized $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x = 0.40, 0.50, 0.55$, and 0.60) nanocrystalline samples by wet-chemical sol-gel route and studied their structural, magnetic, and electrical transport properties. All the samples are single phase and possess orthorhombic structure with space group Pbnm. The samples with $x = 0.40$ and 0.50 show a paramagnetic to ferromagnetic (PM-FM) transition at $T_C \sim 308$ K. However, interestingly, a second transition is observed at $T \sim 273$ and 255 K, respectively, for samples $x = 0.55$ and 0.60 correspond to an A-AFM phase which indicates that, below T_C , samples with $x = 0.55$ – 0.60 have a mixed phase consisting of FM and A-type AFM phases. In these nanocrystalline samples, the resistivity is much larger than the single crystals of corresponding compositions due to large contribution of

grain boundaries. Moreover, the gradual decrease in the MR from ~ 12 to 3.5% is observed with increasing the Sr content from $x = 0.40$ to 0.60 which also indicates the decrease in the metallic component at higher Sr concentrations.

References

- [1] Y. Tokura and Y. Tomioka, "Colossal magnetoresistive manganites," *Journal of Magnetism and Magnetic Materials*, vol. 200, no. 1, pp. 1–23, 1999.
- [2] F. Chen, H. W. Liu, K. F. Wang et al., "Synthesis and characterization of $\text{La}_{0.825}\text{Sr}_{0.175}\text{MnO}_3$ nanowires," *Journal of Physics*, vol. 17, no. 44, pp. L467–L475, 2005.
- [3] S. S. Rao, K. N. Anuradha, S. Sarangi, and S. V. Bhat, "Weakening of charge order and antiferromagnetic to ferromagnetic switch over in $\text{Pr}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ nanowires," *Applied Physics Letters*, vol. 87, no. 18, Article ID 182503, 3 pages, 2005.
- [4] S. S. Rao, S. Tripathi, D. Pandey, and S. V. Bhat, "Suppression of charge order, disappearance of antiferromagnetism, and emergence of ferromagnetism in $\text{Nd}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ nanoparticles," *Physical Review B*, vol. 74, no. 14, Article ID 144416, 5 pages, 2006.
- [5] Z. Q. Wang, F. Gao, K. F. Wang, H. Yu, Z. F. Ren, and J.-M. Liu, "Synthesis and magnetic properties of $\text{Pr}_{0.57}\text{Ca}_{0.43}\text{MnO}_3$ nanoparticles," *Materials Science and Engineering B*, vol. 136, no. 1, pp. 96–100, 2007.
- [6] A. Biswas and I. Das, "Experimental observation of charge ordering in nanocrystalline $\text{Pr}_{0.65}\text{Ca}_{0.35}\text{MnO}_3$," *Physical Review B*, vol. 74, no. 17, Article ID 172405, 4 pages, 2006.
- [7] A. Biswas, I. Das, and C. Majumdar, "Modification of the charge ordering in $\text{Pr}_{1/2}\text{Sr}_{1/2}\text{MnO}_3$ nanoparticles," *Journal of Applied Physics*, vol. 98, no. 12, Article ID 124310, 5 pages, 2005.
- [8] K. S. Shankar, S. Kar, G. N. Subbanna, and A. K. Raychaudhuri, "Enhanced ferromagnetic transition temperature in nanocrystalline lanthanum calcium manganese oxide ($\text{La}_{0.67}\text{Ca}_{0.33}\text{MnO}_3$)," *Solid State Communications*, vol. 129, no. 7, pp. 479–483, 2004.
- [9] T. Zhang, C. G. Jin, T. Qian, X. L. Lu, J. M. Bai, and X. G. Li, "Hydrothermal synthesis of single-crystalline $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ nanowires at low temperature," *Journal of Materials Chemistry*, vol. 14, pp. 2787–2789, 2004.
- [10] S. Dong, F. Gao, Z. Q. Wang, J.-M. Liu, and Z. F. Ren, "Surface phase separation in nanosized charge-ordered manganites," *Applied Physics Letters*, vol. 90, no. 8, Article ID 082508, 3 pages, 2007.
- [11] C. L. Lu, S. Dong, K. F. Wang et al., "Charge-order breaking and ferromagnetism in $\text{La}_{0.4}\text{Ca}_{0.6}\text{MnO}_3$ nanoparticles," *Applied Physics Letters*, vol. 91, no. 3, Article ID 032502, 3 pages, 2007.
- [12] M. Nagao, T. Asaka, D. Akahoshi et al., "Nanostructural evidence at the phase boundary of A- and C-type antiferromagnetic phases in $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$ crystals," *Journal of Physics*, vol. 19, no. 49, Article ID 492201, 2007.
- [13] D. Akahoshi, R. Hatakeyama, M. Nagao, T. Asaka, Y. Matsui, and H. Kuwahara, "Anomalous ferromagnetic behavior and large magnetoresistance induced by orbital fluctuation in heavily doped $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$ ($0.57 \leq x \leq 0.75$)," *Physical Review B*, vol. 77, no. 5, Article ID 054404, 5 pages, 2008.
- [14] R. Prasad, M. M. Singh, P. K. Siwach, P. Fournier, and H. K. Singh, "Anomalous insulator-metal transition and weak ferromagnetism in $\text{Nd}_{0.37}\text{Sr}_{0.63}\text{MnO}_3$ thin films," *Europhysics Letters*, vol. 84, no. 2, Article ID 27003, 2008.

- [15] Y. Tokura, "Critical features of colossal magnetoresistive manganites," *Reports on Progress in Physics*, vol. 69, no. 3, p. 797, 2006.
- [16] Y. Moritomo, T. Akimoto, A. Nakamura, K. Ohoyama, and M. Ohashi, "Antiferromagnetic metallic state in the heavily doped region of perovskite manganites," *Physical Review B*, vol. 58, no. 9, pp. 5544–5549, 1998.
- [17] H. Kawano, R. Kajimoto, H. Yoshizawa, Y. Tomioka, H. Kuwahara, and Y. Tokura, "Magnetic ordering and relation to the metal-insulator transition in $\text{Pr}_{1-x}\text{Sr}_x\text{MnO}_3$ and $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$ with $x \sim 1/2$," *Physical Review Letters*, vol. 78, no. 22, pp. 4253–4256, 1997.
- [18] J. A. Fernandez-Baca, P. Dai, H. Y. Hwang, C. Kloc, and S.-W. Cheong, "Evolution of the low-frequency spin dynamics in ferromagnetic manganites," *Physical Review Letters*, vol. 80, no. 18, pp. 4012–4015, 1998.
- [19] R. N. Bhowmik, R. Nagarajan, and R. Ranganathan, "Magnetic enhancement in antiferromagnetic nanoparticle of CoRh_2O_4 ," *Physical Review B*, vol. 69, Article ID 054430, 5 pages, 2004.
- [20] T. Zhang, T. F. Zhou, T. Qian, and X. G. Li, "Particle size effects on interplay between charge ordering and magnetic properties in nanosized $\text{La}_{0.25}\text{Ca}_{0.75}\text{MnO}_3$," *Physical Review B*, vol. 76, no. 17, Article ID 174415, 7 pages, 2007.



Novel Hybrid Spectrum Handoff for Cognitive Radio Networks

Nisar A. Lala^{a*}, Moin Uddin^b, N A Sheikh^c

^a *Division of Agricultural Engineering, SKUAST(K), Srinagar, J & K, India*

^b *Delhi Technological University, Delhi, India*

^c *Department of Mathematics, NIT Srinagar, J & K, India*

Abstract

Cognitive radio (CR) is projected as a technology (or solution) that will raise the spectrum utilization considerably by allowing low-priority or secondary user (SU) to utilize the spectrum of high-priority or primary user (PU) opportunistically. Spectrum handoff is a different type of handoff necessitated by the reappearance of the primary user on the frequency channels occupied by the secondary user at that time and location. In this paper, a hybrid type of spectrum handoff algorithm is proposed where proactive decision and reactive decision approaches are combined. Depending on the arrival rate of primary user (i.e. PU activity), the algorithm switches from reactive decision mode to proactive decision mode and vice versa. The switching from one mode to another mode depends on threshold value of PU activity and we evaluated the threshold value through analysis for switching of the algorithm to be 0.37. Simulated results show that the proposed hybrid spectrum handoff algorithm reduces the total service time of secondary user considerably compared to conventional proactive decision or reactive decision handoff approaches.

Index Terms: Cognitive radio; spectrum handoff; proactive decision handoff; reactive decision handoff; queueing theory.

© 2013 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science

1. Introduction

The spectrum is a precious natural resource and is presently regulated by governmental agencies to avoid interference among users and is allocated by fixed spectrum allocation policy. The fast growth of new wireless applications and services has resulted in increased demand of radio spectrum access. But most of the radio spectrum stands already allocated by fixed allocation policy and it becomes difficult to find unallocated spectrum for these new upcoming applications and services. As per the survey of Federal Communications Commission (FCC) [1], up to 85% of the assigned spectrum is underutilized. This allocation policy has

* Corresponding author

E-mail address: lalanisar_ae@rediffmail.com

created a situation where there appears an artificial scarcity of the spectrum. This ever increasing demand of spectrum for new applications cannot be fulfilled unless an alternate scheme to regulate the scarce spectrum is not found. Therefore, FCC has suggested a new communication paradigm for accessing the assigned spectrum dynamically [2] known as cognitive radio (CR). CR is a key technology that will make the dynamic spectrum access (DSA) a reality. DSA allows the SU to dynamically adjust its operating parameters (such as transmit power, modulation, operating frequency) in order to adapt to time varying radio environment and utilize the idle spectrum opportunistically [3-6], not used by the primary user at that time and location. In addition to spectrum sensing and management, another important function a CR should perform is spectrum mobility and gives rise to a different type of handoff in cognitive radio networks known as spectrum handoff. The aim of spectrum handoff is to help the SU to release the occupied channels instantly in order to avoid interference to the primary user and find suitable idle channels to restart the interrupted transmission. The handoff to new channels should be seamless so that an application running on the SU perceives minimum quality of service (QoS) degradation [7]. The CR technology allows SUs to sense the environment to find spectrumholes (or idle channel) and utilize these spectrum holes for transmission opportunity with the constraint of non-interference to the primary user. In cognitive radio, the PU has the priority to access the spectrum.

Depending on the decision method used for selecting the idle channels for future handoffs, the spectrum handoff process is classified as the proactive decision and reactive decision handoff approaches.

- In case of proactive decision handoff [8-12], the channels to be used for future handoffs are decided before actual data transmission takes place. The SU senses the wideband spectrum periodically for detection of idle channels so that expected usage pattern of the wideband spectrum over longer period is generated. Then the CR predicts the channels having highest probability of appearing idle at the time of actual handoff.
- In case of the reactive decision handoff [13-14], idle channels are detected through instantaneous sensing of the wideband spectrum after arrival of the PU.

There have been many studies applying queueing theory to study spectrum handoff in cognitive radio networks. The authors of [15], proposed the comparative analysis of two approaches namely proactive decision and reactive decision. The analysis of total service time of proactive decision handoff algorithm with multiple interruptions was proposed in [16] while analysis of extended data delivery time of reactive decision was proposed in [17]. The authors of [18] proposed spectrum management techniques in cognitive radio networks with main focus on QoS provisioning.

In this paper, we focus on the performance analysis of the proposed hybrid spectrum handoff algorithm against the conventional proactive decision or reactive decision handoff approaches. The total service time of the algorithm with multiple interruptions is evaluated using pre-emptive resume priority M/G/1 queueing network model.

The rest of the paper is organized as follows. Section 2 introduces a PRP M/G/1 queueing network model used to evaluate total service time. Section 3 presents proactive decision spectrum handoff. Section 4 presents reactive decision spectrum handoff. Section 5 proposes the hybrid spectrum handoff algorithm. Section 6 presents the simulated results of the hybrid spectrum handoff algorithm and the conclusion is provided in section 7.

2. PRP M/G/1 Queueing Network

A PRP M/G/1 queueing network model [19-25] is used for the characterization of the spectrum usage interactions between primary users and secondary users. The transmission of the SU can be interrupted multiple times by the arrival of PU. Therefore, this model is used to calculate the total service time of the two

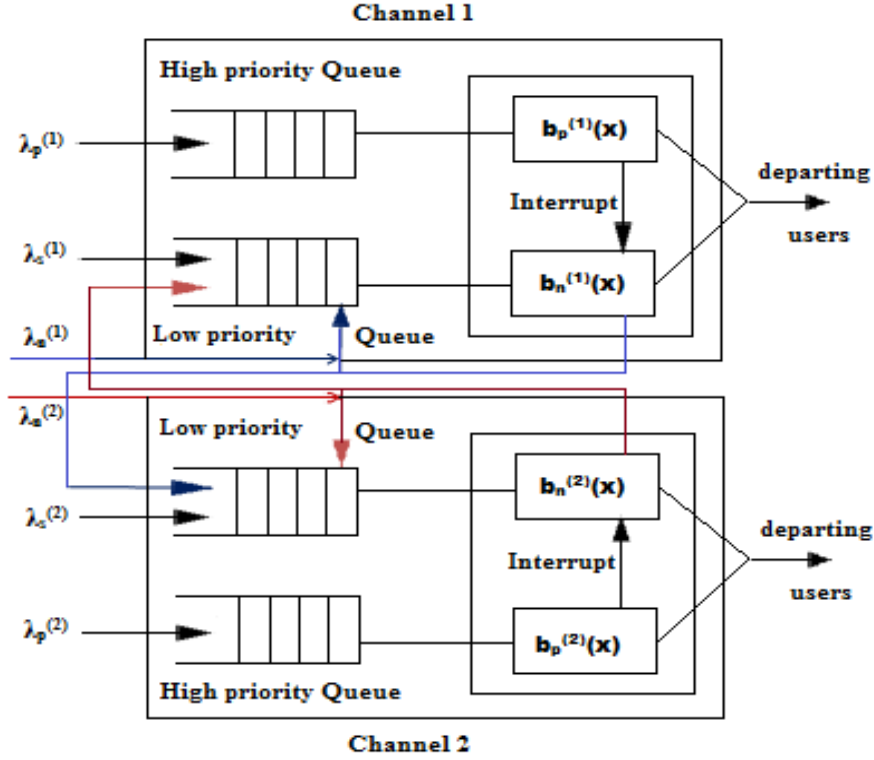


Fig. 1. The PRP M/G/1 queueing network for two channel system where n (Reproduced from [15])

approaches namely proactive decision and reactive decision. Important features of PRP M/G/1 queueing network model are given below:

- The transmissions of secondary users have low priority; therefore, can be interrupted by the arrival of PUs.
- When PU arrives back on the channels presently occupied by SU, the transmission of the SU is paused instantly in order to avoid interference.
- The interrupted SU has a choice to transmit either on the same channels or on other idle channels.
- In case of many secondary users contending for channel access, the access to channels is allowed as on first come first served basis.

Fig.1 shows an example of the PRP M/G/1 queueing network with two channels. The PUs are put into the high priority queue while the SUs are put into the low priority queue. The interrupted SU has two options, either to stay on the current channels or change the operating channels. In case of staying on the same channels, the unfinished transmission is put into the head of low-priority queue while in case of changing the operating channels, the unfinished transmission is put into the tail of low-priority queue. In both cases, the interrupted transmission is resumed as soon as the channels become idle. The parameter of importance in this model is the effective packet length and is defined as the duration of the transmission of the packet until PU arrives back on those channels.

3. Proactive Decision Spectrum Handoff

In case of proactive decision handoff, the decision for channel switching is taken prior to actual handoff. When PU arrives back, the SU pause its transmission and quickly handoff to the predetermined target channel. In this case, the total service time of SU may be reduced as there is no need of instantaneous spectrum sensing

and handshaking between the transmitting and receiving SUs. But at the time of actual handoff, the predetermined channels may be busy. Therefore, the SU has to wait in the queue till all the PUs and SUs complete their service. As a consequence of this, there could be an increase in actual handoff delay.

Let λ_p (arrivals/slot), λ_s (arrivals/slot) be the initial arrival rates of the primary users' and secondary users' connection at each channel and X_p (slots/arrival) and X_s (slots/arrival) be their corresponding service time. It is assumed that X_p and X_s are exponentially distributed with service rates μ_p and μ_s . In this paper, we have decided to change the channels when spectrum handoff takes place. If predetermined channel is idle then the transmission of secondary user is restarted on this new channel and the handoff delay happens to be the switching delay. If the predetermined channel happens to be busy then secondary user has to wait till all primary users and the secondary users in the queue complete their transmission. The handoff delay in this case is switching delay and waiting time on that channel.

Let $T_{proactive}$ be the average total service time, t_s be channel switching time and W_s be waiting time of secondary user on another channel until all primary users and other secondary users finish their transmission. The closed form expression for proactive decision spectrum handoff is derived in [16], we have

$$T_{proactive} = E[X_s] + E[N](W_s + t_s) \quad (1)$$

where $E[N]$ is the average number of interruptions, $E[X_s]$ is the average secondary service time. The average number of interruptions for a secondary user within a period of $E[X_s]$ can be obtained as

$$E[N] = \lambda_p E[X_s] \quad (2)$$

since the transmission of the secondary users will depend on the primary arrival rate. The total service time of the proactive handoff scheme comes out as

$$T_{proactive} = E[X_s] + (\lambda_p E[X_s]) \left(\frac{\lambda_p (E[X_p])^2 + \frac{\lambda_s}{(\lambda_p + \mu_s)\mu_s} + \frac{\rho_p^2}{1 - \rho_p} E[X_p]}{1 - \rho_p - \rho_s} + t_s \right) \quad (3)$$

where t_s is the channel switching time, $E[X_p]$ is the average primary service time, ρ_p is primary utilization is equal to $\rho_p = \lambda_p E[X_p]$ and ρ_s is secondary utilization is equal to $\rho_s = \lambda_s E[X_s]$.

4. Reactive Decision Spectrum Handoff

In case of reactive decision handoff, idle channels are found through instantaneous sensing of the wideband spectrum at the time of the actual handoff. The unfinished transmission of the interrupted SU will be restarted on these idle channels. In this approach, the handoff delay may be short as the target idle channels are found reliably through instantaneous sensing of the spectrum and the handoff delay is the sensing time plus the handshaking time necessary for consensus of the target idle channels between the transmitting and receiving SUs. Hence, time duration of spectrum sensing and handshaking is an important factor for assessing the performance of this approach.

The closed form expression is derived in [17] for total service time for reactive decision spectrum handoff as

$$T_{reactive} = E[X_s] + E[D] \quad (4)$$

where $E[X_s]$ is the average service time of secondary users and $E[D]$ is the average cumulative handoff delay. Then total service of reactive decision handoff ($T_{reactive}$) comes out as

$$T_{reactive} = E[X_s] + \frac{\lambda_p [t_p \mu_s + (E[X_p])^2 \lambda_p \mu_s + E[X_p] (\lambda_s - t_p \lambda_p \mu_s)]}{(1 - \lambda_p E[X_p]) (\mu_s)^2} \quad (5)$$

where t_p is the processing time and is given as $t_p = t_s + t_f$, t_s is channel switching time, t_f channel sensing time and μ_s is the secondary service rate equal to $\mu_s = 1/E[X_s]$ and μ_p is the primary service rate equal to $\mu_p = 1/E[X_p]$.

5. Proposed Hybrid Spectrum Handoff Algorithm

A hybrid type of spectrum handoff is proposed in this section, which is the combination of proactive decision and reactive decision approaches. Depending on the PU activity, the algorithm switches from proactive decision mode to reactive decision mode and vice versa. Equation (3) determines total service time of proactive decision (i.e. $T_{proactive}$) approach, while equation (5) determines total service time of reactive decision ($T_{reactive}$) approach. As can be seen from equations (3) and (5), the total service time of both decision approaches depend on average secondary service rate (μ_s), average primary service rate (μ_p), secondary arrival rate (λ_s) and mean sensing time (t_p). For different arrival rates of PU, the total service time of both decision approaches change due to change in number of perceived handoffs. In order to simplify the analysis of the proposed algorithm, it is assumed that each channel has identical traffic patterns and the hybrid handoff algorithm is simulated for two channel scenario. It is also assumed that channel switching time (t_s) is equal to zero so that channel processing time (t_p) is equal to the sensing time only.

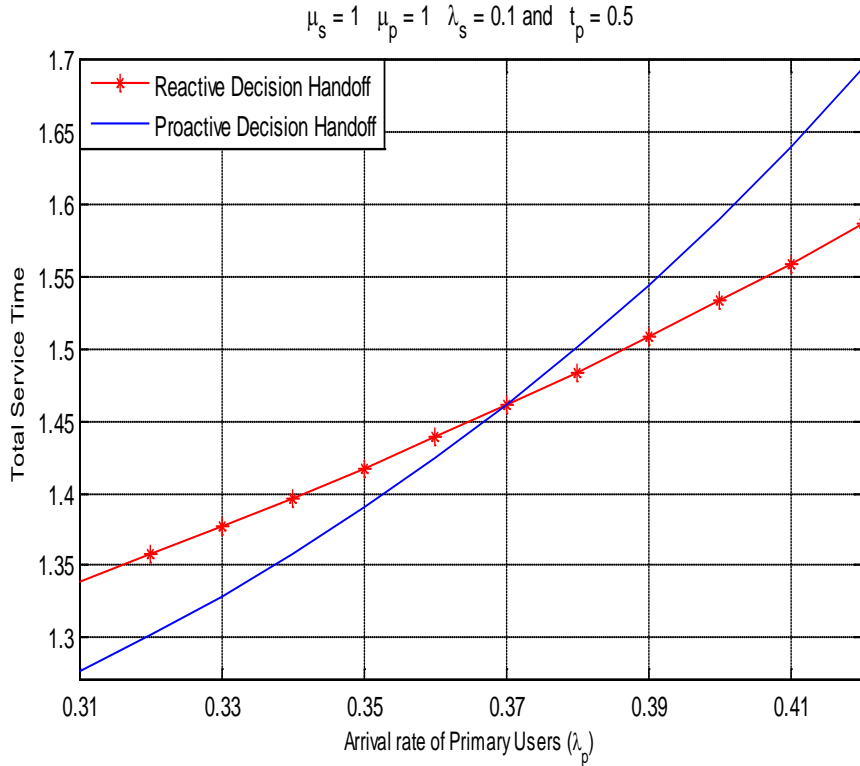


Fig.2. Comparison of reactive decision and proactive decision handoff approaches for threshold evaluation

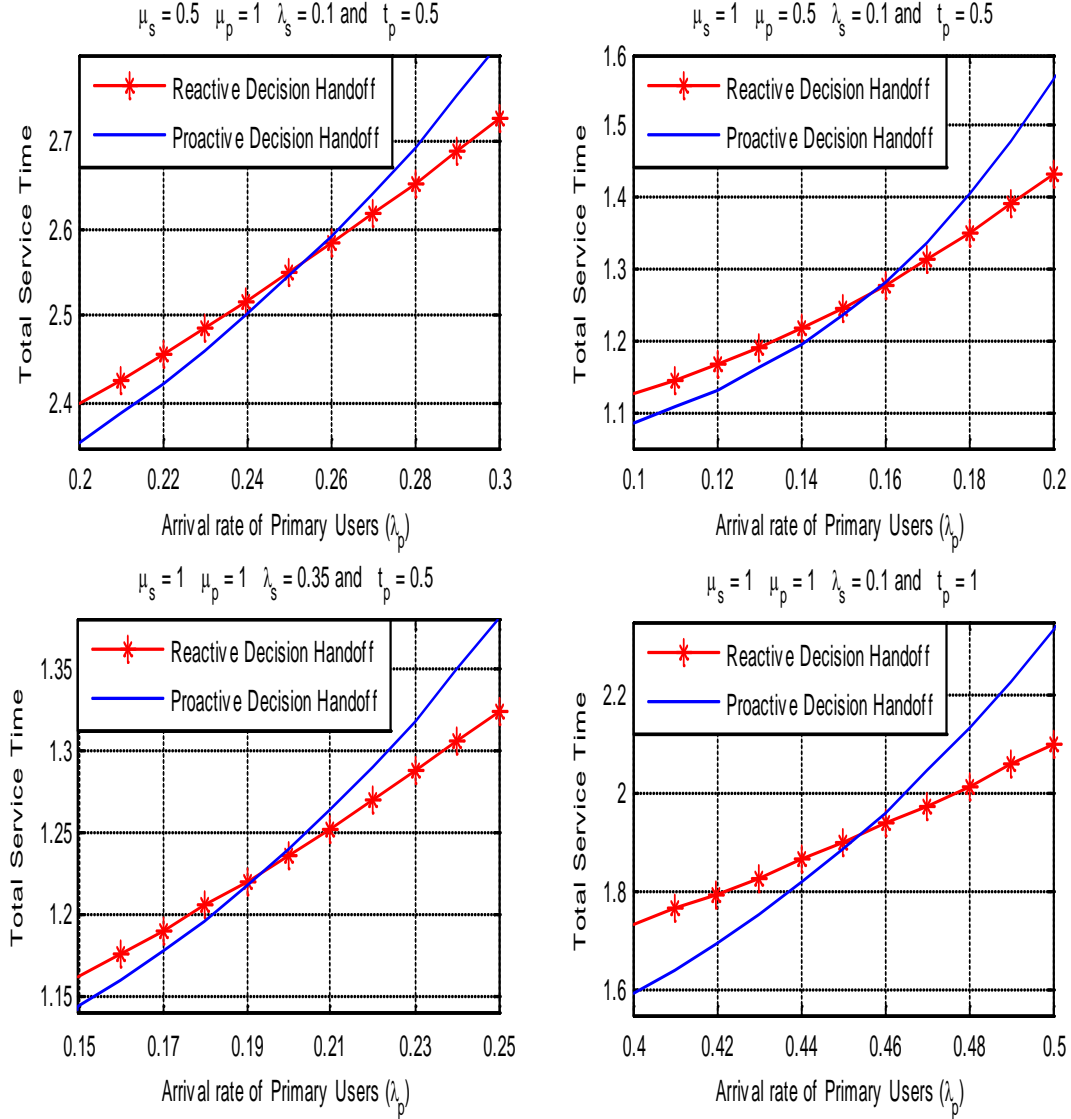


Fig.3. Sensitivity of μ_s , μ_p , λ_s and t_p parameters on threshold value of λ_p

6. Simulated Results

In this section, we presented the simulated results of the proposed hybrid handoff algorithm to support our analysis and the analysis of the algorithm is performed in Matlab 7.6. The first step for the analysis is to determine the value of threshold to be used for switching between proactive decision and reactive decision approaches and vice versa. In Fig. 2, we plot the total service time of proactive decision and reactive decision approaches as a function of primary arrival rate (λ_p). As can be seen from Fig. 2, for parameter values such as $\mu_s=1$, $\mu_p=1$, $\lambda_s=0.1$ and $t_p=0.5$, the threshold value of primary arrival rate (λ_p) comes out to be 0.37. We have analysed, the sensitivity of threshold value to these parameters, such as average secondary service rate (μ_s),

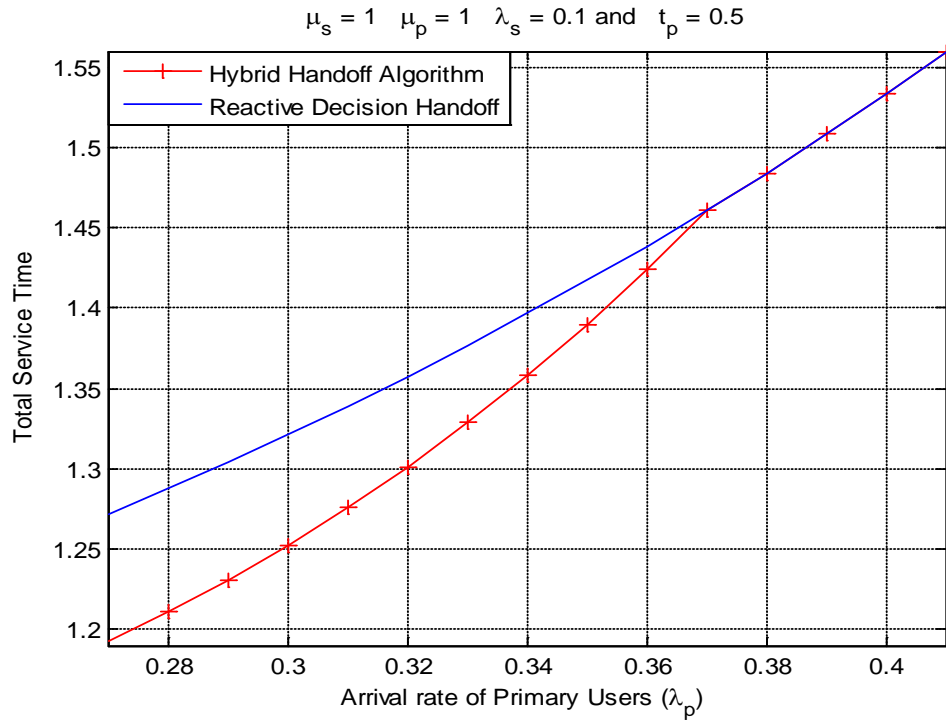


Fig. 4. Comparison of total service time of hybrid handoff and reactive decision handoff algorithms

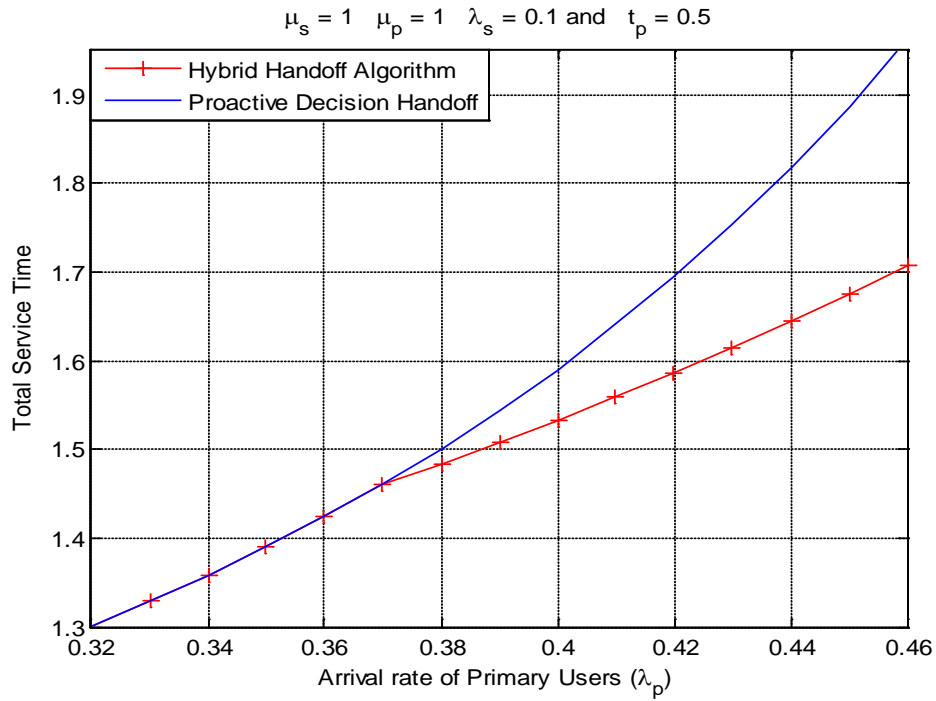


Fig. 5. Comparison of total service time of hybrid handoff and proactive decision handoff algorithms

average primary service rate (μ_p), secondary arrival rate (λ_s) and mean sensing time (t_p) with reference values of parameters such as $\mu_s=1$, $\mu_p=1$, $\lambda_s=0.1$ and $t_p=0.5$. Fig. 3 presents the impact analysis of varying one parameter at a time to threshold value in comparison to reference values. As can be seen from Fig. 3, for parameter values of $\mu_s=0.5$, $\mu_p=1$, $\lambda_s=0.1$ and $t_p=0.5$, the threshold value $\lambda_p = 0.252$, for parameter values $\mu_s=1$, $\mu_p=0.5$, $\lambda_s=0.1$ and $t_p=0.5$, the threshold value $\lambda_p=0.155$, for parameter values of $\mu_s=1$, $\mu_p=1$, $\lambda_s=0.35$ and $t_p=0.5$, the threshold value $\lambda_p = 0.192$ and for parameter values of $\mu_s=1$, $\mu_p=1$, $\lambda_s=0.1$ and $t_p=1$, the threshold value $\lambda_p=0.453$.

Fig. 4 presents the comparison of the proposed hybrid algorithm with reactive decision handoff approach. As can be seen from Fig. 4, the service time of the hybrid algorithm is lower until $\lambda_p < 0.37$. For $\lambda_p \geq 0.37$, the service time of both the algorithms is same. At lower values of λ_p (i.e. $\lambda_p < 0.37$), there is highest probability that the predetermined channels will appear to be idle at the time of actual handoff. In this case, handoff delay is only the switching delay and our hybrid algorithm experiences lesser number of handoffs which results in reduction of total service time of the SU.

Fig. 5 presents the comparison between the proposed hybrid algorithm with proactive decision handoff algorithm. As can be seen from Fig. 5, the service time of the of both algorithms is equal until $\lambda_p \leq 0.37$. For $\lambda_p > 0.37$, the service time of the hybrid algorithm is lower than the proactive decision algorithm. At higher values of λ_p (i.e. $\lambda_p > 0.37$), there is highest probability that the predetermined channels appear busy at the time of actual handoff. Therefore, SU will spend most of the time waiting in queue for service while reactive decision handoff approach will find idle channels reliably by instantaneous sensing. The total delay in this case is sensing and switching delay which is considerably lower than the waiting time in a queue.

Therefore, the proposed hybrid handoff algorithm will intelligently operate in proactive decision handoff mode when $\lambda_p < 0.37$, will switch to reactive decision handoff mode when $\lambda_p = 0.37$ and remains in this mode until $\lambda_p > 0.37$. The results demonstrate that there is considerable reduction in total service time of the secondary users.

7. Conclusion

We have proposed and analysed a hybrid spectrum handoff algorithm for total service time of the secondary user. The proposed hybrid algorithm switches from proactive decision mode to reactive decision mode and vice versa, depending on the threshold value of the primary arrival rate. We have analyzed the sensitivity of selected parameters such as μ_s , μ_p , λ_s and t_p on threshold value used for switching of the hybrid algorithm. The results demonstrate that for the parameter values $\mu_s = 1$, $\mu_p = 1$, $\lambda_s = 0.1$, and $t_p = 0.5$, the threshold value of arrival rate of primary users (λ_p) comes out to be $\lambda_p = 0.37$. The simulated results show that the proposed hybrid handoff algorithm reduces the total service time of SUs considerably as compared to conventional proactive decision or reactive decision handoff algorithms. The reduction in total service time results in increased throughput and will support higher quality of service (QoS) for SUs.

References

- [1] FCC. Notice of proposed rulemaking and order. No. 03-222, Dec. 2003.
- [2] Et docket No. 03-237, Nov. 2003. [Online] Available: <http://hraunfoss.fcc.gov/edocs/public/attachmatch/FCC-03-289A1.pdf>
- [3] Mitola J. "Cognitive radio: an integrated agent architecture for software defined radio". Ph.D. Dissertation: KTH Royal Institute of Technology; 2000.
- [4] Mitola J, Maguire GQ. "Cognitive radio: making software radios more personal". IEEE Personal Communications 1999; 6(4):13-8.
- [5] Haykin S. "Cognitive radio: Brain empowered wireless communications". IEEE Journal on Selected Areas in Communications 2005; 23(2): 201-20.

- [6] Akyildiz IF, Lee WY, Vuran MC, Mohanty S. "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey". *Computer Networks* (Elsevier) 2006; 50:2127–59.
- [7] Liu HJ, Wang ZX, LSF, Yi M. "Study on the performance of spectrum mobility in cognitive wireless Network". In *Proceedings of (IEEE) International Conference on Communication Systems (ICCS)* 2008.
- [8] Song Y, Xie J. "Proactive spectrum handoff in cognitive radio ad hoc networks based on common hopping coordination". In *Proceedings of (IEEE) INFOCOM* 2010.
- [9] Zheng S, Yang X, Chen S, Lou C. "Target channel sequence selection scheme for proactive-decision spectrum handoff". *IEEE communications letters* 2011; 15(12):1332–4.
- [10] Song Y, Xie J. "ProSpect: A proactive spectrum handoff framework for cognitive radio ad hoc networks without common control channel". *IEEE Transactions on mobile computing* 2012; 11(7):1127–39.
- [11] Srinivasa S, Jafar SA. "The throughput potential of cognitive radio: A theoretical perspective". *IEEE Communications Magazine* 2007; 73–9.
- [12] Shi Q, Taubenheim D, Kyperountas S, Gorday P, Correal N. "Link maintenance protocol for cognitive radio system with OFDM PHY". *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)* 2007.
- [13] Willkomm D, Gross J, Wolisz A. "Reliable link maintenance in cognitive radio systems". *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)* 2005.
- [14] Tian J, Bi G. "A new link maintenance and compensation model for cognitive UWB radio systems". In *Proceedings of International Conference on ITS Telecommunications* 2006.
- [15] Wang L-C, Wang C-W. "Spectrum handoff for cognitive radio networks: reactive sensing or proactive sensing?". *IEEE International Performance Computing and Communications Conference (IPCCC)* 2008.
- [16] Wang C-W, Wang L-C. "Modeling and analysis for proactive decision spectrum handoff in cognitive radio networks". *IEEE International Conference on Communications (ICC)* 2009.
- [17] Wang C-W, Wang L-C, Adachi F. "Modeling and analysis for reactive decision spectrum handoff in cognitive radio networks". *IEEE GLOBECOM* 2010.
- [18] Wang L-C, Wang C-W. "Spectrum management techniques with QoS provisioning in cognitive radio networks". *IEEE International Symposium on Wireless Pervasive Computing* 2010.
- [19] Zhang C, Wang X, Li J. "Cooperative cognitive radio with priority queueing analysis". *IEEE International Conference on Communications (ICC)* 2009.
- [20] Suliman I, Lehtomaki J. "Queueing analysis of opportunistic access in cognitive radios". *IEEE International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART)* 2009.
- [21] Li H. "Queueing analysis of dynamic spectrum access subject to interruptions from primary users". *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (Crown Com)* 2010.
- [22] Shiang H-P, Van der Schaar M. "Queueing based dynamic channel selection for heterogeneous multimedia application over cognitive radio networks". *IEEE Transactions on Multimedia* 2008; 10(5):896–909.
- [23] Bose SK. "An Introduction to queueing systems". Kluwer Academic/Plenum Publishers, New York; 2002.
- [24] Wang L-C, Wang C-W, Feng K-T. "A queueing theoretical framework for QoS enhanced spectrum management in cognitive radio networks". *IEEE Wireless Communications* 2011:18–26.
- [25] Cooper RB. "Introduction to queueing theory". 2nd ed. Elsevier North Holland Inc; 1981.

Authors' Profiles

Mr. Nisar Ahmad Lala received his B. Sc. and M.Sc. in Electronics in 1990 and 1994 respectively from University of Kashmir, J&K, India. He is working as Assistant Professor in the division of Agricultural Engineering, S. K. University of Agricultural Sciences and Technology (Kashmir) Srinagar, J & K, India. He is currently pursuing Ph.D at National Institute of Technology Srinagar, J&K, India. His research interests include Cognitive Radio Networks and Fuzzy modelling.



Prof. Moin Uddin is a senior member, IEEE. Moin Uddin did his B.Sc and M.Sc. Engineering in 1972 and 1978 from Aligarh Muslim University, Uttar Pradesh, India. He completed his Ph.D in 1993 from Roorkee University, India. He has more than 30 years of experience in academics and research and is currently serving as Pro-Vice Chancellor, Delhi Technological University, Delhi. Prior to this, he was Director of Dr. B R Ambedkar National Institute of Technology, Jalandhar, India. He has large number of publications in International and national Journals and fourteen research scholars have completed their Ph.D under his guidance and five more are pursuing the same. He has designed the computer engineering curriculums of many international and national universities and institutions and is among the expert panel of these universities. Prof. Moin Uddin is a life member ISTE national society and member, board of studies of many institutions.



Dr. Neyaz Ahmad Sheikh received his M.Sc, M.phil and Ph.D (Applied Mathematics) in 1991, 1993 and 1996 from Aligarh Muslim University, Uttar Pradesh, India. He is working as Assistant Professor in the department of mathematics, National Institute of Technology Srinagar, J & K, India. He has fourteen publications in International and national Journals. His research interests include Walsh functions and applications, Functional analysis and some results on wavelets.

Novel natural structure corrector of ApoE4 for checking Alzheimer's disease: Benefits from high throughput screening and molecular dynamics simulations

Manisha Goyal¹, Sonam Grover², Jaspreet Kaur Dhanjal³, Sukriti Goyal¹, Chetna Tyagi¹,
Sajeev Chacko⁴ and Abhinav Grover^{2*}

¹*Apaji Institute of Mathematics & Applied Computer Technology, Banasthali University,
Tonk, Rajasthan, India – 304022*

²*School of Biotechnology, Jawaharlal Nehru University, New Delhi, India – 110067*

³*Department of Biotechnology, Delhi Technological University, New Delhi, India – 110042*

⁴*Thematic Unit of Excellence on Computational Materials Science, S. N. Bose National
Centre for Basic Sciences, Sector-III, Block – JD, Salt Lake, Kolkata - 700098*

*Corresponding author Email: abhinavgr@gmail.com, agrover@mail.jnu.ac.in

Tel: +91-8130738032

Fax: +91-11-26702040

Abstract

A major genetic suspect for Alzheimer's disease is the pathological conformation assumed by apolipoprotein E4 (ApoE4) through intra-molecular interaction. In the present study, a large library of natural compounds was screened against ApoE4 to identify novel therapeutic molecules that can prevent ApoE4 from being converted to its pathological conformation. We report two such natural compounds PHC and IAH that bound to the active site of ApoE4 during the docking process. The binding analysis suggested that they have a strong mechanistic ability to correct the pathological structural orientation of ApoE4 by preventing repulsion between Arg 61 and Arg 112, thus inhibiting the formation of a salt bridge between Arg 61 and Glu 255. However, when the molecular dynamics simulations were carried out, structural changes in the PHC-bound complex forced PHC to move out of the cavity thus destabilizing the complex. However, IAH was structurally stable inside the binding pocket throughout the simulations trajectory. Our simulations results indicate that the initial receptor-ligand interaction observed after docking could be limited due to the receptor rigid docking algorithm and that the conformations and interactions observed after simulation runs are more energetically favored and should be better representations of derivative poses in the receptor.

Keywords: Alzheimer's; apolipoprotein E4; virtual screening; simulations; molecular dynamics; structure corrector

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia. AD is a harmful neurological disorder that affects about 5.4 million Americans of all ages [1]. One in every eight older Americans has AD, making it the sixth major cause of death in the United States [1]. In India, the annual incidence rate per 1,000 persons for AD is 11.67 for those above 55 years of age and even higher for those above 65 years [2]. AD, which affects memory, thinking ability, and behavior, is characterized by complex neuropathological features that include heaping of amyloid β ($A\beta$) followed by synaptic dysfunction, formation of neurofibrillary tangles, and elements of degenerating neurons [3]. Degeneration causes a decrease in the acetylcholine levels and in the activities of choline acetyltransferase [4].

Although the U.S. Food and Drug Administration (FDA) has approved 5 drugs that temporarily improve the condition of patients suffering with AD, none is fully effective because of associated toxic effects [1]. Tacrine, donepezil, rivastigmine, and memantine, for example, have significant side effects such as elevation of serum aminotransferase concentration, nausea, vomiting, diarrhea, anorexia, anxiety, and agitation [5-7]. The toxic effects of these drugs necessitate the development of new therapeutic compounds.

To develop a new drug, a computational approach is worthwhile and saves time. This approach involves screening new ligands for a specific target within a relatively short span of time. High throughput virtual screening (HTVS) is one of the most effective and rapid approaches for identifying probable inhibitors of the target protein [8]. Various potential drug targets have been reported to improve AD-associated pathological features such as acetylcholine esterases [9], NMDA receptor [10], and apolipoprotein E4 (ApoE4). ApoE plays a significant role in maintaining and repairing neurons. ApoE has three isoforms, namely, ApoE2, ApoE3, and ApoE4. The isoforms differ at residue positions 112 and 158 [11]. ApoE4 is the major genetic risk attributed to AD [12-17]. It acquires a pathological

conformation through an intra-molecular interaction, in which positively charged Arg 112 repels the side chain of Arg 61 in the amino terminal domain, allowing the formation of a salt bridge between Arg 61 and Glu 255 at the carboxyl terminal domain [18, 19]. Forty to eighty percent of patients with AD are estimated to possess at least one ApoE4 allele [20]. ApoE4 is less effective in maintaining and repairing neuronal cells compared to ApoE2 and ApoE3 [21-23]. ApoE4 also disrupts the normal process by which cells release excess A β , resulting in elevated levels of A β leading to its deposition in the brain [24-26]. ApoE4 uniquely performs neuron-specific proteolysis due to which harmful bio-active fragments are formed that can enter the cytosol, disrupt the mitochondrial energy balance, alter the cytoskeleton, and cause cell death [27-29]. ApoE is the only example of a susceptibility gene for AD [30] associated with lower glucose use and is believed to affect the hippocampus and cortex, areas found to be affected in patients with AD [31, 32]. It has been confirmed that the ApoE locus on chromosome 19 is strongly associated with the development of AD [12, 33, 34]. Small molecule structure correctors of ApoE4 have been suggested that effectively modulate the biophysical properties and the function of abnormal proteins. Some examples of ApoE4 structure correctors are GIND25 [35] and phthalazinone derivatives [36]. The evidential association of ApoE4 with increased risk of AD makes it a potential drug target for designing natural drug candidates for AD.

The present study focuses on identifying potential natural drug candidates as structure correctors for ApoE4. Keeping this goal in mind, a large database of natural compounds was screened against the 3D structure of ApoE4 using high throughput technology. *In silico* screening led to the identification of a new class of ApoE4 structure correctors that abolish the ApoE4 domain interaction. The molecular dynamics (MD) were then simulated to examine the dynamic behavior of molecular interactions between the screened compounds and the functional residues of ApoE4. This study paves the way for the development of novel

leads for AD treatment that have improved binding properties and pose low toxicity to humans.

2. Materials and methods

2.1. Protein preparation

The crystal structure of human ApoE4 [PDB ID: 1GS9], determined at a resolution of 1.70 Å, was retrieved from the Protein Data Bank [37]. ApoE4 contains a single domain of 22kD. To preprocess the retrieved structure of ApoE4, Protein Preparation Wizard in Schrodinger's Maestro interface [38] was used, followed by optimization [39].

2.2. Grid generation and ligand library preparation

The prepared protein structure was used to generate a grid using the receptor grid generation utility of the Glide docking module of the Schrodinger suite [40, 41]. Residues Arg-61, Glu-109, and Arg-112 form the catalytic triad in the active cleft of ApoE4 [36, 42]. The ligand library was prepared by extracting approximately 0.2 million natural compounds from the ZINC database [43] and processing them with Schrodinger's LigPrep Wizard [44] and using the Lipinski filter.

2.3. High throughput virtual screening and docking studies

The prepared ligand library was screened with the Glide GScore program [41, 45]. Glide uses a systematic method for virtual screening based on incremental construction searching and provides the output as the GScore scoring function combined with various other parameters. Glide's HTVS and extra precision (XP) algorithms combine to perform docking [46]. The screening against ApoE4 at the desired grid coordinates was performed

through the HTVS docking algorithm [40]. Compounds with a significant docking score were subjected to Glide XP, a more precise docking algorithm for further refined screening.

2.4. Molecular dynamics simulations of docked complexes

The MD were simulated to study the dynamical behavior of the top-scoring docked complexes using the GROMACS package [47]. Initially, amber force fields were applied using the Amber tool package [48]. GROMACS topology files were created by converting amber topology files using the AnteChamber Python Parser interface script. To get electrically neutral complexes, the complexes were solvated in a cubic box of water molecules, and appropriate counter-ions were added. The solvated system was minimized for about 10,000 steps using the steepest descent and conjugate gradient methods until the force on each atom was less than 100 kJ/mol/nm. The geometrically minimized systems were then subjected to isothermal molecular dynamics simulations.

3. Results and discussion

3.1. Outcomes of high throughput virtual screening and docking studies

Human ApoE4, one of the most promising drug targets for treating AD, was virtually screened against approximately 0.2 million compounds of the ZINC database. The screened compounds were ranked according to their binding affinity, calculated as the scoring function called the Glide GScore. Of all compounds, a total of 10,000 compounds were identified from HTVS out of which those with a Glide score of less than -6.0 (64 compounds) were subjected to the Glide XP docking protocol. The top two scoring compounds and their properties are listed in Table 1. The values of the other docking parameters used for evaluating the selection criteria of the top-scoring ligands are shown in Table 2.

The top-scoring compound (4-imimidazoleacetic acid hydrochloride; ZINC19735138; IAH) had a Glide score of -6.79 kcal/mol, while the second compound (2-methyl, 3-hydroxy-4,5-dihydroxymethylpyridin or pyridoxine hydrochloride; ZINC00049154; PHC) had a score of -6.76 Kcal/mol. The results revealed that IAH had a stronger binding affinity for human ApoE4 protein than PHC. Both ligands interacted with the two catalytic triad residues of ApoE4 in addition to other neighboring residues of the active site.

3.2. Binding mode analysis of ligand-docked ApoE4 complexes

3.2.1. ApoE4-IAH complex

In the case of the ApoE4-IAH complex, IAH interacted with the active site residues of ApoE4 (Fig. 1A) with the formation of 3 hydrogen bonds and numerous hydrophobic contacts. Arg 61, Asp 65, and Glu 109 were the residues participating in hydrogen bond formation (Fig. 1B). The NE and NH₂ atoms of basic catalytic amino acid Arg-61 formed 2 hydrogen bonds (3.28 Å, 2.73 Å) with the O atom of IAH. Other hydrogen bonds (2.49 Å, 2.71 Å) were formed by atom N₁ of IAH with the OE₂ atom of acidic active site residue Glu 109; and OD1 of neighboring acidic residue Asp-65 and atom N2 of IAH. In addition, Met-64 was involved in hydrophobic interaction in the ApoE4-IAH complex (Fig. 1C). Among all these interacting residues, Arg 61 and Glu 109 (part of the catalytic triad) are crucial amino acids and play a prominent role in abolishing the structural orientation of ApoE4. IAH bound to these residues, thus preventing interaction among them and improving the functionality of ApoE4. Various chemical properties of IAH were considered that supported its drug-likeness for AD treatment (Table 1). The topological polar surface area was reasonably high, which indicated that it can readily be absorbed in the human intestine and can penetrate the blood–brain barrier (BBB). In IAH, the presence of 12 heavy atoms and a high potential energy of

50.33 kcal/mol suggested that this ligand molecule has a good binding affinity for human ApoE4.

3.2.2. ApoE4-PHC complex

PHC is a single ringed structure with a molecular weight of 169.18 g/mol and lipophilicity value (logP) of -0.55 at pH 7. The topological polar surface area of PHC was also considered as it is very useful for identifying drug transport properties, human intestinal absorption, and BBB infiltration. The presence of a reasonable number of heavy atoms (9) and a good potential energy of 74 kcal/mol suggest PHC is capable of binding strongly with ApoE4 (Fig. 2A). In this study, PHC formed 4 hydrogen bonds and 1 hydrophobic contact with human ApoE4. As can be seen in Fig. 2B, 2 hydrogen bonds were formed between the NH₂ and NE atoms of active site residue Arg 61 and the O₃ atom of PHC with bond length 2.67 Å and 2.74 Å, respectively, while 2 others were formed with the OD1 and OD2 atoms of the neighboring residue Asp 65 and O₁ and O₂ atoms of PHC with bond length of 2.83 Å and 2.74 Å, respectively. However, acidic amino acid Glu 109 of the catalytic triad was involved in making hydrophobic contact with PHC as illustrated in Fig. 2C. Of all these residues, Arg 61 and Glu 109 as part of the catalytic triad are responsible for the structural aberration in the human ApoE4 protein. These interactions of PHC with the crucial residues of ApoE4 suggest that this is a promising ligand that could correct the functionality of abnormal ApoE4.

3.3. Molecular dynamics simulations of ligand-bound ApoE4 complexes

3.3.1. Interaction analysis of the ApoE4-PHC complex

For further refinement and stabilization of both docked complexes, the MD were simulated using the GROMACS package. The simulation lengths used in the study were long enough to allow rearrangement of the side chains of the native and the ligand-complexed

protein thus facilitating the most stable binding mode. As is evident in Fig. 3A, the backbone of the protein acquired stability after 8 ns with a root mean square deviation (RMSD) of only about 2.5 Å from its initial position. However, the MD simulations for ApoE4-PHC complex conducted for up to 24 ns revealed interesting results. PHC moved away from the binding site of ApoE4 during the simulations and lost all interactions formed in the initial docked pose. Fig. 4 illustrates the binding instability snapshots of PHC with ApoE4 during the simulation trajectory. During the MD simulations, the position of PHC in the ligand-bound complex was constantly altered. As can be seen from the snapshots at 6 ns and 8 ns, PHC moved far away from the binding site while staying at the surface of the protein. However, at 20 ns PHC was highly destabilized and split. Thus, it can be inferred that during the docking procedure the interactions of PHC with residues Arg 61, Arg 65, and Glu 109 of ApoE4 were only the result of static contacts. These pseudo-interactions readily vanished when dynamics was considered in the study.

3.3.2. Interaction analysis of MD-stabilized ApoE4-IAH complex

In the energetically stable ApoE4-IAH complex, the IAH molecule interacted with the residues Arg 61, Glu 109, and Arg 112 of the catalytic triad of ApoE4. The IAH molecule also formed contacts with the residues Met 64, Asp 65, Met 68, and Gly 105. Though some deviation of IAH was observed from its initial position leading to a change in its binding mode, the binding was stable inside the ApoE4 cavity. A comparative analysis of the interaction profiles of ApoE4-IAH complex before and after the MD simulations is described in Table 3. The superimposition of the ligand IAH in the pre- and post-MD simulated complex structures inside the active site of ApoE4 is depicted in Fig. 3B. Initially, IAH formed 4 hydrogen bonds with the residues Arg 61, Asp 65, and Glu 109 of ApoE. After the simulations, 3 hydrogen bonds with the residues Arg 61 and Asp 65 had been replaced with 2

new hydrogen bonds involved with amino acids Gly 105 and Met 64. The hydrogen bond with the residue Glu 109 remained consistent with a slight change in the bond length (Fig. 5A). The only hydrophobic contact with Met 64 was present in IAH-bound ApoE4 before MD disappeared during the MD simulations. However, after the MD simulations IAH formed strong hydrophobic contacts with 4 residues of ApoE4 (Fig. 5B). The stability of IAH in the binding pocket of ApoE4 is prominently governed by these hydrophobic contacts. After the MD simulations, IAH acquired a more stable conformation within the active site of ApoE4 by placing itself deep inside the cavity.

4. Conclusion

In the present work, we screened two top-scoring compounds, IAH and PHC, which possess high Glide XP scores of -6.79 kcal/ mol and -6.76 kcal/mol, respectively, against human ApoE4. These compounds interacted with the catalytic triad residues of ApoE4 that are crucial for maintaining its aberrant structure. The binding of these ligands suggests that they have a strong mechanistic ability to correct the pathological structural orientation of ApoE4 by preventing repulsion between Arg 61 and Arg 112, thus inhibiting the formation of a salt bridge between Arg 61 and Glu 255. The chemical properties of these potent structure-correctors are in line with the stipulated requirements of drug-like compounds for further experimental analysis. After the MD simulations, the interactions formed by IAH were consistent. However, a comparison between the conformations obtained from docking and that from molecular dynamics simulations for the second ligand PHC revealed substantial changes in binding conformations. Our simulation results indicate that the initial receptor-ligand interaction observed after docking can be limited due to the receptor rigid docking algorithm and that the conformations and interactions observed after the simulation runs are more energetically favored and should be better representations of the derivative poses in the

receptor. Our detailed binding analysis of IAH substantiated by its dynamic structural stability provides considerable evidence for use as a potent natural lead against Alzheimer's. Results from this study would also be helpful in designing novel neuro-regenerative drugs with improved binding properties and low toxicity.

Acknowledgements

AG is thankful to Jawaharlal Nehru University for usage of all computational facilities. AG also thanks support from DST as a recipient of INSPIRE Faculty Award and UGC as a recipient of Faculty Recharge Award.

References

- [1] "Alzheimer's Association: Alzheimer's Disease Facts and Figures," *Alzheimer's & Dementia*, vol. 8, pp. 1-67, 2012.
- [2] P. S. Mathuranath, A. George, N. Ranjith et al., "Incidence of Alzheimer's disease in India: a 10 years follow-up study," *Neurology India*, vol. 60, pp. 625-30, Nov-Dec 2012.
- [3] M. J. Sadowski, J. Pankiewicz, H. Scholtzova et al., "Blocking the apolipoprotein E/amyloid-beta interaction as a potential therapeutic approach for Alzheimer's disease," *Proceedings of National Academy of Sciences USA*, vol. 103, pp. 18787-92, Dec 5 2006.
- [4] T. G. Price DL, Borchlet DR, Martin LJ, Crain BJ, Sisodiya SS and Troncoso JC "Neuropathology of Alzheimer's disease and animal models," in *In "Neuropathology of dementing disorders"*, M. WR, Ed., 1998, pp. 121-141.

- [5] E. L. Conway, "A review of the randomized controlled trials of tacrine in the treatment of Alzheimer's disease: methodologic considerations," *Clinical Neuropharmacology*, vol. 21, pp. 8-17, Jan-Feb 1998.
- [6] S. I. Gracon, M. J. Knapp, W. G. Berghoff et al., "Safety of tacrine: clinical trials, treatment IND, and postmarketing experience," *Alzheimer Disease and Associated Disorders*, vol. 12, pp. 93-101, Jun 1998.
- [7] R. Mayeux and M. Sano, "Treatment of Alzheimer's disease," *New England Journal of Medicine*, vol. 341, pp. 1670-9, Nov 25 1999.
- [8] D. A. Ostrov, J. A. Hernandez Prada, P. E. Corsino, K. A. Finton, N. Le, and T. C. Rowe, "Discovery of novel DNA gyrase inhibitors by high-throughput virtual screening," *Antimicrobial Agents and Chemotherapy*, vol. 51, pp. 3688-98, Oct 2007.
- [9] A. Lleo, S. M. Greenberg, and J. H. Growdon, "Current pharmacotherapy for Alzheimer's disease," *Annual Review of Medicine*, vol. 57, pp. 513-533, 2006.
- [10] S. K. Sonkusare, C. L. Kaul, and P. Ramarao, "Dementia of Alzheimer's disease and other neurodegenerative disorders--memantine, a new hope," *Pharmacological Research*, vol. 51, pp. 1-17, Jan 2005.
- [11] K. H. Weisgraber, "Apolipoprotein E: structure-function relationships," *Advances in Protein Chemistry*, vol. 45, pp. 249-302, 1994.
- [12] W. J. Strittmatter, A. M. Saunders, D. Schmechel et al., "Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease," *Proceedings of National Academy of Sciences USA*, vol. 90, pp. 1977-81, Mar 1 1993.
- [13] E. H. Corder, A. M. Saunders, W. J. Strittmatter et al., "Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families," *Science*, vol. 261, pp. 921-3, Aug 13 1993.

- [14] R. W. Mahley, K. H. Weisgraber, and Y. Huang, "Apolipoprotein E4: a causative factor and therapeutic target in neuropathology, including Alzheimer's disease," *Proceedings of National Academy of Sciences USA*, vol. 103, pp. 5644-51, Apr 11 2006.
- [15] R. W. Mahley, K. H. Weisgraber, and Y. Huang, "Apolipoprotein E: structure determines function, from atherosclerosis to Alzheimer's disease to AIDS," *Journal of Lipid Research*, vol. 50 Suppl, pp. S183-8, Apr 2009.
- [16] R. W. Mahley and Y. Huang, "Alzheimer disease: multiple causes, multiple effects of apolipoprotein E4, and multiple therapeutic approaches," *Annals of Neurology*, vol. 65, pp. 623-5, Jun 2009.
- [17] J. Kim, J. M. Basak, and D. M. Holtzman, "The role of apolipoprotein E in Alzheimer's disease," *Neuron*, vol. 63, pp. 287-303, Aug 13 2009.
- [18] L. M. Dong, C. Wilson, M. R. Wardell et al., "Human apolipoprotein E. Role of arginine 61 in mediating the lipoprotein preferences of the E3 and E4 isoforms," *Journal of Biological Chemistry*, vol. 269, pp. 22358-65, Sep 2 1994.
- [19] L. M. Dong and K. H. Weisgraber, "Human apolipoprotein E4 domain interaction. Arginine 61 and glutamic acid 255 interact to direct the preference for very low density lipoproteins," *Journal of Biological Chemistry*, vol. 271, pp. 19053-7, Aug 9 1996.
- [20] L. A. Farrer, L. A. Cupples, J. L. Haines et al., "Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium," *Journal of the American Medical Association* vol. 278, pp. 1349-56, Oct 22-29 1997.
- [21] R. W. Mahley, "Apolipoprotein E: cholesterol transport protein with expanding role in cell biology," *Science*, vol. 240, pp. 622-30, Apr 29 1988.
- [22] R. W. Mahley and S. C. Rall, "Apolipoprotein E: Far more than a lipid transport protein," *Annual Review of Genomics and Human Genetics*, vol. 1, pp. 507-537, 2000.

- [23] K. H. Weisgraber and R. W. Mahley, "Human apolipoprotein E: the Alzheimer's disease connection," *Faseb Journal*, vol. 10, pp. 1485-94, Nov 1996.
- [24] R. E. Tanzi and L. Bertram, "Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective," *Cell*, vol. 120, pp. 545-55, Feb 25 2005.
- [25] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics," *Science*, vol. 297, pp. 353-6, Jul 19 2002.
- [26] K. Blennow, M. J. de Leon, and H. Zetterberg, "Alzheimer's disease," *Lancet*, vol. 368, pp. 387-403, Jul 29 2006.
- [27] M. E. Risner, A. M. Saunders, J. F. Altman et al., "Efficacy of rosiglitazone in a genetically defined population with mild-to-moderate Alzheimer's disease," *Pharmacogenomics Journal*, vol. 6, pp. 246-54, Jul-Aug 2006.
- [28] Y. Huang and R. W. Mahley, "Commentary on "Perspective on a pathogenesis and treatment of Alzheimer's disease." Apolipoprotein E and the mitochondrial metabolic hypothesis," *Alzheimers & Dementia*, vol. 2, pp. 71-3, Apr 2006.
- [29] A. D. Roses and A. M. Saunders, "Perspective on a pathogenesis and treatment of Alzheimer's disease," *Alzheimers & Dementia*, vol. 2, pp. 59-70, Apr 2006.
- [30] A. M. Saunders, M. K. Trowers, R. A. Shimkets et al., "The role of apolipoprotein E in Alzheimer's disease: pharmacogenomic target selection," *Biochimica et Biophysica Acta*, vol. 1502, pp. 85-94, Jul 26 2000.
- [31] E. M. Reiman, R. J. Caselli, K. Chen, G. E. Alexander, D. Bandy, and J. Frost, "Declining brain activity in cognitively normal apolipoprotein E epsilon 4 heterozygotes: A foundation for using positron emission tomography to efficiently test treatments to prevent Alzheimer's disease," *Proceedings of National Academy of Sciences USA*, vol. 98, pp. 3334-9, Mar 13 2001.

- [32] E. M. Reiman, K. Chen, G. E. Alexander et al., "Functional brain abnormalities in young adults at genetic risk for late-onset Alzheimer's dementia," *Proceedings of National Academy of Sciences USA*, vol. 101, pp. 284-9, Jan 6 2004.
- [33] A. M. Saunders, W. J. Strittmatter, D. Schmechel et al., "Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease," *Neurology*, vol. 43, pp. 1467-72, Aug 1993.
- [34] M. A. Pericak-Vance, L. H. Yamaoka, C. S. Haynes et al., "Genetic linkage studies in Alzheimer's disease families," *Experimental Neurology*, vol. 102, pp. 271-9, Dec 1988.
- [35] S. Ye, Y. Huang, K. Mullendorff et al., "Apolipoprotein (apo) E4 enhances amyloid beta peptide production in cultured neuronal cells: apoE structure as a potential therapeutic target," *Proceedings of National Academy of Sciences USA*, vol. 102, pp. 18700-5, Dec 20 2005.
- [36] H. K. Chen, Z. Liu, A. Meyer-Franke et al., "Small molecule structure correctors abolish detrimental effects of apolipoprotein E4 in cultured neurons," *Journal of Biological Chemistry*, vol. 287, pp. 5253-66, Feb 17 2012.
- [37] *Protein Data Bank*. <http://www.rcsb.org/pdb/home/home.do>
- [38] Schrodinger, "maestro," 9 ed. LLC, New York, NY, 2009.
- [39] S. Sreeramulu, H. R. Jonker, T. Langer, C. Richter, C. R. Lancaster, and H. Schwalbe, "The human Cdc37.Hsp90 complex studied by heteronuclear NMR spectroscopy," *Journal of Biological Chemistry*, vol. 284, pp. 3885-96, Feb 6 2009.
- [40] T. A. Halgren, R. B. Murphy, R. A. Friesner et al., "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening," *Journal of Medicinal Chemistry*, vol. 47, pp. 1750-9, Mar 25 2004.

- [41] R. A. Friesner, J. L. Banks, R. B. Murphy et al., "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy," *Journal of Medicinal Chemistry*, vol. 47, pp. 1739-49, Mar 25 2004.
- [42] C. Frieden and K. Garai, "Structural differences between apoE3 and apoE4 may be useful in developing therapeutic agents for Alzheimer's disease," *Proceedings of National Academy of Sciences USA*, vol. 109, pp. 8913-8, Jun 5 2012.
- [43] J. J. Irwin and B. K. Shoichet, "ZINC - A free database of commercially available compounds for virtual screening," *J. Chem. Inf. Model.*, vol. 45, pp. 177-182, Jan-Feb 2005.
- [44] Schrodinger, "ligprep," 2.3 ed. LLC, New york, NY, 2009.
- [45] Schrodinger, "Glide," 5.5 ed. LLC, New York, NY, 2009.
- [46] M. Sandor, R. Kiss, and G. M. Keseru, "Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes," *Journal of Chemical Information and Modeling*, vol. 50, pp. 1165-72, Jun 28 2010.
- [47] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, pp. 1701-18, Dec 2005.
- [48] D. A. Case, T. A. Darden, T. E. Cheatham et al., "AMBER 12." San Francisco: University of California, 2012.

Figure Captions

Figure 1: Molecular interactions between IAH (orange) and ApoE4 before MD simulations

(a) Position of IAH in the ligand-bound ApoE4 complex (b) Hydrogen bond interactions (c) Hydrophobic interactions

Figure 2: Molecular interactions between PHC (yellow) and ApoE4 before MD simulations

(a) Position of PHC in the ligand-bound docked complex (b) Hydrogen bond interactions (c) Hydrophobic interactions

Figure 3: MD simulations trajectories (a) RMSD trajectory of IAH in complex with ApoE4

obtained after MD simulations (b) Superimposition of pre- (orange) and post- (red) MD complexes of IAH with ApoE4.

Figure 4: Snapshots depicting the binding instability of PHC with APoE4 during the MD simulations trajectory

Figure 5: Molecular interactions between IAH (orange) and ApoE4 after MD simulations (a)

Hydrogen bond interactions (b) Hydrophobic interactions

Figures

Figure 1

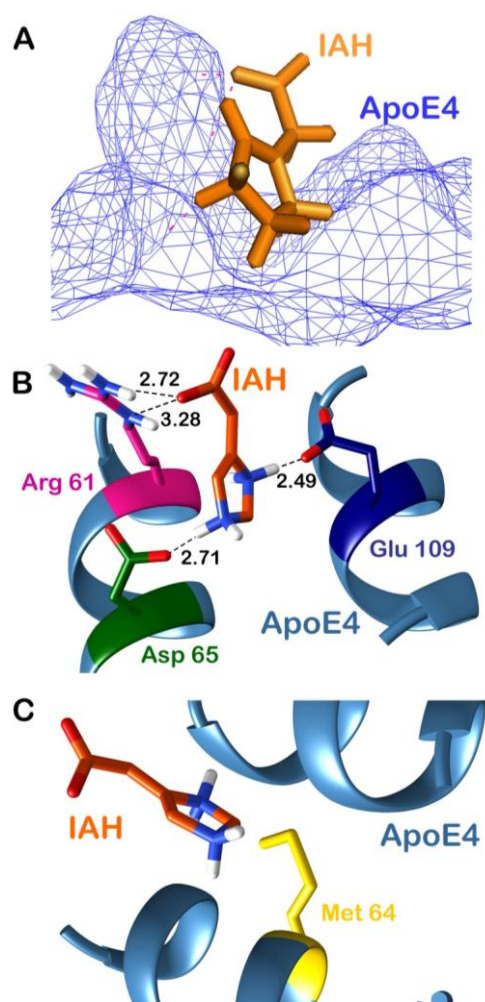


Figure 2

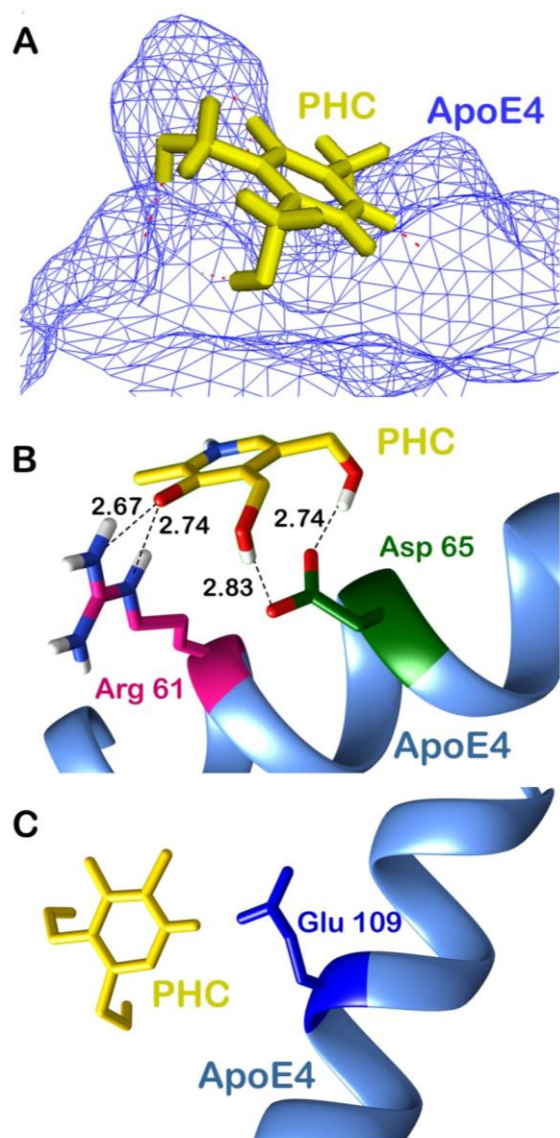


Figure 3

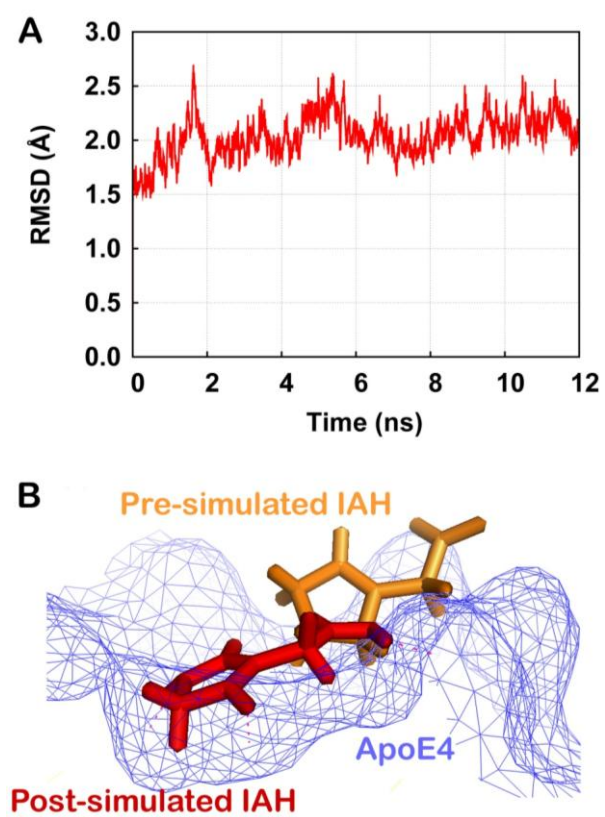


Figure 4

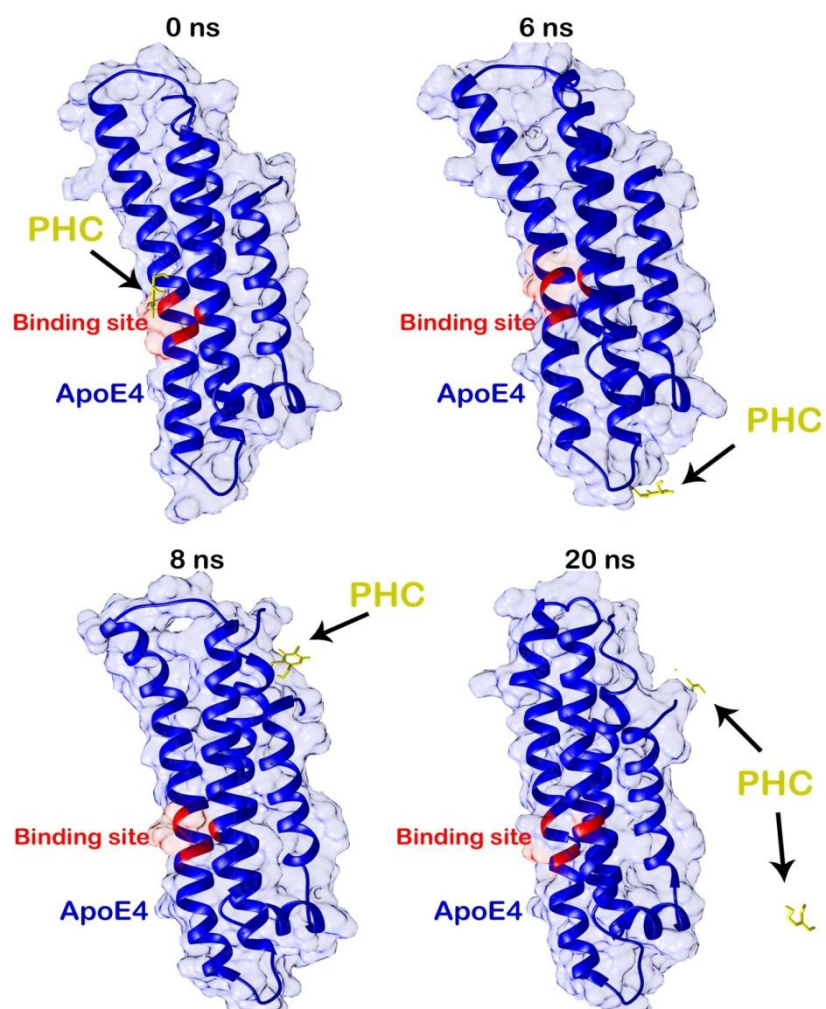
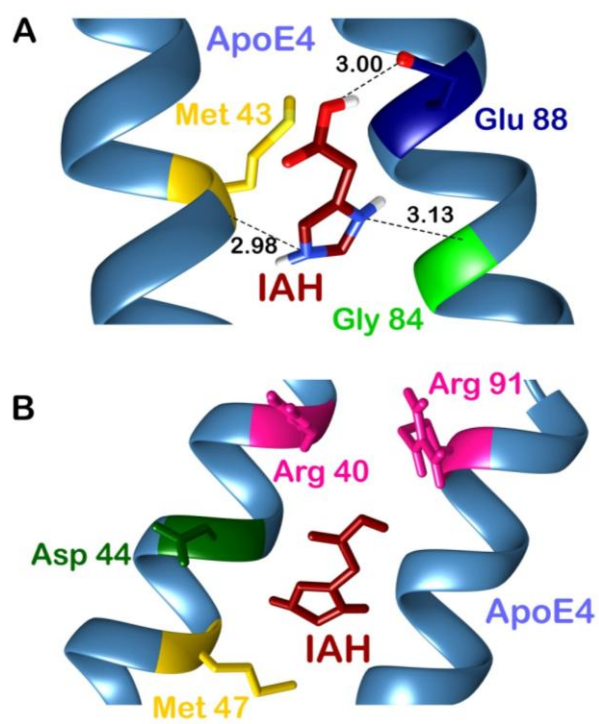
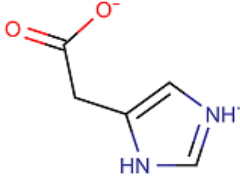
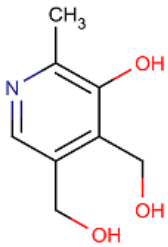


Figure 5



Tables

Table 1: Physical properties of potential structure correctors identified using virtual screening

Compound ZINC ID	Structure	logP value	Mol.wt. (g/ mol)	HBD	HBA	tPSA (Å ²)	Heavy atoms
ZINC19735138		-0.41	126.11	2	4	70	9
ZINC00049154		-0.55	169.18	3	4	74	12

Mol.wt.; Molecular weight, HBD; hydrogen bond donor, HBA; hydrogen bond Acceptor, tPSA; topological polar surface area.

Table 2: Binding affinity scores and energies of ApoE4 in complex with IAH and PHC.

Compound	ZINC ID	docking score	XP Gscore	glide ligand efficiency	glide evdw	glide emodel	glide energy
IAH	ZINC19735138	-6.79	-6.79	0.75	-3.28	-23.90	-28.96
PHC	ZINC00049154	-6.76	-6.76	-0.56	-6.18	-32.97	-26.97

Table 3: Molecular interactions present in pre- and post-MD simulated IAH-bound ApoE4 complexes.

ApoE4-IAH complex	Residues participating in hydrogen bonding	Residues governing hydrophobic contacts	Hydrogen bond length
Pre MD	Arg-61	Met-64	3.28, 2.73
	Asp-65		2.71
	Glu-109		2.49
Post MD	Met-64	Arg-61, Asp-65, Met-68, Arg-112	2.98
	Gly-105		3.14
	Glu-109		3.08

Prediction of Slot-Size and Inserted Air-Gap for Improving the Performance of Rectangular Microstrip Antennas using Artificial Neural Networks

Taimoor Khan, *Member, IEEE*, Asok De, *Senior Member, IEEE*, and Moinuddin, *Senior Member, IEEE*

Abstract—Artificial neural networks have been getting popularity for predicting various performance parameters of microstrip antennas due to their learning and generalization features. In this letter, a neural networks based synthesis model is presented for predicting the “slot-size” on the radiating patch and inserted “air-gap” between the ground plane and the substrate sheet, simultaneously. Different performance parameters like resonance frequencies, gains, directivities, antenna efficiencies and radiation efficiencies for dual-resonance are observed by varying the dimensions of slot and inserted air-gap. For validation, a prototype of microstrip antenna is fabricated using Roger’s substrate and its performance parameters are measured. Measured results show a very good agreement to their predicted and simulated values.

Index Terms—Slot-size, cross-slotted geometry, inserted air-gap, rectangular microstrip patch, neural networks, synthesis model.

I. INTRODUCTION

Microstrip antennas (MSAs) are proved to be the excellent radiators in different arena of wireless communications like satellite communication, radar communication, global positioning system (GPS) and many more applications. Because of their operation in dual-frequency mode, the MSAs have eliminated two single-frequency operated antennas in these applications. In spite of having many attractive features like low profile, conformable to planar and non-planar surfaces, low fabrication cost etc., the MSAs suffer from the drawbacks of poor radiation characteristics (narrow bandwidth, low gain, low efficiency etc.) which require more attention [1]. Different researchers have proposed different techniques for designing the MSAs for dual-resonance such as multi-layered stacked patch [2]–[3], slotted rectangular patch [4], square patch with notches [5], patch loaded with shorting posts [6] or varactor diodes [7] and rectangular patch fed by an inclined slot [8]. These techniques [2]–[8] can be divided into analytical methods and numerical methods. The analytical methods offer comprehensive details for the operation of MSAs and are based on the physical assumptions for simplifying the radiating phenomenon of the MSAs. For this, such approaches are not suitable for the MSAs in which the thickness of the substrate is not very thin. On the other hand, the numerical methods provide fairly good results but only at the cost of mathematical burden in the form of complex integral equations. The choice of test functions and path integrations emerge to be more significant without initial assumptions in the final stage of the numerical results. Also, these approaches require a new solution even for a very small change in the geometry. Thus, the obligation for having a new solution for every minor change in the geometry and the problems associated with the thickness of the substrates in analytical methods lead to complexities and processing cost [9].

In last decade, artificial neural network (ANN) models have acquired remarkable importance in wireless communications. It is so because of the ability and adaptability to learn and generalization features of the ANNs [9]–[12]. The ANN model can be trained by measured, calculated and/or simulated samples. The purpose of training an ANN model is to minimize the error between the reference and the actual output of the

ANN model. The trained neural model predicts the results very fast for every small variation in the geometry both for electrically thin and thick MSAs. Different ANN models [13]–[20] have been proposed for analyzing and/or synthesizing the conventional MSAs. Few neural models [21]–[23] have been proposed for synthesizing the slotted MSAs. A synthesis modeling using ANN for predicting the size of truncated corners of square patch MSAs has been suggested by Wang et al. [24]. Robustillo et al. [25] have designed contoured-beam reflectarray for a EuTELSAT European Coverage using ANN. An ANN modeling has been suggested by Freni et al. for modeling the behavior of the reflectarray reradiating elements [26]. In the literature [10]–[26], a neural networks based synthesis model has not been proposed for predicting the appropriate size of introduced slot on the radiating patch and inserted air-gap between the ground plane and the substrate sheet, simultaneously. It is very much essential for the antenna designers to predict the “slot-size” and inserted “air-gap” for achieving the desired level of performance parameters like $1.5\text{GHz} \leq \text{resonance frequencies} \leq 3.0\text{GHz}$, $6.2\text{dBi} \leq \text{gains} \leq 9.6\text{dBi}$, $6.6\text{dBi} \leq \text{directivities} \leq 9.9\text{dBi}$, $83\% \leq \text{antenna efficiencies} \leq 100\%$ and $85\% \leq \text{radiation efficiencies} \leq 100\%$ for dual-resonance. For this purpose, a simple and innovative neural networks based synthesis modeling is proposed in this letter.

Section II describes the geometry for samples generation. The neural networks based synthesis model is discussed in Section III. Section IV illustrates the experimental results and discussion. A conclusion followed by the references is discussed in section V.

II. PROPOSED GEOMETRY FOR SAMPLES GENERATION

The cross-sectional view of the proposed antenna is shown in Fig. 1 in which the terms GP, S and SP represent ground plane, substrate and slotted patch, respectively.

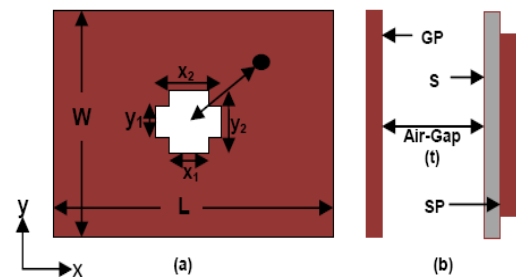


Fig. 1: Proposed Antenna (a) top-view (b) side-view

Here a rectangular patch of dimensions $61\text{mm} \times 56\text{mm}$ is used with a ground-plane of dimensions $200\text{mm} \times 175\text{mm}$. The antenna is simulated in method of moment (MoM) based IE3D software [27] using RT-Duroid substrate RO3003 ($\epsilon_r = 3$, $h = 0.762\text{ mm}$ and $\tan\delta = 0.0045$) for dual-resonance. The proposed antenna is excited by a single probe. Total 1350-samples for dual-resonance (DR), dual-frequency gain (DFG), dual-frequency directivity (DFD), dual-frequency antenna efficiency (DFA) and dual-frequency radiation efficiency (DFR) are generated by varying the slot-size between $1\text{mm} \leq \text{slot-size} \leq 50\text{mm}$ and inserted air-gap between $1\text{mm} \leq \text{air-gap} \leq 10\text{mm}$, simultaneously. These samples are generated in IE3D software on a computer system with system configuration; Dell Optiplex 780 Core 2 Duo CPU E8400, 3.0 GHz with 4.0 GB RAM. These generated samples are used for training as well as testing of the proposed neural networks model which is to be discussed in section III.

The authors are with the department of electronics and communication engineering, Delhi Technological University, Delhi-110 042, India (Corresponding author to provide E-mail: ktaimoor@gmail.com).

III. NEURAL NETWORK MODELING

Artificial neural networks (ANNs) are becoming powerful techniques for obtaining solutions to the problems which are cross-disciplinary in nature. The neural networks are extremely distributed analogous processors that have usual tendency for storing the empirical knowledge during training and making it available for use during testing. ANN models resemble the brain since knowledge is acquired by the neural networks through a training process, and inter-neuron connection strengths are used to store that knowledge [9]. Few known examples of a problem are used to acquire this knowledge during training of the ANN model. The trained model is then set to use this acquired knowledge effectively in solving ‘unknown’ or ‘untrained’ instances of the problems [9]. Multi-layered perceptron (MLP) neural networks consist an input layer, a hidden layer (or a number of hidden layers) and an output layer. Each layer in the model is having entirely different role. Three common steps are used to train the MLP neural networks. Firstly, the training samples are generated, the structural configuration of hidden layer is then selected in the second step and finally, in the third step the weights and biases are optimized using training algorithm. The trained ANN model is then tested on the arbitrary sets of samples which are not included in the training samples. Coding for training and testing algorithms of the ANN model is created in MATLAB software [28].

(A.) Training and Testing Samples Generation

It is a common fact that the neural networks model, trained on a set of samples (known as training samples) produces results very fast. But generation of samples and allocating them into training and testing samples is a challenging task for a complicated geometry where excitation-response is multi-dimensional. In section II, total 1350-samples are generated by varying the slot-size (x_1 , y_1 , x_2 and y_2) and inserted air-gap (t) for the geometry shown in Fig. 1. The sampling used for creating these samples is shown in Table I. Different performance parameters like resonance frequencies (f_1 & f_2), gains (G_1 & G_2), directivities (D_1 & D_2), antenna efficiencies (A_1 & A_2) and radiation efficiencies (R_1 & R_2) for dual-resonance are observed by varying the slot-size and air-gap. Hence, a 10-dimensional response matrix is achieved by varying the 5-dimensional excitation matrix. For ANN modeling, a reverse strategy is used for predicting the 5-dimensional geometrical parameters $[r] \rightarrow [x_1 y_1 x_2 y_2 t]$ for the given 10-dimension performance parameters $[x] \rightarrow [f_1 f_2 G_1 G_2 D_1 D_2 A_1 A_2 R_1 R_2]$.

TABLE I: SAMPLING OF SLOT-SIZE AND INSERTED AIR-GAP

Sampling of Slot-Size and Air-Gap			Generated Samples
Slot-Size	Specified Range	Step-Size	
x_1	$1\text{mm} \leq x_1 \leq 50\text{mm}$	$36\mu\text{m}$	For Training= 1000 Nos. For Testing = 350 Nos.
y_1	$1\text{mm} \leq y_1 \leq 50\text{mm}$	$36\mu\text{m}$	
x_2	$1\text{mm} \leq x_2 \leq 50\text{mm}$	$36\mu\text{m}$	
y_2	$1\text{mm} \leq y_2 \leq 50\text{mm}$	$36\mu\text{m}$	
Air-Gap	Specified Range	Step-Size	
T	$1\text{mm} \leq t \leq 10\text{mm}$	$7\mu\text{m}$	

(B.) Proposed Structure of the Neural Model and Training Algorithms

An ANN model for predicting the “slot-size” and inserted “air-gap” is shown in Fig. 2.

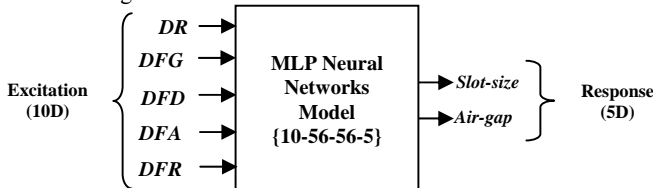


Fig. 2: Synthesis MLP Neural Model.

Training of an ANN model basically, consists of adjusting weights and biases for the applied input sample to get the desired response. This adjustment is carried out by using a training algorithm. The training performance is observed by varying the number of hidden layers as well as the neurons in each hidden layer. After several trials, the structural configuration of the model is optimized as {10-56-56-5} for the best performance. It means that there are 10-neurons in the input layer, 56-

neurons each in first and second hidden layer and 5-neurons in the output layer. Further, the training performance of the model is observed for seven different algorithms; BFGS quasi-Newton (BFG), Bayesian regulation (BR), scaled conjugate-gradient (SCG), Powell-Beale conjugate gradient (CGB), conjugate gradient with Fletcher-Peeves (CGF), one step secant (OSS), and Levenberg-Marquardt (LM), respectively [29]. The samples generated in IE3D simulator, are normalized between +0.1 to +0.9 in MATLAB software before applying training. For an applied input pattern, the arbitrary numbers between 0 and 1 are assigned to initialize the weights and biases. The output of the model is then calculated for that input pattern. The mean square error (MSE) between the predicted and the simulated outcomes is also computed and the weights and biases are then updated accordingly. This updating process is carried out after presenting each set of input sample until the calculated accuracy of the model is estimated satisfactory for 1000-training samples. Training of the model is done by taking some initial parameters: $MSE = 3.45 \times 10^{-6}$, learning rate (η) = 0.631 and momentum coefficient (μ) = 0.518.

The testing algorithm is then realized for the remaining 350-samples. For this realization, all the initial weights and biases are replaced by their corresponding optimized values. The trained model then predicts the “slot-size” introduced on the radiating patch and the inserted “air-gap” between the ground plane and the substrate sheet, simultaneously within a fraction of a second for any arbitrary set of resonance frequencies, gains, directivities, antenna efficiencies and radiation efficiencies for dual-resonance within their specified ranges mentioned in section I.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The training performance of the ANN model discussed in section III, is observed for seven different algorithms; BFG, BR, SCG, CGB, CGF, OSS, and LM algorithms, respectively. The average absolute error in predicting the “slot-size” and inserted “air-gap” both for training and testing of the model is mentioned in Table II. The time elapsed in training of the model for each algorithm is specified in the sixth column of Table II whereas the number of iterations required in the seventh column of same the table. Hence the LM-algorithm is proved to be the most accurate training algorithm for the proposed problem as it produces the least error both for training and testing of the model. It is also proved to be the fastest training algorithm as it requires least training time (1307 sec.) and minimum number of iterations (31739) during training of the ANN model.

TABLE II: PERFORMANCE OF PROPOSED ANN MODEL

Algo.	Average Absolute Error during Training of ANN		Average Absolute Error during Testing of ANN		Training Time	Iteration Required
	Slot-Size	Air-Gap	Slot-Size	Air-Gap		
	mm	mm	mm	mm		
BFG	23.29	13.41	30.12	39.44	3467 sec.	79342
BR	16.25	23.23	27.53	36.36	3269 sec.	72391
SCG	25.46	19.21	29.40	35.49	5392 sec.	76184
CGB	17.59	20.12	25.12	37.18	2485 sec.	76734
CGF	16.04	22.88	29.38	29.41	7226 sec.	78832
OSS	21.41	19.65	24.66	33.39	5638 sec.	83136
LM	7.18	6.99	9.96	8.47	1307 sec.	31739

The weights and biases for the proposed neural model are selected randomly for initialization. Hence, the stochastic behavior (mean and standard deviation) of the five-dimensional absolute error corresponding to five-dimensional response is also calculated. The mean is the sum of the observations divided by the number of observations and identifies the central location of the error set. The standard deviation is the most common measure of variability, measuring the spread of the error set and the relationship of the mean to the rest of the error. The error points closer to the mean indicate fairly uniform responses with small value of the standard deviation. Conversely, the error points far from the mean show a wide variance in the response and large value of the standard deviation. Another term i.e. coefficient of variation (CoV) is also defined to know the relationship between the mean and the standard deviation of an error set. It is defined as the ratio of standard deviation to the mean value. CoV closer to 0, represents greater uniformity of

error whereas CoV closer to 1, represents the large variability of the error. For LM-algorithm, the means for five-dimensional errors are computed as: 1.5061, 1.4927, 1.5217, 1.4956 and 1.4932, respectively whereas their corresponding standard deviations are found as: 0.2910, 0.3062, 0.2824, 0.3229 and 0.3033, respectively. Hence, the coefficients of variations for these five values are coming out to be 0.1932, 0.2051, 0.1856, 0.2159 and 0.2031, respectively. Thus, it is concluded here that all the error points over a full validation set of 1350 samples are uniformly distributed [26].

During simulation in IE3D software, 1.5 GHz to 3.0 GHz frequency range with total 100-sampling points for each simulating structure is used and total 1350-such structures are simulated. The simulation time in the IE3D software depends on the complexity inserted in the geometry. For the proposed geometry, it is computed as ~6700sec. (or ~1h 53min.) per structure. 36MB system memory (RAM) is required during each simulation. By using the ANN modeling, the computational time and the required memory storage is fairly reduced. The time elapsed during training of the ANN model is already computed as 1307 sec. (see Table II). The time elapsed during testing of the ANN model is computed in the MATLAB software on a computer system using MATLAB syntax 'cputime' as mentioned below:

```

clc; clear all;
time_1 = cputime;
Start of testing algorithm
Program statement-1
Program statement-2
.....
End of Testing Algorithm
Computational Time_ANN = cputime-time_1

```

This procedure is repeated for 350-independent runs and finally concluded that the ANN model requires ~53msec. in producing the results after training. Thus, the neural approach after training is much faster than the simulation approach.

Further, the training of the model requires only 29KB RAM of a system and for testing the performance, only 1.4KB RAM is required. Hence, the required memory space in training as well as testing of the model is also lesser as compared to 36MB required for simulation.

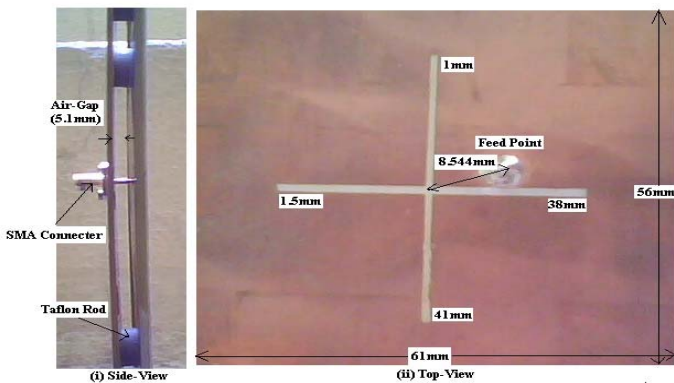


Fig. 3: Enlarged view of fabricated antenna.

For validating the work, a prototype is fabricated using RT-Duroid substrate. The patch of dimensions 61mm×56mm is etched on upper side of the substrate whereas an air-gap of 5.1mm between the substrate and the ground plane is inserted using Teflon rods (see Fig. 3). The prototype is excited by probe feed which guides the electromagnetic waves to the feed point. An SMA connector with 6.8mm long pin (1mm for ground plane+5.1mm for air-gap+0.762mm for substrate) is used for RF connection. The S-parameters of the fabricated prototype are measured using Agilent N5230A network analyzer. A comparison between measured and simulated S-parameters is depicted in Fig. 4 which shows a good convergence between the two. It is also confirmed here that the Taflon rods do not affect the antenna performance. A comparison between simulated and predicted parameters for the

fabricated prototype is given in Table III which shows a good agreement between the two. A very good conformity is also achieved if the simulated dual-resonance (DR) is compared with its measured counterparts as mentioned in Table IV.

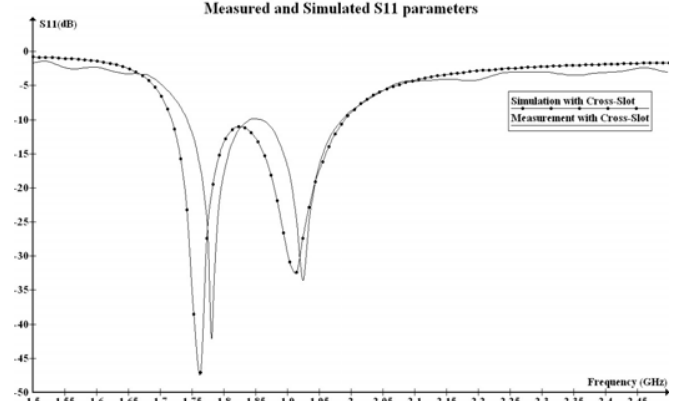


Fig. 4: Comparison of Measured and Simulated S-Parameters.

TABLE III: COMPARISON OF SIMULATED AND PREDICTED RESULTS

Name of Parameter	Simulated value in IE3D	Predicted value using ANN
x_1 (mm)	38.0000	37.9209
y_1 (mm)	1.5000	1.5021
x_2 (mm)	1.0000	0.9567
y_2 (mm)	41.0000	41.0561
t (mm)	5.1000	5.1023

TABLE IV: COMPARISON OF SIMULATED AND MEASURED DR

Specifications of slot-size and inserted air-gap	Simulated DR in IE3D	Measured DR
Slot-size: $x_1 = 38.0$ mm, $y_1 = 1.5$ mm, and $x_2 = 1.0$ mm & $y_2 = 41.0$ mm	1.7708 GHz and 1.9221 GHz	1.7800 GHz and 1.9250 GHz
Air-gap: 5.1mm		

Table V shows the improvement in the performance parameters by inserting the cross-slot on the patch antenna. By introducing the slot, the excited surface current path lengthens, increasing the antenna length, hence decreasing the resonance frequency. For the case of slotted patch, both the resonance frequencies (1.763GHz and 1.914GHz) are lowered by ~13% as compared to those (2.0256GHz and 2.1970GHz) of the patch geometry without slot which can result to a patch size reduction of ~35% for a given dual-frequency design. Hence, a good rank of compactness is achieved in the fabricated prototype. Further, the gains and efficiencies are also improved but only at the cost of slight reduction in the directivities. The gain of an antenna is defined as the product of directivity and radiation efficiency. Hence, it is improved by increasing the radiation efficiencies.

TABLE V: IMPROVEMENT IN PERFORMANCE PARAMETERS

Performance Parameter	Without cross-slot	With cross-slot	Remarks
DR (GHz)	2.0256 and 2.1970	1.7630 and 1.9140	Desired
DFG (dBi)	8.2026 and 8.2129	8.9365 and 8.9387	Desired
DFD (dBi)	9.1428 and 9.4193	9.1122 and 9.2079	Un-desired
DFA (%)	80.5329 and 85.3708	95.8065 and 93.9897	Desired
DFR (%)	90.2454 and 93.7718	96.3604 and 96.6599	Desired

A bandwidth of 55.5MHz is achieved (see Fig. 4) for $S_{11} \leq -10$ dB during simulation but it increases to 250MHz (1.99GHz-1.74GHz) in the fabricated prototype during measurement. It may be probably due to irregularity in fabricating the prototype and/or inserting the air-gap using Teflon rods. Further, the size and losses of the solder joints are not considered during the simulation, which may cause this difference between them.

The ratio (f_2/f_1) of the two frequencies is also computed as 1.0846 (= 2.1970/2.0256), 1.0857 (=1.914/1.763) and 1.0815 (=1.925/1.780) for the original geometry without any slot, simulated slotted geometry and fabricated slotted prototype, respectively. Thus, the ratio (f_2/f_1) is slightly affected by inserting the cross-slot, and mainly determined by the aspect ratio ($1.89 = 61/56$) of the patch; which makes the design simpler and easier to implement.

An extremely good impedance matching is also achieved by inserting the air-gap between the substrate sheet and the ground plane. Consequently, the antenna and radiation efficiencies both significantly are improved because of very small amount of dielectric loss in the air-substrate.

V. CONCLUSION

In this letter, an accurate, simple and fast ANN based synthesis modeling scheme has been proposed to predict the required "slot-size" and inserted "air-gap", simultaneously in order to achieve desired level of radiation patterns of rectangular microstrip antennas. Such a neural approach is rarely proposed in the open literature to the best of our knowledge for the ten-dimensional excitation and five-dimensional response. It takes only a fraction of a second in predicting the required results after doing training properly. As the approach has produced more accurate results, hence it can also be recommended to include in antenna computer-aided-design (CAD) algorithms.

The computational time and required system memory have been calculated and concluded that the neural approach requires lesser amount of computational time and the system memory after training.

In general, computing 5-dimensional response may require 5-different neural networks modules, whereas in the present work, only one module is fulfilling the requirement of 5-independent modules. During synthesizing the antenna, it is desirable for the design engineers to know different performance parameters of an antenna, simultaneously instead of knowing individual parameters, alternatively. Hence, the present approach has been considered more generalize and efficient.

For validation purpose, a prototype has also been fabricated. A very good convergence between measured and simulated results has been achieved which supports the effectiveness of the proposed work.

REFERENCES

- [1] I. J. Bahl and P. Bhartia, "Microstrip antennas", Artech House, Dedham, MA, 1980.
- [2] J. S. Dahele, K. F. Lee, and D. P. Wong, "Dual-frequency stacked annular-ring microstrip antennas," *IEEE Trans on Antenna Propag*, vol. 35, pp.1281-1285, 1987.
- [3] S. A. Long and M. D. Walton, "A dual-frequency stacked circular-disk antenna," *IEEE Trans on Antennas Propag*, vol. 27, pp. 270-273, 1979.
- [4] S. Maci, G. B. Gentili, and G. Avitabile, "Single-layer dual-frequency patch antenna," *Electron. Lett.*, vol. 29, pp.1441-1443, 1993.
- [5] H. Nakano and K. Vichien, "Dual-frequency square patch antenna with rectangular notch," *Electron. Lett.*, vol. 25, pp.1067-1068, 1989.
- [6] D. H. Schaubert, F. G. Farrar, A. Sindoris, and S. T. Hayes, "Microstrip antennas with frequency agility and polarization diversity," *IEEE Trans on Antennas Propagation*, vol. 29, pp.118-123, 1981.
- [7] R. B. Waterhouse and N. V. Shuley, "Dual-frequency microstrip rectangular patches," *Electron. Lett.*, vol. 28, 1992.
- [8] Y. M. M. Antar, A. I. Ittipiboon, and A. K. Bhattacharyya, "A dual-frequency antenna using a single patch and an inclined slot," *Microwave and Opt. Technol Lett.*, vol. 8, pp. 309-311, 1995.
- [9] Q.J. Zhang and K.C. Gupta, *Neural Networks for RF and Microwave Design*, Artech House Publishers, 2000.
- [10] P.M. Watson and K.C. Gupta, "EM-ANN models for microstrip vias and interconnects in database circuits," *IEEE Trans Microw Theory Tech*, vol. 44, pp.2495-2503, 1996.
- [11] P.M. Watson and K.C. Gupta, "Design and optimization of CPW circuits using EM-ANN models for CPW components," *IEEE Trans on Microw Theory Tech*, vol. 45, pp.2515-2523, 1997.
- [12] P.M. Watson, K.C. Gupta and R.L. Mahajan "Development of knowledge based artificial neural network models for microwave components," *Microwave Symposium Digest*, 1998 IEEE MTT-S International, vol. 1, pp.9-12, 1998.
- [13] D. Karaboga, K. Guney, S. Sagiroglu and M. Erler, "Neural computation of resonance frequency of electrically thin and thick rectangular microstrip antennas," *Microwave, Antennas and Propagation*, 1999 IEE Proceedings, vol. 146, pp.155-159, 1999.
- [14] Seref Sagiroglu, Karim Guney and Mehmet Erler, "Resonant frequency calculation for circular microstrip antennas using artificial neural networks," *International Journal of RF, Microw and CAE*, vol. 8, pp. 270-277, 1998.
- [15] S. Sagiroglu and K. Guney, "Calculation of resonant frequency for an equilateral triangular microstrip antenna with the use of artificial neural networks," *Microw and Opt Technol Lett*, vol. 14, pp.89-93, 1997.
- [16] R. Gopalakrishnan and N. Gunasekaran, "Design of equilateral triangular microstrip antennas using artificial neural networks," *Antenna Technology: Small Antennas and Novel Metamaterials*, 2005. IWAT 2005. *IEEE International Workshops on*, pp. 246-249, 2005.
- [17] Karim Guney, Seref Sagiroglu and Mehmet Erler, "Generalized neural method to determine resonant frequencies of various microstrip antennas," *International Journal of RF and Microw CAE*, vol. 12, pp.131-139, 2002.
- [18] K. Guney and N. Sarikaya, "A hybrid method based on combining artificial neural network and fuzzy interference system for simultaneous computation of resonance frequencies of rectangular, circular, and triangular microstrip antennas," *IEEE Trans on Antenna Propag*, vol. 55, pp. 659-668, 2007.
- [19] K. Guney and N. Sarikaya, "Concurrent neuro-fuzzy systems for resonant frequency computation of rectangular, circular, and triangular microstrip antennas," *Progress in Electromagnetics Research*, vol. 84, pp. 253-277, 2008.
- [20] Nurhan Turker, Filiz Gunes, and Tulay Yildirim, "Artificial neural design of microstrip antennas," *Turk J Elec. Engin*, vol. 14, 2006.
- [21] D. K. Neog, S. S. Pattnaik, D. C. Panda, S. Devi, B. Khuntia, and M. Dutta, "Design of a Wideband Microstrip Antenna and the use of Artificial Neural Networks in Parameter Calculation" *Antenna Propag Magazine, IEEE*, vol. 47, pp.60-65, 2005.
- [22] V. V. Thakare and P. K. Singhal, "Bandwidth analysis by introducing slots in microstrip antenna design using ANN" *Progress In Electromagnetics Research M*, vol. 9, pp.107-122, 2009.
- [23] Vandana Vikas Thakare and Pramod Singhal, "Microstrip antenna design using artificial neural networks", *Int J RF and Microwave CAE*, vol. 20, pp.76-86, 2010.
- [24] Zhongbao Wang, Shaojun Fang, Qiang Wang, and Hongmei Liu, "An ANN-based synthesis model for the single-feed circularly-polarized square microstrip antenna with truncated corners" *IEEE Trans on Antennas Propag*, vol. 60, pp. 5989-5992, 2012.
- [25] P. Robustillo, J. Zapata, J.A. Encinar and M. Arrebola, "Design of a Contoured-Beam Reflectarray for a EuTELSAT European Coverage using a Stacked-Patch Element Characterized by an Artificial Neural Network," *IEEE Antennas wireless Propag. Letters*, vol.11, pp.977-980, 2012
- [26] Angelo Freni, Marco Mussetta and Paola Pirinoli, "Neural Network Characterization of Reflectarray Antennas," *International Journal of Antennas and Propagation*, vol. 2012, Article ID 541354.
- [27] IE3D Version 14.0, Fremont, CA, Zeland Software Inc., Oct. 2007.
- [28] D. J. Higham, and N. J. Higham, *MATLAB Guide*, SIAM, PA, 2005.
- [29] Martin T. Hagan and Mohammad B. Menhaj, "Training feed forward networks with the Marquardt algorithms," *IEEE Trans. Neural Networks*, vol. 5, pp. 989-993, 1994.

STACKED PATCH ANTENNA WITH CROSS SLOT ELECTRONIC BAND GAP STRUCTURE

N.S. Raghava¹, Asok De², Nitish Kataria³, Sarthak Chatterjee⁴

^{1,2} Dept. Of Electronics and Communication Engg., Delhi Technological University, Delhi, India

^{3,4} Dept. Of Electronics and Communication Engg., GPMCE, IP University, Delhi, India

ABSTRACT

A cross slotted electronic band gap (EBG) with stacked rectangular patches shorted with a shorting pin is proposed in this paper. The study is being done on how the various parameters are varied by changing the probe feed location. The design is constructed by using stacking of patches, shorting pin and cross slotted EBG to form an optimized antenna design with antenna efficiency of approximately 99.06%. The radiation patterns are given at 2.586 GHz which can be used for wireless communications.

Keywords - Electronic band gap (EBG), shorting pin, reflector plane, stacking, antenna efficiency, gain and directivity.

1. INTRODUCTION

As the wireless mobile communication technologies is growing very rapidly, microstrip antenna have various applications because of their small size, low weight, low cost, low planar configuration and can be easily integrated with other microwave circuits. They are very compact and wideband antennas for the transmission of video, voice and data information. The various applications for microstrip antennas are wireless communication systems, satellite communication systems, cellular phones, pagers, radars, etc.

The 2.6 GHz band is generally considered to cover the frequency range between 2500-2690MHz, although there are some minor national variations. The band provides an opportunity to meet rapidly rising demand for capacity to deliver mobile broadband services on widespread, common basis across the world. The applications of the frequencies in this band, which vary by country and region, includes satellite services, fixed mobile and broadcast as well as terrestrial and video broadcasting in certain countries. In this paper, the frequency of 2.586GHz is used which can be implemented in GPRS, WLAN, mobile broadcasting and road vehicle communications (Raghava and Asok De, 2006).

Small antenna design is always compromised between size, bandwidth and efficiency. The bandwidth can be increased with the use of stacked patch structure in a microstrip antenna (MSA) (Ollikainen, Fischer and Vainikainen, 1999) design whereas the size of the antenna can be decreased by the use of shorting pin. The stacked structure can be shorted with the use of shorting pin and the overall size of the antenna is reduced drastically (Sanad, 1994). The radiation characteristics can be further improved by an electronic band gap structure on one side of the substrate. This reduces the surface waves induced in the antenna and also increases the radiation pattern (Gonzalo, MAagt, Sorolla, 1999). The use of the reflector plane is fed on the finite size ground plane at the rear end of the antenna to reduce the level of back radiations (Raghava and Asok De, 2009).

In this paper, a stacked rectangular microstrip antenna using a shorting post backed with Cross Slot Electronic Band Gap (CSEBG) structure is studied. The effect on antenna efficiency and gain is also compared by changing the probe feed location while the shorting post remained fixed.

2. ANTENNA DESIGN

A stacked rectangular microstrip antenna using a shorting post backed with Cross Slot EBG structure (CSEBG) is designed. The top view of typical EBG stacked rectangular microstrip antenna with a shorting post is shown in Fig. 1. The antenna consists of two dielectric layers with stacked rectangular patches of dimensions (16 x 16) mm. The thickness of the upper dielectric (i) having dielectric constant of air is 2.6mm and that of the lower one (ii) having dielectric constant 2.6 is 1.6mm. The square patches used are mounted

over the substrate creating a stacked structure. The stacking of patches increases the antenna efficiency and gain of the microstrip antenna. The 3-D view of CSEBG MSA is given in Fig. 2.

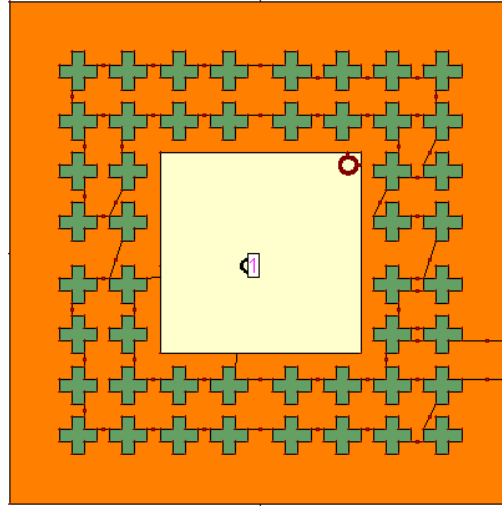


Fig. 1 Top View of CSEBG MSA

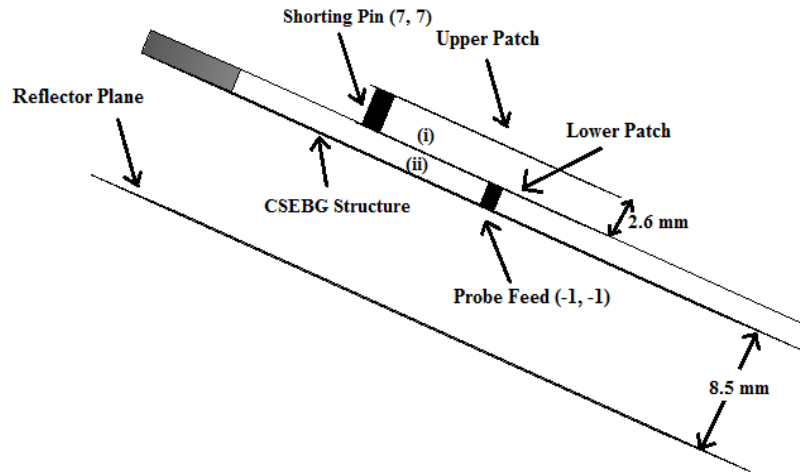


Fig. 2 3-D view of CSEBG MSA

The EBG structure is constructed by etching cross shaped slots on a square patch of size (40 x 40) mm. The cross shaped slots of dimensions (3 x 1) mm which are spaced 1 mm apart forms an 8 x 8 matrix. The distance from the edge is 4mm and there are no slots beneath the patch area. The EBG structure is used for better radiation characteristics.

This antenna is excited by a coaxial probe feed of diameter 1mm given at location (-1, -1). The two rectangular patches are shorted by a shorting post of diameter 1.3mm at location (7, 7). The use of shorting post decreases the size of the antenna (M.Sanad, 1994). A reflector plane is placed at a distance of 8.5mm behind the ground plane as given in Fig. 2. The reflector plane is used to reduce the back lobes generated by the antenna. In this paper, the various parameters like antenna efficiency, gain and directivity are improved by using stacked rectangular patches and shorting them with shorting post backed with cross slotted EBG

structure (CSEBG). These parameters are also compared at different probe feed locations. The proposed CSEBG MSA at probe feed location (-1, -1) with antenna efficiency of 99.06 % and gain 4.75 dB is able to work at 2.586 GHz frequency and can be used for various wireless communication application.

4. RESULTS AND DISCUSSIONS

The CSEBG MSA is simulated using IE3D simulator which is based on the method of moments (MoM) (Zeland, ver. 14.0, 2007). We have done a comparison of antennas various parameters at different locations of probe feed. The table 1 shows that the antenna efficiency and gain for CSEBG MSA at location (-1,-1) are approximately 99.06% and 4.75 dB which are highest with respect to others. The antenna parameters like antenna efficiency and gain show strange behaviour when the probe feed location is near shorting post as the parameters decrease drastically as we approach it. The 3-D radiation pattern for CSEBG MSA is shown in Fig. 3. The Fig. 4 shows the graph between frequency and return loss (S11) and it can be seen that highest negative return loss -19.96 is at frequency 2.586 GHz. The resonant frequency of CSEBG MSA is 2.586 GHz which has various applications in wireless and road vehicle communications.

Table. 1 Comparison at different probe feed locations

S.No.	Location Of Probe Feed	Frequency (GHz)	Radiation Efficiency (%)	Antenna Efficiency (%)	Gain (dB)	Directivity (dB)	Return Loss (S11)
1	(-7,-7)	2.616	100	97.1155	4.5495	4.67658	-15.29
2	(-6,-6)	2.606	100	96.9724	4.5354	4.66893	-15.22
3	(-5,-5)	2.606	100	97.337	4.5748	4.69197	-15.76
4	(-4,-4)	2.606	100	97.7917	4.6228	4.71975	-16.53
5	(-3,-3)	2.6	100	98.3209	4.6612	4.73471	-17.75
6	(-2,-2)	2.586	100	98.8704	4.7254	4.77476	-19.64
7	(-1,-1)	2.586	100	99.0584	4.7488	4.78993	-19.96
8.	(1,1)	2.556	100	98.0513	4.7814	4.86684	-16.81
9	(2,2)	2.606	100	87.5899	4.2951	4.87054	-9.056
10	(3,3)	2.6	100	73.3518	3.5751	4.92096	-5.743
11	(4,4)	2.596	100	55.5322	2.4456	5.00016	-3.521
12	(5,5)	2.576	100	11.6286	-4.204	5.14061	-0.5345

5. CONCLUSIONS

The cross slotted EBG with stacked rectangular microstrip antenna (CSEBG MSA) is proposed in this paper. The antenna efficiency and gain are 99.06 % and 4.75 dB respectively. The probe feed locations are changed along the diagonal. The different values of antenna parameters are compared at different probe feed locations. The highest value of antenna efficiency and gain is at location (-1, 1) but as we approach towards shorting post their values decrease drastically. The use of stacked structure and shorting them with shorting pin also increases the gain and efficiency of the antenna. For the calculation of radiation efficiency no losses are taken into account, whereas in the case of antenna efficiency losses due to mismatch, ohmic and dielectric etc. are taken into account. With good antenna efficiency and gain of the proposed antenna at frequency 2.586 GHz, it can be used for GPRS, WLAN and road vehicle communications.

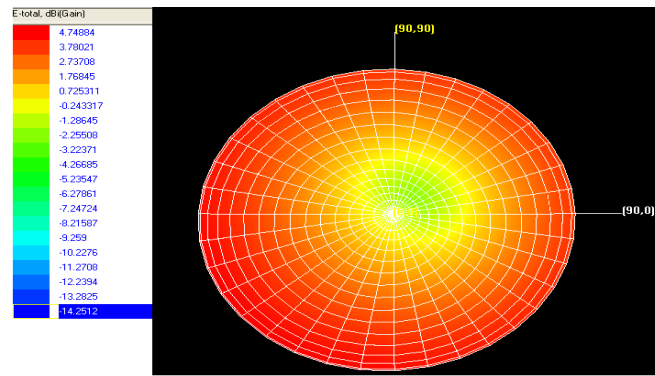


Fig. 3 3-D Radiation Pattern for CSEBG MSA

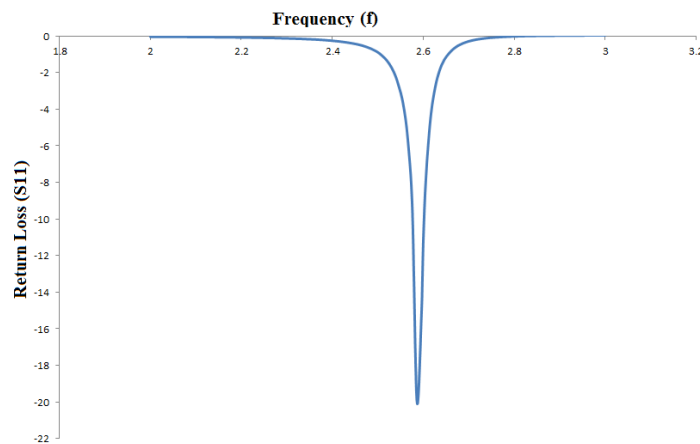


Fig. 4 Plot between return loss (S_{11}) and frequency of CSEBG MSA

6. REFERENCES

IE3D version 14.0 Zeland Software Inc, Fremont CA, 2007.

J.Ollikainen, M. Fischer and P.Vainikainen,"Thin dual-resonant stacked shorted patch antenna for mobile communications," Electron. Lett. March 1999, Volume 35,(6), pp. 437-438.

M.Sanad, "Effect of shorting posts on short circuited microstrip antennas, " Proceedings,IEEE Antennas and Propagation Society International Symposium,pp. 794-797,1994.

N S Raghava and Asok De, "Effect of air gap width on the performance of a Stacked Electronic Band Gap Antenna " , IJMOT, VOL 4,NO. 5, Sep. 2009.

N S Raghava and Asok De,"Photonic Bandgap Stacked Rectangular Microstrip Antenna for Road Vehicle Communication," IEEE Trans. Antennas and Wireless Propagation Letters, Volume 5, pp.421-423. Dec.2006.

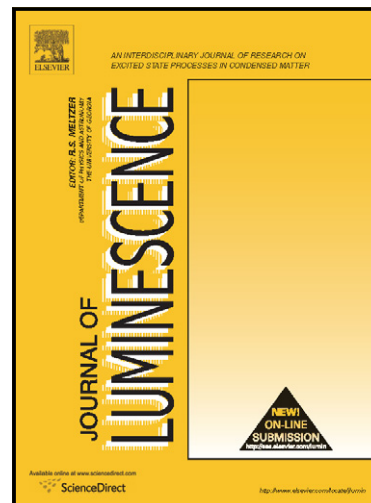
Romon Gonzalo, Peter de MAagt, Mario Sorolla," Enhances Patch-Aantenna Performance by Suppressing Surface Waves Using Photonic-Bandgap Substrates," IEEE Trans. on Microwave Theory and Techniques, Vol.47, No. 11, Nov 1999, pp.2131-2138.

Scott fox,Martyn Roetter,Jeffrey Walkenhorst,Peter Rysang, "The 2.6 GHz Spectrum Band" , Report prepared for the GSM association,Dec 2009.

Author's Accepted Manuscript

Visible Luminescence Characteristics of Sm^{3+} doped Zinc Alumino Bismuth Borate Glasses

K. Swapna, Sk. Mahamuda, A. Srinivasa Rao, T. Sasikala, L. Rama Moorthy



www.elsevier.com/locate/jlumin

PII: S0022-2313(13)00599-1
DOI: <http://dx.doi.org/10.1016/j.jlumin.2013.09.035>
Reference: LUMIN12187

To appear in: *Journal of Luminescence*

Received date: 5 June 2013
Revised date: 11 August 2013
Accepted date: 17 September 2013

Cite this article as: K. Swapna, Sk. Mahamuda, A. Srinivasa Rao, T. Sasikala, L. Rama Moorthy, Visible Luminescence Characteristics of Sm^{3+} doped Zinc Alumino Bismuth Borate Glasses, *Journal of Luminescence*, <http://dx.doi.org/10.1016/j.jlumin.2013.09.035>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Visible Luminescence Characteristics of Sm³⁺ doped Zinc Alumino Bismuth Borate Glasses

K.Swapna¹, Sk.Mahamuda¹, A. Srinivasa Rao^{1,2}, T.Sasikala³ and L. Rama Moorthy³*

¹*Department of Physics, K L University, Green Fields, Vaddeswaram 522 502, A P, India.*

²*Department of Applied Physics, Delhi Technological University, Delhi-110 042, India.*

³*Department of Physics, S V University, Tirupathi- 517 502, A P, India.*

Abstract

Zinc Alumino Bismuth Borate (ZnAlBiB) Glasses at different compositions doped with 1 mol% of Sm³⁺ ions were prepared by using melt quenching technique and investigated by XRD, optical absorption, emission and decay curve analysis to understand the visible luminescence of these glasses. From the absorption spectra the JO parameters are evaluated and are used to calculate the radiative properties such as transition probability, radiative lifetime, branching ratio and absorption cross-sections for various fluorescent levels of Sm³⁺ ions. The emission spectra of Sm³⁺ ions doped ZnAlBiB glasses show two intense emission bands $^4G_{5/2} \rightarrow ^6H_{7/2}$ (orange), $^4G_{5/2} \rightarrow ^6H_{9/2}$ (red) for which the stimulated emission cross-section and branching ratios are evaluated to understand the potentiality of these materials as visible lasers. The decay profiles for the $^4G_{5/2}$ fluorescent level of Sm³⁺ doped ZnAlBiB glasses have been recorded to measure the quantum efficiency (η) of these glasses. The strong visible emissions, large stimulated emission cross-sections, high branching ratios and good quantum efficiencies observed for the present ZnAlBiB glasses suggests the suitability of these glasses as laser and photonic devices operating in visible region.

***Corresponding author:**

Email: drsrallam@yahoo.co.in., Tel: +918586039007.

1. Introduction

Optical properties of rare-earth (RE) doped glasses are extensively studied for their potential applications in the fields of lasers, fluorescent display devices, optical detectors, wave guides and fibre amplifiers [1-4]. The versatility of glasses, the broader emission and absorption spectra they provide when compared with crystalline hosts, have increasing interest from the perspective of spectroscopic [5] as well as technological point of view [6,7]. The excitation and emissions spectra of RE ions in glasses are due to 4f electronic transitions, which are sensitive to the symmetry, structure of the local environment and on phonon energies of the host glass. The design and development of an optical device for the above applications require a thorough investigation of optical absorption and luminescence of RE doped glasses. Such studies give the information pertaining to lifetimes and quantum efficiency of rare-earth ions. In order to optimise the efficiency of rare-earth doped glasses, the local environment around RE ions have to be changed by changing the glass composition or by changing the concentration of RE ions in a host glass. The ligand fields produced by a host matrix will have considerable influence on the absorption, emission and quantum efficiency of the doped ions. In general a host glass with low phonon energies decreases non-radiative relaxation rates and gives high quantum efficiencies useful for obtaining efficient lasers and optical fibre amplifiers [8].

It is well known that Boric acid (B_2O_3) is one of the good glass formers and can form glass alone with good transparency, high chemical durability, thermal stability and good rare-earth ion solubility [9]. However a glass with B_2O_3 alone will possess high phonon energies ($\sim 1300\text{ cm}^{-1}$) which cannot suppress the non-radiative decay process and hence rare-earth ion emissions are strongly reduced. Addition of heavy metal oxides to B_2O_3 can reduce phonon energies considerably and can give intense fluorescence suitable for lasers and fibre

amplifiers [8]. A heavy metal oxide Bi_2O_3 besides being a higher valence oxide, low field strength and high polarizability cannot form glass alone, however in presence of a good glass former like B_2O_3 , it give a glass with less phonon energies. [10,11]. Moreover, Bismuth containing glasses can exhibit broad band luminescence in the near-infrared (NIR) region, which finds applications in the fields of telecommunications, tuneable fibre lasers and spectral converters [12]. In recent years glasses containing ZnO have attracted the scientific community because of its role in increasing chemical stability and to reduce the thermal expansion of the glass. Its wide band gap, large exciton binding energy and intrinsic emitting property makes them as promising candidates for the development of optoelectronic devices, solar energy concentrators, ultraviolet emitting lasers and gas sensors [13]. Presence of Al_2O_3 in a glass can increase mechanical and thermal stability of the glass. Hence for the present work we prepared a glass with a glass former B_2O_3 by adding ZnO, Al_2O_3 and a heavy metal oxide such as Bi_2O_3 at different compositions, keeping the concentration of Sm^{3+} ion content constant.

Spectroscopic studies of Sm^{3+} ions have been extensively investigated to characterize them for optical device applications [14-16]. The $^4\text{G}_{5/2}$ excited level of Sm^{3+} ions emit in visible region through different emission channels with relatively high quantum efficiency [17]. This makes the Sm^{3+} ions an interesting case to analyze the energy transfer process. Sm^{3+} ion with $4f^5$ configuration exhibit a strong orange-red fluorescence in the visible region and having applications in high density optical storage, underwater communication, colour displays and visible solid state lasers [18]. Although the trivalent samarium ion exhibits a strong orange luminescence in the visible region, the complicated electronic structure of the $4f^5$ configuration is responsible for their neglect. A very large number of energy levels lying close to each other render the interpretation of the absorption spectra rather cumbersome. The

fluorescence intensity of Sm^{3+} ion depends on the concentration of RE ions as well as on the glass network.

The main aim of the present work is to find out a suitable glass composition for Sm^{3+} ions for better luminescence efficiency and to ascertain their utility for solid state laser devices. The present study mainly includes absorption, emission spectral studies and lifetime measurement of Sm^{3+} ions in Zinc Alumino Bismuth Borate (ZnAlBiB) glasses at different compositions of the glass, keeping the concentration of Sm^{3+} constant at 1 mole%. From the absorption, emission and decay curve measurements, various spectral properties such as oscillator strengths, JO parameters, radiative properties such as transition probability, absorption cross-section, branching ratio's, lifetimes, emission cross-sections and quantum efficiency for the excited levels of Sm^{3+} ions are estimated by using Judd-Ofelt theory.

2. Experimental

Sm^{3+} ions doped Zinc Alumino Bismuth Borate (ZnAlBiB) glasses were prepared with the chemical composition $60 \text{ B}_2\text{O}_3 + 9 \text{ ZnO} + (30-x) \text{ Al}_2\text{O}_3 + x \text{ Bi}_2\text{O}_3 + 1 \text{ Sm}_2\text{O}_3$ (where $x=5, 10, 15$ and 20) using melt quenching technique. For convenience these glasses are designated as glass A to D in accordance with the x value. High purity chemicals were mixed thoroughly, grinded in an agate motor, to get homogeneous powder and was placed in an aluminium crucible at 1200°C in an electric furnace. The liquid was then air quenched by pouring it on a thick brass plate and pressed with another brass plate. The glass samples were annealed at 400°C for 1h to get thermal and structural stability. These samples were well polished to get uniform thickness before going to physical and optical measurements. The densities were measured by the Archimedes method using water as an immersion liquid. The refractive indices of these glasses are measured by using the Brewster angle method using He-Ne laser (650 nm line) as the source. The optical absorption spectra of the samples were

recorded at room temperature in the spectral wavelength range covering 1100-1900 nm with a spectral resolution of 0.1nm using JASCO model V-670 UV-VIS-NIR spectrophotometer. The luminescence spectra and lifetime measurements were carried out at room temperature using JOBIN YVON Fluorolog-3 spectrofluorimeter with Xenon arc lamp as radiation source.

3. Results and Discussion

3.1. XRD Spectral measurements

The recorded XRD profile of an un doped ZnAlBiB glass gives a hump without any sharp peaks as shown in Fig.1. Such spectral profile with hump confirms the amorphous nature of the prepared glasses.

3.2. Physical Properties

From the practically measured densities and refractive index, the other related properties such as average molecular weight (M), mean atomic volume (V), molar refractivity (R_M), molecular electronic polarizability (α_e), rare-earth ion concentration (N) and inter ionic distance have been computed by using the relevant expressions available in literature [19] are presented in Table 1. From Table 1, it can be observed that, from glass A to D as Bi_2O_3 content increases the refractive index decreases, whereas the density of the glass, molar refraction, polaron radius, interionic distance and molecular electronic polarizability increases. From Table 1, it is also evident that the optical basicity value which indicate the ability of a glass to donate negative charges to the probe ion decreases from glass A to D.

3.3 Absorption Spectral Analysis

Fig.2 shows the absorption spectrum of 1 mol% of Sm^{3+} doped glass A recorded in the wavelength range 1100-1900 nm. The absorption spectra for the remaining glasses are quite

similar to that of glass A with slight difference in intensities of various bands and hence they were not given in the manuscript. Each absorption spectrum consists of six absorption bands in NIR region which originate from the ground state ${}^6H_{5/2}$. Due to strong absorption of the host glass in UV-VIS region, some of the absorption bands of Sm^{3+} in the UV-VIS region were disappeared. The experimental oscillator strength (f_{exp}), which is a measure of the strength of an absorption transitions in a given matrix, has been determined from the area under the absorption curves using the formula given in the literature [20,21] and such oscillator strengths for all the glasses are given in Table 2. In case of Sm^{3+} ion with $4f^5$ electronic configuration, the absorption band corresponding to ${}^6H_{5/2} \rightarrow {}^6F_{1/2}$ transition is identified as hypersensitive transition. From Table 2 it can be seen that, f_{exp} values corresponding to this hypersensitive transition is smaller than the other transitions. This might be due to the overlapping of ${}^6H_{5/2} \rightarrow {}^6F_{3/2}$ and ${}^6H_{5/2} \rightarrow {}^6H_{15/2}$ transitions.

The Judd-Ofelt theory has been applied to the experimentally evaluated oscillator strengths to find the JO intensity parameters (Ω_2 , Ω_4 and Ω_6) by least square fit analysis. Utilization of the JO theory to evaluate the JO intensity parameters were discussed in detailed in our previous paper [22]. The JO parameters thus evaluated are used to calculate the oscillator strengths (f_{cal}) and such calculated oscillator strengths are also given in Table 2 along with f_{exp} . The small rms deviation values obtained between calculated and experimental oscillator strengths indicates the good fit between the two values and also the validity of JO theory. The JO parameters obtained for the present glasses are given in Table 3 along with the values reported for the other glasses. From Table 3 it can be observed that the JO parameters reported for the present glasses are in consistent with the values reported for other glasses. The JO parameters are found to be following the same trend ($\Omega_6 > \Omega_4 > \Omega_2$) in all ZnAlBiB glasses. The JO parameters play an important role in studying the coordination

environment around the trivalent rare-earth ions and also on the covalency between the rare-earth and oxygen ions in any host glass. In general, the Ω_2 parameter is related to the covalency, structural change and symmetry of ligand field around the Sm^{3+} ion site [23]. On the other hand, the magnitudes of Ω_4 and Ω_6 are related to the bulk properties such as viscosity and rigidity of the medium in which the ions are situated. The lower value of Ω_2 observed for all these glasses indicates that the symmetry of the ligand field at the rare-earth site is higher [24]. The larger size of the modifier ion causes the average Sm-O distance to increase and such increase in the bond length produces a weaker field around the Sm^{3+} ions leading to a lower value of Ω_2 for the present glasses. The covalency between the Sm^{3+} ion and the ligand oxygen ion also contributes to Ω_2 [25]. The Ω_2 parameter exhibits the dependency of the covalence between Sm^{3+} ions and ligand ions, because Ω_2 is connected to the asymmetry of the local environment around the Sm^{3+} sites. The lesser the values of Ω_2 , the more Centro symmetrical the ion site and the more ionic its chemical bond with the ligands. Jacob and Weber [26] reported that the magnitude of spectroscopy quality factor ($\chi = \Omega_4/\Omega_6$) is used to characterize the stimulated emission in any host glass matrix. The reasonably higher values of spectroscopic quality factor observed for the present glasses predicts efficient stimulated emission in the present glasses, and is comparable to the other glasses [27-32].

To know the nature of bonding between the Sm^{3+} ions and the ligands, the nephelauxetic ratios and bonding parameters have been evaluated for all the ZnAlBiB glasses using the absorption spectral data. The nephelauxetic ration (β) is given by [33]

$$\beta = \frac{\gamma_a}{\gamma_b} \quad (1)$$

Where γ_a and γ_b are the energies (in cm^{-1}) of the corresponding transitions in the complex and the free ion or aqua ion respectively. The value of β for all the observed transitions are

computed and their average value of $\bar{\beta}$ is used to estimate the bonding parameter δ using the relation

$$\delta\% = \left[\frac{(1-\bar{\beta})}{\bar{\beta}} \right] \times 100 \quad (2)$$

The bonding will be covalent or ionic depending upon the positive or negative sign of $\delta\%$. For the present glasses the δ value is found to be 0.1238, 0.1231, 0.1224 and 0.1219 for glass A to D respectively. The positive sign of δ observed for all the glasses indicates that Sm-O bond in these glasses is covalent.

3.4. Radiative properties and visible luminescence analysis

The radiative properties such as radiative transition probability (A), total transition probability (A_T), radiative lifetime (τ_R), branching ratio (β_R) and integrated absorption cross-section (σ_a) for the electric dipole transitions between an excited level to its lower lying levels of Sm^{3+} ions in ZnAlBiB glasses are theoretically predicted using the expressions given in our previous paper [22]. The total radiative transition probability (A_T) for the excited states $^4G_{5/2}$, $^6F_{11/2}$, $^6F_{9/2}$, $^6F_{7/2}$, $^6F_{5/2}$, $^6F_{3/2}$ and $^6F_{1/2}$ of Sm^{3+} in all the ZnAlBiB glasses are given in Table 4 along with radiative lifetimes (τ_R) and absorption cross-sections (σ_a). It is observed that the order of magnitude of A_T parameter is $^6F_{11/2} > ^6F_{9/2} > ^6F_{7/2} > ^6F_{5/2} > ^6F_{3/2} > ^6F_{1/2}$. Among all the glasses studied, the glass D indicates higher radiative transition probabilities. Table 5 gives the electric dipole line strength, radiative transition probabilities and predicted radiative branching ratios (β_R) for certain excited states of Sm^{3+} ions in ZnAlBiB glasses. The magnitude of branching ratio (β_R), characterizes the lasing power of an emission transition. It is well established that, an emission transition with branching ratio greater than 0.50 is more potential for laser emission. From Table 5, it can be seen that $^4G_{5/2} \rightarrow ^6H_{7/2}$ transition has

higher branching ratios than the other transitions. Among all the glasses studies here, glass C and D are having higher branching ratios for the ${}^6G_{5/2} \rightarrow {}^6H_{7/2}$ transition.

To record the luminescence spectra for the prepared glasses, it is necessary to know the excitation wavelengths of Sm^{3+} ions. For this purpose, we recorded the excitation spectra from 350-450 nm for all ZnAlBiB glasses by fixing the emission at 602 nm and are shown in Fig.3. Totally five excitation bands are observed in all the glasses at wavelengths approximately 361nm, 374nm, 402nm, 417nm and 436nm corresponding to transitions ${}^6H_{5/2} \rightarrow {}^4D_{3/2}$, ${}^6P_{7/2}$, ${}^4F_{7/2}$, ${}^6P_{5/2}$ and ${}^4G_{5/2}$ respectively. In general, the Sm^{3+} ion will exhibit several overlapped excitation bands in the spectral range 450 to 500 nm due to the closely spaced energy levels located at these wavelengths. For this reason we measured the Sm^{3+} excitation spectra from 350 to 450 nm by fixing the emission at 602 nm. The sharpness of the emission bands depends on the host matrix phonon energies. Sometimes, Stark splitting may cause broadening of these observed bands. It is well known that the wavelength corresponding to the intense excitation band can give intense emissions. Among all excitation bands, a band corresponding to the transition ${}^6H_{5/2} \rightarrow {}^4F_{7/2}$ at 402 nm is more intense and is used as an excitation wavelength to record the emission spectra for all ZnAlBiB glasses. Fig.4 shows the emission spectra of ZnAlBiB glasses at room temperature. The emission spectra consists of three peaks centered at 566, 602 and 648 nm corresponds to the transitions ${}^4G_{5/2} \rightarrow {}^6H_{5/2}$, ${}^6H_{7/2}$ and ${}^6H_{9/2}$ respectively. Among the three transitions, the transition ${}^4G_{5/2} \rightarrow {}^6H_{7/2}$ is more intense when compared with the remaining two transitions. Upon excitation at 402 nm, the intense reddish-orange luminescence is observed even with naked eye from all these ZnAlBiB glasses. When Sm^{3+} ions are excited to any level above ${}^4G_{5/2}$, a fast non-radiative decay takes place due to the closely spaced energy levels above ${}^4G_{5/2}$. The excited Sm^{3+} ions thus reaching to ${}^4G_{5/2}$ relax radiatively through ${}^4G_{5/2}$ transition. The large energy separation ($\sim 7600\text{ cm}^{-1}$) between ${}^4G_{5/2}$ emission level and its lower lying ${}^6F_{11/2}$ state

prevents the possibility of multiphonon relaxation. The partial energy level diagram shown in Fig. 5 describes the emission mechanism involved in ZnAlBiB glasses along with the prominent cross relaxation channel from $^4G_{5/2}$ level.

To understand the Sm^{3+} luminescence in the glasses under study, the JO theory has been applied [34,35] to determine the stimulated emission cross-section (σ_e) for the emission transitions $^4G_{5/2} \rightarrow ^6H_{5/2}$, $^6H_{7/2}$, and $^6H_{9/2}$ using the relevant expression given in our previous paper [22] and are documented in Table 6 along with gain bandwidth and optical gain. The experimental branching ratio (β_{exp}) are obtained by using relative intensities of the individual peaks to that of the total intensity of emission peaks and are given in Table 6. The variation in calculated (Table 5) and experimental branching ratios may be attributed to the non radiative contributions from the $^4G_{5/2}$ level of Sm^{3+} ions in these glasses. The value of stimulated emission cross-section has been used to identify the potential laser transition of rare earth ions in glasses. A good laser transition can have a larger stimulated emission cross-section. The stimulated emission cross-section value is higher for the $^6G_{5/2} \rightarrow ^6H_{7/2}$ transition among all the emission transitions for the present glasses. The higher stimulated emission cross-section value is an attractive feature for low threshold, higher gain laser applications, which are used to obtain continuous wave (CW) laser action [30]. A glass material doped with RE ions can act as a good optical amplifier, if gain bandwidth ($\sigma_e \times \Delta\lambda_p$) and optical gain ($\sigma_e \times \tau_R$) parameters are relatively high. From Table 6, it can be observed that the reasonably higher values of these parameters for $^6G_{5/2} \rightarrow ^6H_{7/2}$ transition suggests that ZnAlBiB glasses are suitable for reddish-orange optical amplification applications.

3.5. Decay curve analysis and quantum efficiencies

The luminescence decay curves of the $^4G_{5/2}$ fluorescent level of Sm^{3+} doped ZnAlBiB glasses have been measured at room temperature and are shown in Fig. 6. All the measured

decay curves are single exponential in nature. The measured lifetime (τ_m) of the ${}^4G_{5/2}$ excited level has been found by taking the first e-fold times of the decay curves [33]. The measured lifetimes are found to be 1851, 1366, 2133 and 1977 μ s for glass A, B, C and D respectively. The radiative lifetimes (τ_R) obtained for the present glasses are 2854, 2815, 2447 and 2405 μ s for glass A, B, C and D respectively. The measured lifetimes are significantly smaller than the radiative lifetimes evaluated by using JO theory may be due to the energy transfer through cross relaxation, but not due to multi-phonon relaxation because the energy gap existing between ${}^4G_{5/2}$ energy level to its next lower level ${}^6F_{11/2}$ is very large (7600 cm^{-1}). In such case the non-radiative decay rate, W_{NR} is given by

$$W_{NR} = \frac{1}{\tau_m} - \frac{1}{\tau_R} \quad (3)$$

The W_{NR} (s^{-1}) for the ${}^4G_{5/2}$ level is found to be 190, 376, 60 and 90 for glass A to glass D respectively. The high magnitudes of non-radiative decay rates are due to difference between the experimental and radiative lifetimes. For direct excitation, the luminous quantum efficiency is defined as the ratio of the emitted light to the absorbed pump intensity. For a rare-earth ion, it is equal to the ratio for the measured lifetime (τ_m) to the radiative lifetime (τ_R) and is an important parameter to predict the laser host material [36].

$$\eta = \frac{\tau_m}{\tau_R} \times 100 \quad (4)$$

The quantum efficiency (η) depends on emission cross-section, radiative transition probabilities, lifetimes of the excited states, concentration of rare-earth ions and ligand field

produced by network modifiers. The radiative lifetimes (τ_R), measured lifetimes (τ_m), quantum efficiencies (η) and non-radiative relaxation rates (W_{NR}) for $^4G_{5/2} \rightarrow ^6H_{7/2}$ transition of Sm^{3+} ions are given in Table 7 along with the values reported for other glasses[16,25,37,38].

4. Conclusions

Zinc Alumino Bismuth Borate glasses with good optical quality at different compositions were prepared by melt quenching technique with 1 mol% of Sm^{3+} ions. The covalent character of Sm^{3+} ions with its surrounding ligands have been confirmed through optical spectra. The room temperature absorption and emission spectra of Sm^{3+} ions doped ZnAlBiB glasses were recorded and analysed using the Judd-Ofelt theory. The reasonably small rms deviations observed between the measured and calculated oscillator strengths indicates the accuracy of JO parameters. From the evaluated JO parameters, the radiative properties such as transition probability, radiative lifetimes, branching ratio's and integrated absorption cross-sections were calculated for various fluorescent levels of Sm^{3+} ions and are found to be comparable with other reported values. The room temperature emission spectra recorded for all the glasses show two intense emissions one at orange region ($^4G_{5/2} \rightarrow ^6H_{7/2}$) and the other one at red region ($^4G_{5/2} \rightarrow ^6H_{9/2}$). Based on the visible emission spectra, the large stimulated emission cross-section and high branching ratio's observed for $^4G_{5/2} \rightarrow ^6H_{7/2}$ transition for all the glasses, suggest the utility of these materials as lasing materials. Quantum efficiency (η), an another important parameter to characterize the materials is also evaluated by measuring the experimental lifetimes (τ_m) of $^4G_{5/2}$ level and combining them with the radiative lifetimes (τ_R) measured from JO theory. Based on the physical, optical properties, strong visible emissions, large stimulated emission cross-sections, high branching

ratios and good quantum efficiencies, it is concluded that ZnAlBiB glasses are suitable luminescent materials for the development of laser and photonic device operating in visible region.

Acknowledgements

One of the authors Swapna Koneru is very much thankful to Department of Science and Technology (DST), Government of India, New Delhi, for awarding her with a Women Scientist's scheme under DST – WOS (A) File No. SR/WOS-A/PS-35/2011(G)/2011 programme.

References

- [1] B. J. Chen, E. Y. B. Pun, H. Lin, J. Alloys Compd. 479(2009) 352-356.
- [2] B. Lal, J. Wang, Q. Su, Appl. Phys. B. 98(2010) 41-47.
- [3] A. Mohan Babu, B.C. Jamalaiah, J. Suresh Kumar, T. Sasikala, L. Rama Moorthy, J. Alloys compd. 509 (2011) 457-462.
- [4] Y. N. Ch. Ravi Babu, P. Sree Ram Naik, K. Vijaya Kumar, N. Rajesh Kumar and A. Suresh Kumar, J. Quant. Spectrosc. Rad. Trans. 113(2012) 1669-1675.
- [5] Yanlin Huang, Chuanfang Jiang, Kiwan Jang, HoSueb Lee, Eunjin Cho, M. Jayasimhadri, Soung-Soo Yi, J. Appl. Phys. 103 (2008) 113519 (7 pages).
- [6] M. Jayasimhadri, L. Rama Moorthy, R.V.S.S.N. Ravikumar, Opt. Mater. 29 (2007) 1321-1326
- [7] S.A. Saleem, B.C. Jamalaiah, M. Jayasimhadri, A. Srinivasa Rao, Kiwan Jang, L. Rama Moorthy, J. Quant. Spectrosc. Rad. Trans. 112 (2011) 78-84.
- [8] Y. G. Choi, J. Heo, J. Non-Cryst. Solids 217(1997)199-207.

- [9] Yan Zhang, Chunhua Lu, Liyan Sun, Zhongzi Xu, Yaru Ni, Mater. Res. Bull. 44(2009) 179-183.
- [10] A. Wells, Structural Inorganic Chemistry, 4th Ed.(1975) (Oxford, Clarendon).
- [11] G. Keizer, Optical Fibre Communications, Tata Mc. Graw-Hill, India, 2009.
- [12] M. Peng, L. Wondraczek, Optic Lett. 35(2010)2544-2546.
- [13] K. Annapurna, M. Das, M. Kundu, R.N.Dwivedhi, S.Buddhudu, J.Molecular Structure 741(2005) 53-60.
- [14] R. S. Quimby, P. A. Tick, N. F. Borrelli, L. K. Cornelius, J. Appl. Phys. 83(1998) 1649-1653.
- [15] R. Van Deun, K. Binnemans, C. Gorllerwalrand, J. L. Adam, J. Alloys. Comp 283 (1999) 59-65.
- [16] O. Ravi, C. Madhukar Reddy, L. Manoj, B. Deva Prasad Raju, J. Molecular Struct. 1029(2012) 53-59.
- [17] C.K. Jayasankar, K. Upendra Kumar, V. Venkatramu, P. Babu, Th. Troster, W. Sievers, G. Wortmann, J. Lumin. 128 (2008) 718-720.
- [18] K. K. Mahato, D. K. Rai, Solid State Commun. 108 (1998) 671-676.
- [19] A. S. Rao, Y. N. Ahammed, R. R. Reddy and T. V. R. Rao, Opt. Mater. 10 (1998) 245-252.
- [20] W. T. Carnall, P. R. Fields, K. Rajnak, J. Chem. Phys. 49(1968) 4424-4442.
- [21] C. Gorller-walrand, K. Binnemans, K. A. Gschneidner Jr, I. Eyring (Eds.) Handbook on the physics and chemistry of rare-earths, Vol.25, North-Holland, Amsterdam, 1998, PP.101-264 (Chapter 167)
- [22] A. Srinivasa Rao, B. Rupa Venkateswara Rao, M. V. K. Srinivasa Prasad, J. V. Shanmukha Kumar, M. Jayasimhadri, J. L. Rao and R. P. S. Chakradhar, Physica B, 404 (2009) 3717-3721.

- [23] S. Tanabe, T. Ohayagi, N. Soga, T. Hanada, Phys. Rev. B 46 (1992) 3305-3310.
- [24] P. Srinivasa and S. B. Rai, Spectrochim Acta A 60 (2004) 637-642.
- [25] Y. C. Ratnakar, D. Tirupati Naidu, A. Vijay Kumar and N. O. Gopal, Physica B 358 (2005) 296-307.
- [26] R. R. Jacobs and M. J. Weber, IEEE J. Quantum Electron. 12 (1976) 102-111.
- [27] Y. C. Ratnakaram, D. Tirupati Naidu, R. P. S. Chakradhar, J. Non-Cryst. Solids, 352 (2006) 3914-3922.
- [28] C. K. Jayasankar, E. Rukmini, Opt. Mater. 8 (1997) 193-205.
- [29] K. Maheshvaran, K. Linganna, K. Marimuthu, J. Lumin, 131 (2011) 2746-2753.
- [30] S. P. Sinha, In: Complexes of the Rare-Earths, Pergamon Press, Oxford, 1966.
- [31] M. S. Saisudha, J. Ramakrishna, Opt. Mater. 18 (2002) 403-417.
- [32] L. Boehm, R. Reisfield, N. Spector, J. Solid State Chem. 28 (1979) 75-78.
- [33] V. M. Orera, P. J. Alonso, R. Cases, R. Alkala, Phys. Chem. Glasses. 29 (1988) 59-62.
- [34] B. R. Judd, Phys. Rev 127 (1962) 750-761
- [35] G. S. Ofelt, J. Chem. Phys. 37 (1962) 511-521.
- [36] S. Surendra Babu, C. K. Jayasankar, P. Babu, T. Troster, W. Sievers, G. Wortmann, Phys. Chem. Glasses 47 (2006) 548-552.
- [37] S. Shanmuka Sundari, K. Marimuthu, M. Sivaraman, S. Surendra Babu, J. Lumin, 130 (2010) 1313-1319.
- [38] M. Jayasimhadri, Eun-Jin Cho, Ki-Wan Jang, HoSueb Lee, Sun il Kim, J. Phys. D: Appl. Phys. 41 (2008) 175101-175107.

Table 1
Various physical properties of Sm^{3+} - doped Zinc Aluminium Bismuth Borate (ZnAlBiB) glasses.

Physical property	Glass A	Glass B	Glass C	Glass D
Refractive index (n_d)	1.803	1.801	1.799	1.797
Density d(g/cc)	3.5475	3.794	4.0405	4.287
Average molecular weight, M (g)	96.6	114.8	133.09	151.2
Sm^{3+} ion concentration, $N(\times 10^{20} \text{ ions/cm}^3)$	4.418	3.977	3.657	3.413
Mean atomic volume ($\text{g/cm}^3/\text{atom}$)	6.195	6.882	7.486	8.020
Dielectric constant, (ϵ)	3.250	3.243	3.236	3.229
Optical dielectric constant ($\epsilon - 1$)	2.250	2.243	2.236	2.229
Reflection losses (%)	8.207	8.117	8.148	8.119
Molar refraction	11.686	12.957	14.068	15.045
Polaron radius (\AA)	1.160	1.200	1.235	1.260
Interionic distance (\AA)	2.880	2.980	3.064	3.140
Molecular electronic polarizability, $\alpha_e (\times 10^{-24} \text{cm}^3)$	2.320	2.570	2.787	2.990
Field strength, $F (\times 10^{16} \text{cm}^{-2})$	2.230	2.080	1.966	1.880
Optical basicity (Λ_{Th})	0.674	0.664	0.654	0.644

Table 2

Measured ($f_{\text{exp}} \times 10^{-6}$) and calculated ($f_{\text{cal}} \times 10^{-6}$) oscillator strengths of Sm^{3+} doped ZnAlBiB glasses.

Transitions	Glass A		Glass B		Glass C		Glass D	
from $^4\text{I}_{9/2}$	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}
$^6\text{F}_{9/2}$	2.7046	2.5378	2.61	2.872	3.318	3.832	3.451	4.024
$^6\text{F}_{7/2}$	3.3878	3.5050	4.10	3.923	5.416	5.052	5.755	5.380
$^6\text{F}_{5/2}$	1.2665	1.4927	1.363	1.537	1.383	1.643	1.643	1.787
$^6\text{F}_{3/2}$	1.4424	0.9156	1.037	0.824	1.117	0.873	0.808	0.821
$^6\text{H}_{15/2}$	0.0440	0.0215	0.689	0.025	0.165	0.034	0.032	0.036
$^6\text{F}_{1/2}$	0.0160	0.3404	0.064	0.187	0.055	0.188	0.0510	0.023
$\delta_{\text{rms}}(\times 10^{-6})$	± 0.282		± 0.324		± 0.305		± 0.293	

Table 3

Judd-Ofelt parameters ($\times 10^{-20} \text{ cm}^2$), trends of Ω_λ parameters and spectroscopic quality factor ($\chi = \Omega_4 / \Omega_6$) of the Sm^{3+} doped ZnAlBiB glasses and reported Sm^{3+} glasses.

Glass	J O Parameters			Trend	$\chi = \Omega_4 / \Omega_6$	References
Composition	Ω_2	Ω_4	Ω_6			
Glass A	1.029	2.511	2.866	$\Omega_6 > \Omega_4 > \Omega_2$	0.876	Present Work
Glass B	0.553	2.655	3.260	$\Omega_6 > \Omega_4 > \Omega_2$	0.812	Present Work
Glass C	0.557	2.839	4.401	$\Omega_6 > \Omega_4 > \Omega_2$	0.645	Present Work
Glass D	0.067	3.162	4.610	$\Omega_6 > \Omega_4 > \Omega_2$	0.686	Present Work
SPG12	1.47	2.0	3.09	$\Omega_6 > \Omega_4 > \Omega_2$	0.647	[27]
KZnBS(Set A)	0.180	11.37	11.45	$\Omega_6 > \Omega_4 > \Omega_2$	0.99	[28]
KZnBS(Set B)	0.930	10.97	11.41	$\Omega_6 > \Omega_4 > \Omega_2$	0.96	[28]
NaZnBS	0.550	9.680	9.770	$\Omega_6 > \Omega_4 > \Omega_2$	0.99	[28]
B0TS	0.200	3.478	3.023	$\Omega_4 > \Omega_6 > \Omega_2$	1.15	[29]
B1TS	0.460	3.435	2.327	$\Omega_4 > \Omega_6 > \Omega_2$	1.47	[29]
B2TS	0.370	3.273	2.312	$\Omega_4 > \Omega_6 > \Omega_2$	1.41	[29]
B3TS	0.372	2.980	1.867	$\Omega_4 > \Omega_6 > \Omega_2$	1.59	[29]
B4TS	0.066	2.819	1.362	$\Omega_4 > \Omega_6 > \Omega_2$	1.78	[29]
Phosphate	4.310	4.280	5.780	$\Omega_6 > \Omega_4 > \Omega_2$	0.74	[30]
LCN Borate	0.84	4.00	5.02	$\Omega_6 > \Omega_4 > \Omega_2$	0.80	[31]

Table 4

Total radiative transition probabilities (A_T) (s^{-1}), radiative lifetimes (τ_R) (μs) and absorption cross section (σ_a) for fluorescent levels of Sm^{3+} doped ZnAlBiB glasses

Level	Glass A			Glass B			Glass C			Glass D		
	A_T	τ_R	σ_a	A_T	τ_R	σ_a	A_T	τ_R	σ_a	A_T	τ_R	σ_a
$^4G_{5/2}$	350.3	2854	0.804	355.17	2815	0.636	408.64	2447	0.728	415.75	2405	0.736
$^6F_{11/2}$	1285.7	777	4.338	1413.2	707	1.21	1816.3	550	4.32	1895.8	527	2.881
$^6F_{9/2}$	1084.3	922	2.823	1178.3	848	3.41	1490.7	670	1.999	1554.4	643	2.326
$^6F_{7/2}$	785.43	1273	3.362	834.49	1198	3.736	1029.8	971	3.691	1067.7	936	2.975
$^6F_{5/2}$	532.37	1878	3.238	542.00	1844	3.438	645.01	1550	2.982	656.16	1524	3.148
$^6F_{3/2}$	388.93	2571	4.19	370.63	2698	2.741	423.52	2361	2.055	411.64	2429	3.14
$^6F_{1/2}$	317.55	3149	1.406	281.52	3552	1.667	310.12	3224	2.107	281.84	3548	2.194

Table 5

Electric dipole line strength ($S_{\text{ed}} \times 10^{-22} \text{cm}^2$), radiative transition probabilities (A_{R}) (s^{-1}), radiative branching ratio (β_{R}) and total radiative transition probability (A_{T}) (s^{-1}) for Sm^{3+} doped ZnAlBiB glasses

Transition	Glass A			Glass B			Glass C			Glass D		
	S_{ed}	A_{R}	β_{R}	S_{ed}	A_{R}	β_{R}	S_{ed}	A_{R}	β_{R}	S_{ed}	A_{R}	β_{R}
$^4\text{G}_{5/2} \rightarrow ^6\text{H}_{5/2}$	0.17	29.36	0.0838	0.17	29.10	0.0819	0.18	29.40	0.0719	0.19	29.51	0.0710
$^4\text{G}_{5/2} \rightarrow ^6\text{H}_{7/2}$	4.71	163.7	0.4674	5.18	177.5	0.4999	6.35	212.6	0.5204	6.81	225.8	0.5432
$^4\text{G}_{5/2} \rightarrow ^6\text{H}_{9/2}$	3.42	83.27	0.2377	3.07	74.26	0.2091	3.42	82.65	0.2022	3.13	75.34	0.1812
	$A_{\text{T}} = 350.357$			$A_{\text{T}} = 355.178$			$A_{\text{T}} = 408.648$			$A_{\text{T}} = 415.758$		

Table 6

Emission peak wavelength (λ_p)(nm), effective band widths($\Delta\lambda_p$)(nm), branching ratio (β_{exp}), stimulated emission cross-sections (σ_{se})($\times 10^{-22}$) (cm^2), gain band width ($\sigma_{se} \times \Delta\lambda_p$)($\times 10^{-28}$) (cm^3) and optical gain ($\sigma_{se} \times \tau_R$) ($\times 10^{-25} \text{cm}^2 \text{s}$) for the emission transitions of Sm^{3+} doped ZnAlBiB glasses

Transition	Parameters	Glass A	Glass B	Glass C	Glass D
$^4\text{G}_{5/2} \rightarrow ^6\text{H}_{5/2}$	λ_p	561.31	565.59	565.59	565.86
	$\Delta\lambda_p$	10.638	9.2592	9.2592	10.294
	β_{exp}	0.223	0.211	0.216	0.227
	σ_{se}	1.12	1.32	1.33	1.21
	$\sigma_{se} \times \Delta\lambda_p$	1.19	1.22	1.23	1.24
	$\sigma_{se} \times \tau_R$	3.19	3.70	3.26	2.90
$^4\text{G}_{5/2} \rightarrow ^6\text{H}_{7/2}$	λ_p	601.35	601.36	601.36	601.36
	$\Delta\lambda_p$	12.765	11.111	11.111	11.029
	β_{exp}	0.545	0.531	0.537	0.558
	σ_{se}	6.85	8.55	11.3	11.0
	$\sigma_{se} \times \Delta\lambda_p$	8.74	9.50	11.4	12.1
	$\sigma_{se} \times \tau_R$	19.5	24.1	27.7	26.2
$^4\text{G}_{5/2} \rightarrow ^6\text{H}_{9/2}$	λ_p	647.65	647.62	647.62	647.67
	$\Delta\lambda_p$	12.765	12.962	12.962	12.499
	β_{exp}	0.213	0.198	0.209	0.216
	σ_{se}	4.68	4.12	4.60	4.36
	$\sigma_{se} \times \Delta\lambda_p$	5.98	5.34	5.96	5.45

$\sigma_{se} \times \tau_R$	13.4	11.6	11.3	10.5
-----------------------------	------	------	------	------

Table 7

Radiative lifetimes (τ_R) (μs), measured lifetimes (τ_m) (μs), quantum efficiency (η) and Non radiative relaxation rate (W_{NR}) (μs) for Sm^{3+} doped ZnAlBiB glasses.

Host	τ_m	τ_R	η (%)	W_{NR}	Reference
Glass A	1851	2854	64	190	Present glass
Glass B	1366	2815	46	376	Present glass
Glass C	2133	2447	87	60	Present glass
Glass D	1977	2405	82	90	Present glass
TMZNB	450	1900	24	1695	[16]
TCZNB	520	2180	27	1411	[16]
TSNB	770	1850	46	758	[16]
B4TS	1888	3936	48	276	[29]
N5BS	2175	2852	76	109	[38]
LGT10	795	1550	51	613	[39]

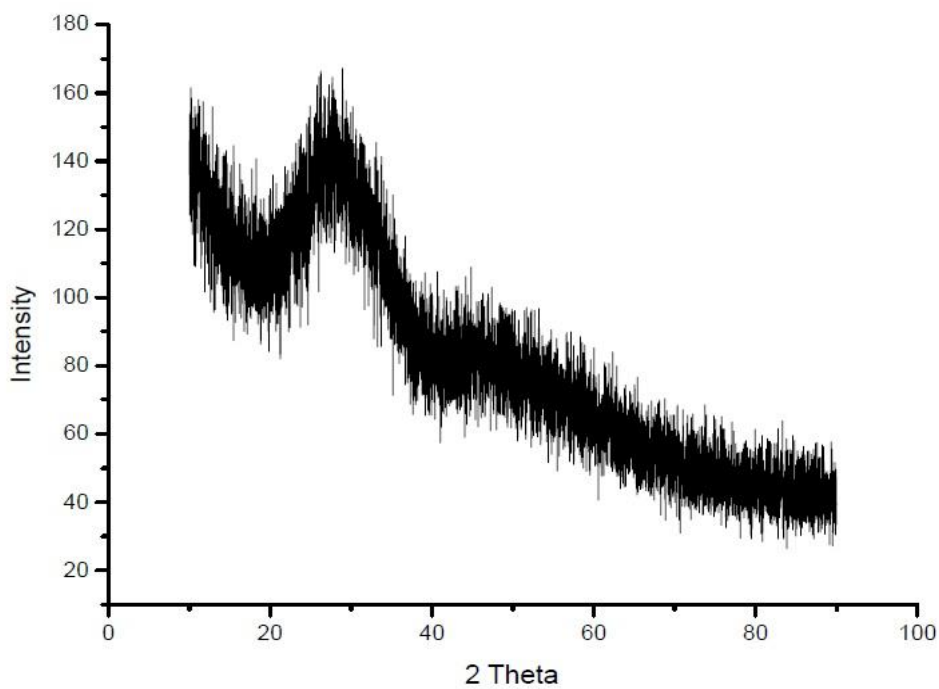


Fig 1. XRD spectrum for undoped ZnAlBiB glass

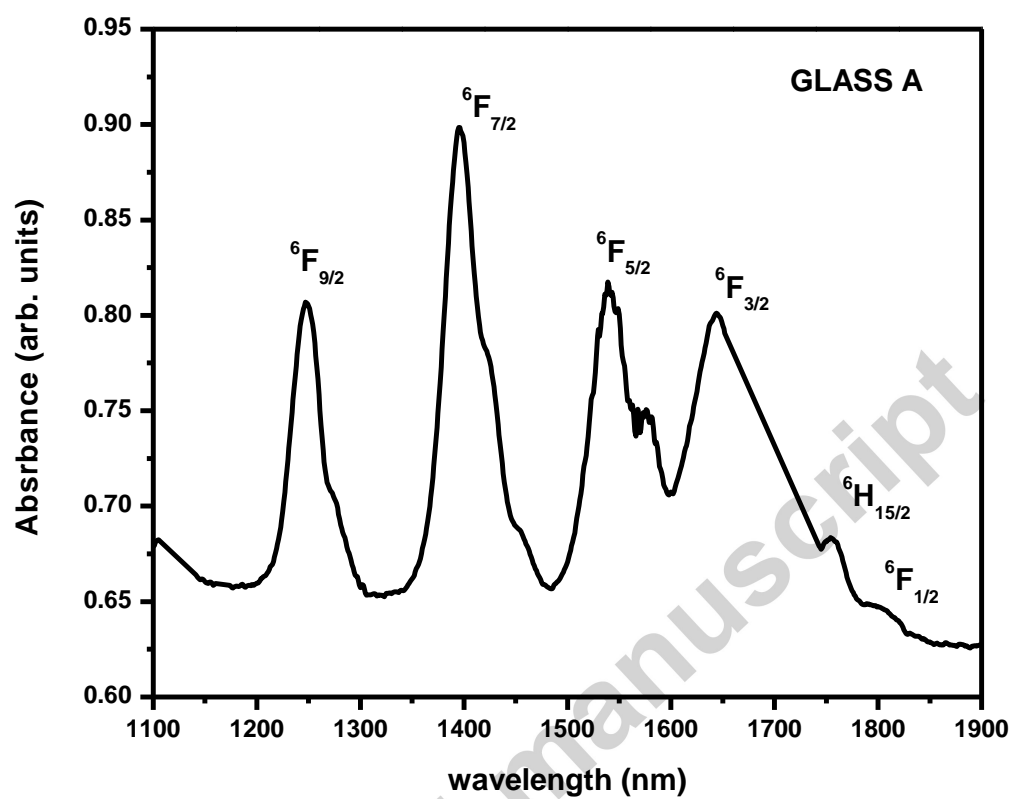


Fig 2. Absorption spectra of Sm^{3+} : ZnAlBiB glass

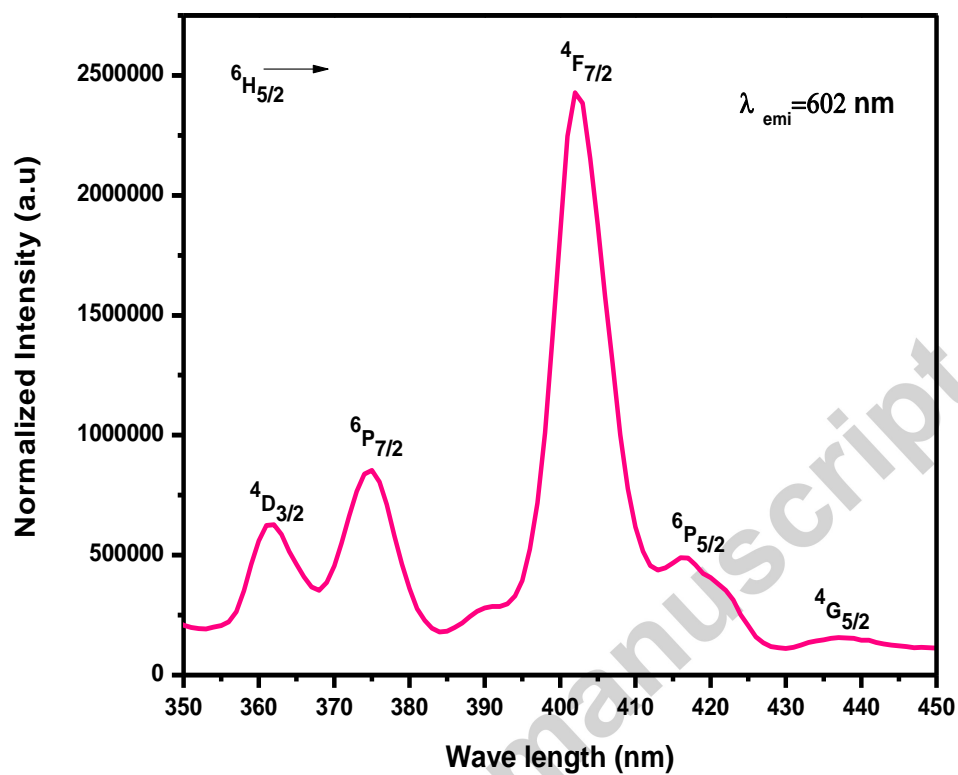


Fig 3. Excitation spectra of Sm^{3+} : ZnAlBiB glasses

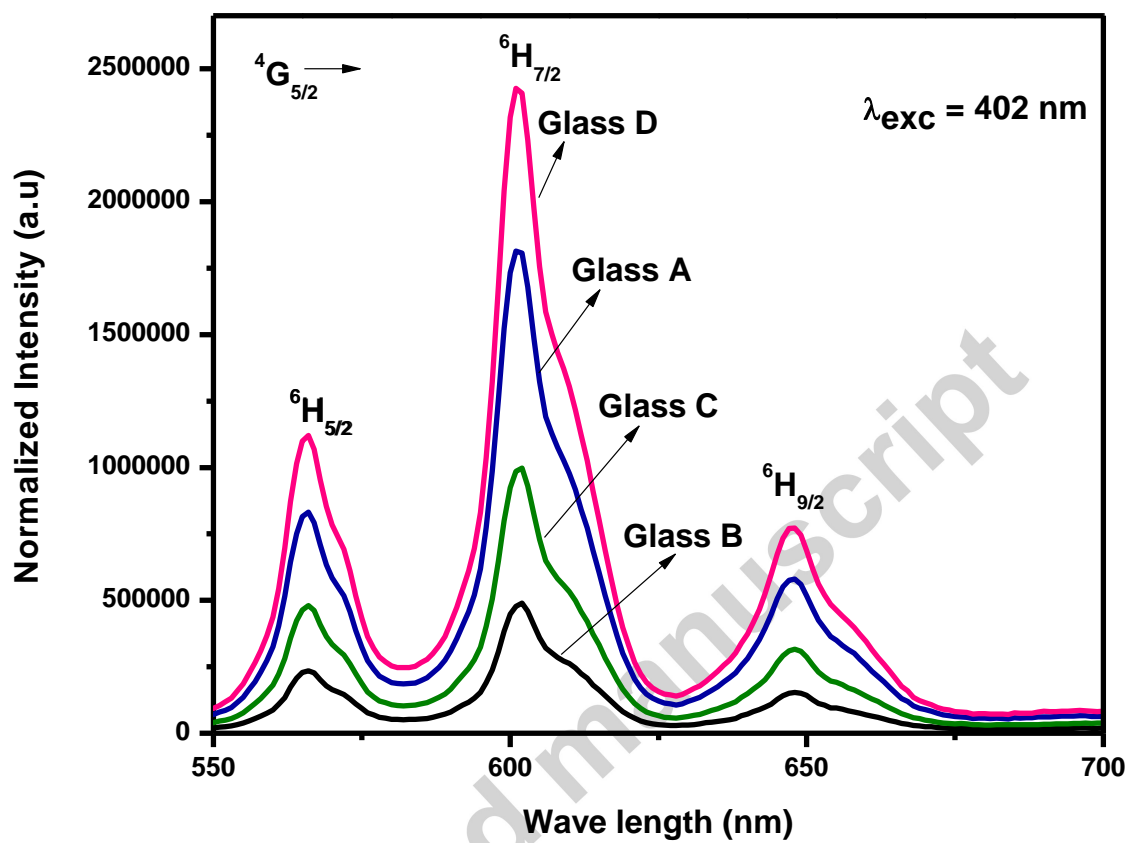


Fig 4. Emission spectra of Sm^{3+} : ZnAlBiB glasses.

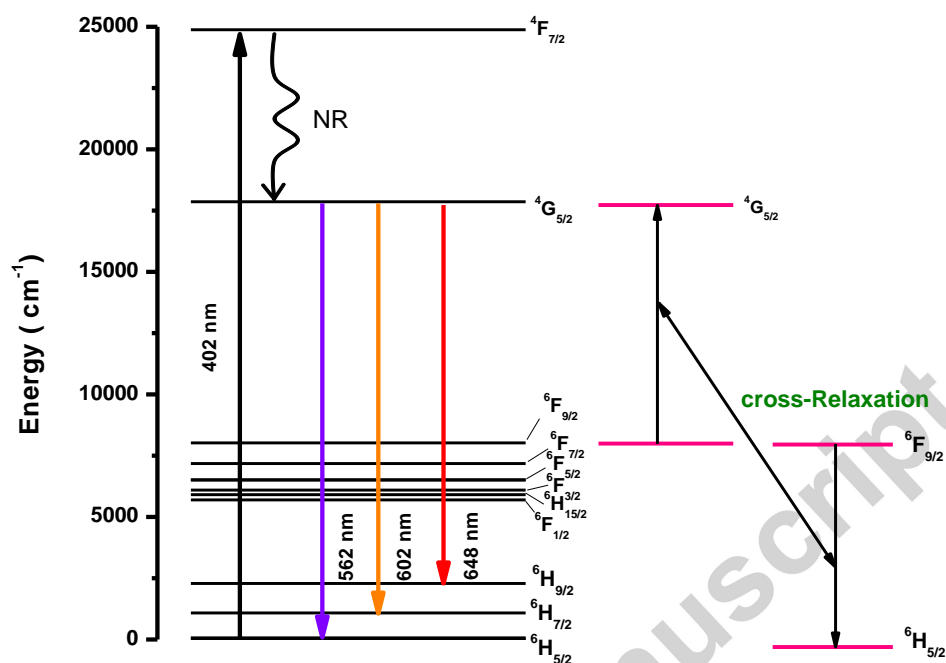


Fig 5. Energy levels depicting the various lasing transitions of Sm^{3+} ions. The main cross relaxation channel for $^4\text{G}_{5/2}$ level is also shown separately.

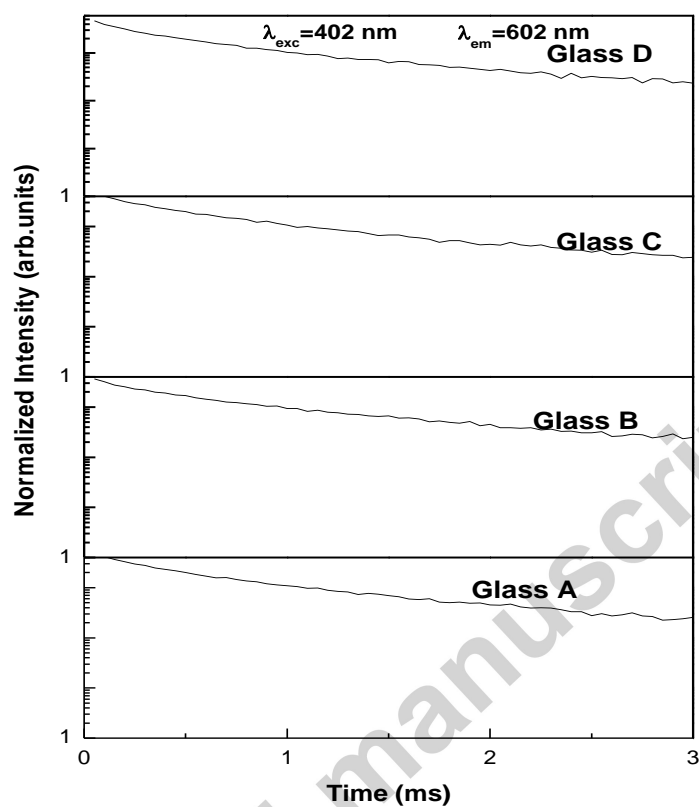
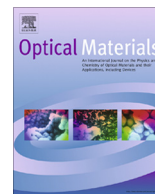


Fig 6. Decay profiles of $^4G_{5/2}$ level for same concentrations of Sm^{3+} ion in ZnAlBiB glasses.



Contents lists available at ScienceDirect

Optical Materials

journal homepage: www.elsevier.com/locate/optmat

Visible red, NIR and Mid-IR emission studies of Ho³⁺ doped Zinc Alumino Bismuth Borate glasses

Sk. Mahamuda^a, K. Swapna^a, P. Packiyaraj^c, A. Srinivasa Rao^{a,b,*}, G. Vijaya Prakash^c

^a Department of Physics, KL University, Green Fields, Vaddeswaram 522 502, Guntur (Dt), AP, India

^b Department of Applied Physics, Delhi Technological University, New Delhi 110 042, India

^c Nanophotonics Laboratory, Department of Physics, Indian Institute of Technology – Delhi, New Delhi 110 016, India

ARTICLE INFO

Article history:

Received 8 August 2013

Received in revised form 21 September 2013

Accepted 23 September 2013

Available online xxxx

Keywords:

Glasses

Holmium

JO parameters

Radiative properties

Emission cross-section

Optical gain

ABSTRACT

Zinc Alumino Bismuth Borate (ZnAlBiB) glasses doped with different concentrations of Holmium were prepared by conventional melt quenching technique. The glassy nature of these glasses has been confirmed through the XRD spectral measurements. The FTIR spectra recorded for undoped glass revealed the information related to the functional groups involved in the host glass. Optical absorption, excitation and photoluminescence spectra of these glasses have been recorded at room temperature. The Judd–Ofelt theory has been applied successfully to characterize the absorption spectra of the ZnAlBiB glasses. From this theory various radiative properties such as radiative transition probability (A_R), radiative lifetimes (τ_R), branching ratios (β_R) and spectroscopic quality factor (χ) for the prominent emission levels $^5F_5 \rightarrow ^5I_7$, $^5F_5 \rightarrow ^5I_8$ and $^5I_7 \rightarrow ^5I_8$ have been evaluated. The photoluminescence spectra revealed the quenching of luminescence intensity beyond 1.0 mol% of Ho³⁺ ion concentration in ZnAlBiB glasses. To investigate the lasing potentiality of $^5F_5 \rightarrow ^5I_7$, $^5F_5 \rightarrow ^5I_8$ and $^5I_7 \rightarrow ^5I_8$ transitions, the effective band width ($\Delta\lambda_p$) and the stimulated emission cross-section (σ_{se}) were determined. The CIE chromaticity co-ordinates were also evaluated from the emission spectra for all the glasses to understand the suitability of these materials for visible red laser emission in principle.

© 2013 Published by Elsevier B.V.

1. Introduction

Rare earth doped glasses based on the radiative transitions from the intra 4f shell energy levels of rare earth ions have been successfully utilized as an active media for solid state lasers operating in the visible to infrared (IR) spectral ranges, fiber lasers, optical amplifiers due to their wide applications in laser remote sensing, tele-communications, atmosphere transmissions, Raman laser, amplifiers, optical parametric oscillators, eye-safe LIDAR and biomedical applications [1–5]. The main advantage of glasses is, it is easy to produce large homogeneous pieces having low non-linear refractive indices with large doping concentration. Among all the glasses, oxide glasses are more stable and harder [6]. The relationship between local structure and fluorescence properties of oxide glasses containing rare earth ions have been investigated widely to make new laser glasses and phosphors. The optical properties of rare earth ions are strongly influenced by the ligand field surrounding them and on phonon energies of the host glass. In general a host glass with low phonon energies can give high quan-

tum efficiencies which are useful to design lasers and optical fibers [7]. Among all the oxide glasses, borate glasses (B₂O₃) have several advantages over conventional silicate and phosphate glasses such as high transparency, low melting point, high thermal stability and good rare earth ion solubility [8,9]. However these glasses cannot act as good laser hosts due to their high phonon energies ($\sim 1300\text{ cm}^{-1}$). Such high phonon energies possessed by the borate glasses can be reduced by adding heavy metal oxides like Bi₂O₃ to give high quantum efficiencies. The glasses based on the heavy metal oxides have linear (>2) and non-linear (up to 10^2 times that of silicate glasses) refractive index and a broad transparency range from the visible up to the mid infrared ($0.4 < \lambda < 8\text{ }\mu\text{m}$) [10–12]. The broad transparency range in the mid infrared is related to the lowest phonon energy among all the oxide glasses [13]. All these properties make heavy metal oxide glasses very attractive for photonic applications such as optical wave guides in the 2.5–3.0 μm range and more efficient gain or non-linear optical devices [10–12]. The glasses based on ZnO have attracted many researchers because of its optical, electrical and magnetic properties in combination with its non-toxicity, non-hygroscopic nature. The glasses containing ZnO have also been used in the development of opto-electronic devices, solar energy converters, UV lasers and gas sensors [14]. The presence of Al₂O₃ in oxide glasses enhances the mechanical strength, gives greater chemical durability, lowers

* Corresponding author at: Department of Applied Physics, Delhi Technological University, New Delhi 110 042, India. Tel.: +91 85860 39007; fax: +91 08662555989.

E-mail address: drsallam@yahoo.co.in (A. Srinivasa Rao).

the coefficient of thermal expansion and increases freedom from devitrification.

Among various rare earth ions, Ho^{3+} ions in glasses can emit bands in the ultra violet, visible, near infrared and mid infrared regions at $\sim 2 \mu\text{m}$ [15,16]. In recent years, the lasers emitting light at mid infrared region are widely used in the areas of laser ranger finders, LIDAR devices and remote sensing because of their 'eye safe' characteristics. The lasers of wavelength around $2 \mu\text{m}$ can be absorbed strongly by OH^- and the heating effects generated can make the body tissues concentered, carbonized and vaporized. They can be used for operation with less skin being damaged, less bleeding and less pollution than using the metal bistouries. The mid infrared laser sources also find application in atmospheric sensing of environmentally important gases such as methane, nitrogen dioxide, carbon dioxide, carbon monoxide, on-line monitoring of exhaust gas pollution by industrial plants, vehicles and jet engines. The other possible applications include optical radar, e.g., collision avoiding radar, ammonia and water vapour sensing in agriculture, molecular spectroscopy and point-to-point atmospheric communication. Holmium doped lasers operating in mid infrared region at around $2 \mu\text{m}$ are quite suitable for the above said applications when compared with the other rare-earth ions operating in the same region. Most of the commercially available lasers operate in the visible to near infrared spectrum from 0.4 to $1.4 \mu\text{m}$. This wavelength band usually called as the retinal hazard region because the laser light might cause damage to the retina resulting in a blind spot in the fovea. The laser light outside the retinal hazard region, the ultraviolet between 0.29 and $0.40 \mu\text{m}$ or mid to far infrared between 1.40 and $10.60 \mu\text{m}$ spectrum will not focus on the retina. Erbium doped lasers can generate outputs at 1.55 , 1.65 , 2.7 and $3.46 \mu\text{m}$, but the performance is not very impressive in terms of efficiency or energy output [17]. Thulium doped lasers operating in the wavelength of $1.9 \mu\text{m}$ have also been available commercially with output power 100 watt . However, the lower stimulated emission cross-section is unfavorable for the energy extraction of Q-switched oscillator and pulsed amplifiers. This drawback restricts their applications in LIDAR and remote sensing. On the other hand Holmium doped lasers have large stimulated emission cross-section and is good for the energy extraction of Q-switch oscillator and pulse amplifier. The Holmium ions are most important active ions for lasers because of their favorable energy level structure. It is well known that Ho^{3+} ion has $^5\text{I}_8$ ground state and $^5\text{I}_7$, $^5\text{I}_6$, $^5\text{I}_5$ and $^5\text{I}_4$ as lower lying excited states at about 5100 , 8600 , $11,200$ and $13,200 \text{ cm}^{-1}$ respectively. Most of these states are Meta-stable and are suitable for lasing action at 0.55 and $2.0 \mu\text{m}$ wavelengths [18–20] which are useful for the development of visible and mid infrared solid state lasers respectively. In addition to that, the Ho^{3+} ion provides eye-safe potential laser emission even at room temperature (RT) with a low threshold action that has attractive applications in atmospheric communication systems [21].

In the present investigation, we have analyzed the spectroscopic properties of different concentrations of Ho^{3+} doped Zinc Alumino Bismuth Borate (ZnAlBiB) glasses to study the suitability of these glasses for solid state laser devices and optical fiber amplifiers.

2. Experimental techniques

The glass samples used in the present work were prepared by conventional melt quenching technique with the following compositions.

Glass A : $20\text{ZnO} - 10\text{Al}_2\text{O}_3 - 9.9\text{Bi}_2\text{O}_3 - 60\text{B}_2\text{O}_3 - 0.1\text{Ho}_2\text{O}_3$

Glass B : $20\text{ZnO} - 10\text{Al}_2\text{O}_3 - 9.5\text{Bi}_2\text{O}_3 - 60\text{B}_2\text{O}_3 - 0.5\text{Ho}_2\text{O}_3$

Glass C : $20\text{ZnO} - 10\text{Al}_2\text{O}_3 - 9.0\text{Bi}_2\text{O}_3 - 60\text{B}_2\text{O}_3 - 1.0\text{Ho}_2\text{O}_3$

Glass D : $20\text{ZnO} - 10\text{Al}_2\text{O}_3 - 8.5\text{Bi}_2\text{O}_3 - 60\text{B}_2\text{O}_3 - 1.5\text{Ho}_2\text{O}_3$

Glass E : $20\text{ZnO} - 10\text{Al}_2\text{O}_3 - 8.0\text{Bi}_2\text{O}_3 - 60\text{B}_2\text{O}_3 - 2.0\text{Ho}_2\text{O}_3$

Glass F : $20\text{ZnO} - 10\text{Al}_2\text{O}_3 - 7.5\text{Bi}_2\text{O}_3 - 60\text{B}_2\text{O}_3 - 2.5\text{Ho}_2\text{O}_3$

Appropriate amounts of analytical reagents ZnO , Al_2O_3 , Bi_2O_3 , B_2O_3 and Ho_2O_3 were used as starting materials. About 10 g of batch compositions were crushed in an agate mortar to obtain homogeneous mixture. The homogeneous mixture was then taken into an alumina crucible and heated to 1200°C for 50 min in an electric furnace. The melt was thoroughly shaken at 5 min interval in order to ensure homogeneity. The resultant bubble free melt was then poured into a brass mould and subsequently annealed at 400°C for 4 h to remove thermal strains. After annealing, they were optically polished before measuring their optical properties.

The refractive indices of glass samples were measured at $\lambda = 650 \text{ nm}$ (He–Ne laser) by using Brewster's method with an accuracy ± 0.01 . By classical Archimedes principle, the densities were determined with water as an immersion liquid with an accuracy ± 0.01 . The FTIR spectrum for an undoped ZnAlBiB glass has been recorded with a spectral resolution of 4 cm^{-1} using Perkin–Elmer Spectrum II FTIR spectrophotometer in the wave number range $400\text{--}4000 \text{ cm}^{-1}$ following KBr Pellet technique to identify the presence of functional groups. The optical absorption spectra of the glasses were recorded using a double beam JASCO V-670 UV–vis–NIR spectrophotometer in the wavelength range $350\text{--}2500 \text{ nm}$ with a spectral resolution of 0.1 nm . The photoluminescence emission and excitation spectra for all these glasses were recorded at room temperature using Shimadzu RF-5301 PC Spectrofluorophotometer. The time-resolved photoluminescence (PL) measurements were carried out using home-built setups with 980 nm (CW laser) laser as an excitation source with 1 W power. The emission from sample was coupled into a monochromator (ANDOR SR-500i-B2) coupled to InGaAs detector through the appropriate lenses and filters.

3. Results and discussion

3.1. XRD spectral analysis

In order to check the amorphous nature of the prepared glasses, the XRD spectral measurement were taken for an undoped ZnAlBiB glass using XPERT-PRO diffractometer with $\text{Cu K}\alpha$ radiation ($k = 1.54\text{\AA}$) in the 2θ angle range $5\text{--}70^\circ$ with step size $0.02^\circ \text{ s}^{-1}$ and is shown in Fig. 1. The XRD spectrum shows a broad hump which is the characteristic of an amorphous material. Hence the prepared ZnAlBiB glasses have amorphous nature.

3.2. Physical properties

From the estimated values of density and refractive index, various other physical properties were calculated by using the relevant expressions available in literature [22] and are given in Table 1. From Table 1 it is observed that, the trend of inter atomic distance in the prepared glasses show a decrease of the available volume per atom. This implies that the atoms are more tightly packed as the Ho^{3+} ion concentration increases in the glass. The molar refraction is considered as the sum of the contributions of the cationic refraction and oxygen ionic refraction and it is related to the structure of the glass which measures the bonding condition in the glass. The low field strength values observed for the present glasses indicate the high solubility of rare earth ions in the glass. Molecular electronic polarizability is related to many macro and microscopic physical and chemical properties such as optical UV

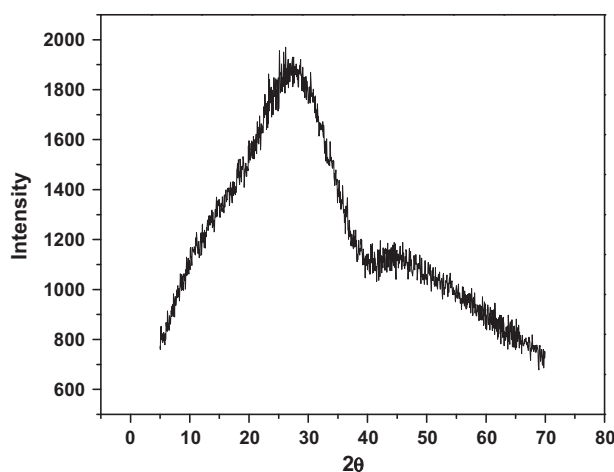


Fig. 1. XRD spectrum of an undoped ZnAlBiB glass.

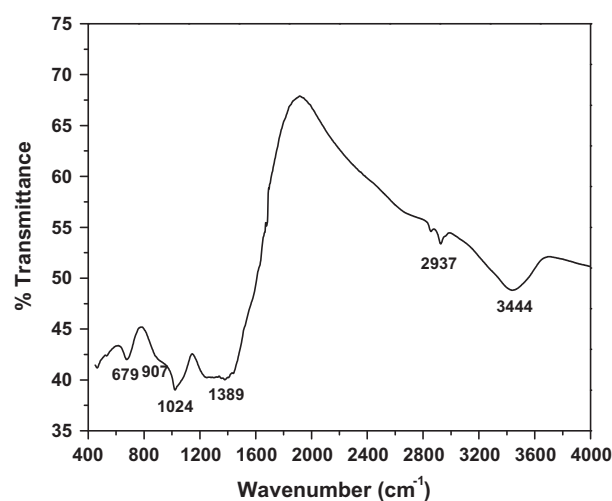


Fig. 2. FTIR spectrum of an undoped ZnAlBiB glass.

absorption, ionic refraction, dielectric properties and chemical stability. From Table 1, it is also observed that these glasses possess low molecular electronic polarizability which makes them more stable. The optical basicity of the oxide medium is the average electron donor power of all the oxide atoms comprising the medium. Increasing values of optical basicity results in increasing negative charge on the oxygen atom and thus increasing covalency in the cation–oxygen bonding.

3.3. FTIR spectral analysis

The FTIR spectrum of undoped ZnAlBiB glass in the region 400–4000 cm^{-1} is shown in Fig. 2. Fig. 2 shows the Characteristic IR absorption band at 3444 cm^{-1} , which is attributed to the fundamental vibrations of OH^- group. It is well known that the presence of OH^- content increases the non-radiative losses and then decreases the quantum efficiency of rare earth doped materials. From Fig. 2 it is observed that, the low intensity of OH band indicates that the ZnAlBiB glasses possess low OH content. The coefficient of OH^- vibration (α_{OH}) at 3000 cm^{-1} is used as a measure of the OH^- content [23]. In a glassy material, the content of OH^- has been estimated using the equation, $\alpha_{\text{OH}} = \ln \left(\frac{I_0}{I} \right) / l$ here T_0 is the highest transmission, T_0 is the transmission of the glass at 3000 cm^{-1} and l is the thickness of the glass sample. For ZnAlBiB host glass, the OH^- content is estimated as 65.48 (2.18 cm^{-1}), which is found to be smaller than LBTAF (123.00 ppm) [24], GeS2 (175.5 ppm) [25]

and LHG-8L (128 ppm) glasses [26]. Relatively less content of OH^- in these glasses indicates that these glasses are quite suitable for efficient lasers with less non-radiative loss.

The peak identified at 2937 cm^{-1} is attributed to hydrogen bonding. The absorption peaks in the FTIR spectra can be divided into three main groups in the ranges 1600–1200 cm^{-1} , 1200–800 cm^{-1} and 700–400 cm^{-1} corresponding to the asymmetric stretching relaxation of the B–O bond stretching of trigonal BO_3 units, B–O stretching of the BO_4 units and B–O–B linkages or vibrational modes of network modifier, respectively. The absence of the absorption band at 806 cm^{-1} is the characteristic peak of boroxyl ring which indicates that, there is no boroxyl ring formation in the prepared glasses [27]. The absorption band appears around 1389 cm^{-1} is due to asymmetric stretching vibrations of B–O and B–O $^-$ bond in trigonal BO_3 units. The IR absorption peaks around 1200–1400 cm^{-1} are characteristic of $\text{B}=\text{O}-\text{B} \equiv$ linkage with one of the boron in tetrahedral coordination [28]. Hence in the present series of glasses there is a growth of BO_4 tetrahedral units at the expense of triangular BO_3 units. The presence of peaks in the group 1017–1100 cm^{-1} corresponds to presence of tri, tetra, penta borate and diborate groups belonging to BO_4 groups. At the same time the glass structure also contains varying amounts of penta, meta, pyroborate units (peaks around 1374–1562 cm^{-1}) of BO_3 groups [29]. The peaks 618–650 cm^{-1} may be assigned to Bi–O bending vibrations in BiO_6 units [30,31]. The band assignments for FTIR spectrum are given in Table 2.

Table 1

Various physical properties of Ho^{3+} doped ZnAlBiB glasses.

Physical properties	Glass A	Glass B	Glass C	Glass D	Glass E	Glass F
Refractive index (n_d)	1.8190	1.8170	1.8150	1.8130	1.8110	1.8090
Density, d (g/cm^3)	3.9365	3.9345	3.9321	3.9296	3.9272	3.9247
Average molecular weight, \bar{M} (g)	114.75	114.40	113.96	113.52	113.08	112.64
Ho^{3+} ion concentration N ($\times 10^{22}$ ions/ cm^3)	0.2070	1.0400	2.0800	3.1300	4.1800	5.2500
Mean atomic volume ($\text{g/cm}^3/\text{atom}$)	6.6252	6.6082	6.5868	6.5655	6.5441	6.5227
Optical dielectric constant ($p\partial t/\partial p$)	3.3087	3.3014	3.2942	3.2869	3.2797	3.2724
Dielectric constant (ϵ)	2.3087	2.3014	2.2942	2.2869	2.2797	2.2724
Reflection losses ($R\%$)	0.0844	0.0841	0.0838	0.0835	0.0832	0.0829
Molar refraction (R_m) (cm^{-3})	12.677	12.622	12.559	12.496	12.432	12.369
Polaron radius (r_p) (Å)	3.2100	1.8800	1.4900	1.3000	1.1800	1.1000
Inter-atomic distance (r_i) (Å)	7.9800	4.6600	3.7000	3.2300	2.9300	2.7200
Molecular electronic polarizability α ($\times 10^{-23} \text{ cm}^3$)	5.0200	0.9970	0.4980	0.3300	0.2470	0.1960
Field strength F ($\times 10^{15} \text{ cm}^{-2}$)	2.9000	8.5100	13.500	17.700	21.500	25.000
Optical basicity (Δ_{th})	0.438	0.44	0.442	0.444	0.446	0.448

3.4. Absorption spectral analysis

The absorption spectra of 1 mol% of Ho^{3+} doped ZnAlBiB glass (Glass C) in UV–vis and NIR regions are shown in Fig. 3a. The assignments of the absorption bands originating from the ground level $^5\text{I}_8$ to various excited levels within the $4f$ shell are also shown in Fig. 3a. The absorption spectra contains twelve bands and are assigned to the transitions $^5\text{I}_8 \rightarrow ^5\text{I}_7$, $^5\text{I}_8 \rightarrow ^5\text{I}_6$, $^5\text{F}_5$, $^5\text{S}_2 + ^5\text{F}_4$, $^5\text{F}_3$, $^5\text{F}_2$, $^3\text{K}_8$, $^5\text{G}_6$, $(^5\text{G}, ^3\text{G})_5$, $^5\text{G}_4$, $^3\text{K}_7$ and $^3\text{H}_6$. The identification and assignment of energy levels are made as per the procedure outlined by Carnall et al. [32]. The observed band positions (cm^{-1}) and bonding parameters (β and δ) of Ho^{3+} doped ZnAlBiB glasses are given in Table 3. The bonding parameters (β , δ) for all Ho^{3+} doped glasses have been calculated using the expressions given in our previous paper [22]. From Table 3 it is observed that, for the present glasses, the bonding parameter is negative i.e., the prepared glasses possess ionic nature and the ionic nature decreases gradually as the Ho^{3+} ion concentration increases.

From Fig. 3a it is observed that the absorption bands $^5\text{I}_8 \rightarrow ^3\text{H}_6$ (360 nm), $^5\text{I}_8 \rightarrow (^5\text{G}, ^3\text{G})_5$ (417 nm), $^5\text{I}_8 \rightarrow ^5\text{G}_6$ (451 nm), $^5\text{I}_8 \rightarrow ^5\text{S}_2 + ^5\text{F}_4$ (537 nm) and $^5\text{I}_8 \rightarrow ^5\text{F}_5$ (642 nm) are relatively intense and sharp than the other transitions. The bands corresponding to $^5\text{I}_8 \rightarrow ^3\text{K}_7$ (381 nm), $^5\text{I}_8 \rightarrow ^5\text{G}_4$ (387 nm), $^5\text{I}_8 \rightarrow ^3\text{K}_8$ (467 nm), $^5\text{I}_8 \rightarrow ^5\text{F}_2$ (473 nm), $^5\text{I}_8 \rightarrow ^5\text{F}_3$ (484 nm) are found to be very weak while the bands due to the $^5\text{I}_8 \rightarrow ^5\text{I}_6$ (1150 nm) and $^5\text{I}_8 \rightarrow ^5\text{I}_7$ (1944 nm) are weak and broad. In majority of the rare earth ions, the absorption transitions are induced electric dipole in nature and magnetic dipole contribution is negligible. But in case of Ho^{3+} ion, the transition $^5\text{I}_8 \rightarrow ^5\text{I}_7$ (1944 nm) contains some magnetic dipole contribution [20].

The analysis of the absorption spectra can also provide insights on the local structure of the rare earth ions. Fig. 3b shows the ground state absorption transition $^5\text{I}_8 \rightarrow ^5\text{I}_7$ (1944 nm) in different concentrations of Ho^{3+} doped ZnAlBiB glasses. From this figure, it is observed that the band width of the absorption decreased sharply when the Ho^{3+} concentration increases in the present glasses. The strongest band in each absorption spectrum was used to obtain the values of the FWHM intensity. From Fig. 3b, it is observed that FWHM values decreased with the increase of Ho^{3+} concentration. Spectral line widths of optical transitions in rare earth ions embedded in materials mainly depend on two factors. The first contribution is due to the temperature dependent lattice vibration referred to as the homogeneous broadening and it is common for both amorphous and crystalline material. The second feature, inhomogeneous broadening which is pronounced in non-crystalline materials and it further increases the spectral line width. It is resulted from the site to site variation of the ligand field surrounding rare earth ions. The decrease in FWHM values in Fig. 3b was attributed to the decrease in the homogeneous broadening with the increase in Ho^{3+} concentration. This decrease was unusual since the site-to-site variation was expected to increase as the concentration of the rare earth ions increased [33,34].

Table 2
Band positions (cm^{-1}) of FTIR spectra of an undoped ZnAlBiB glass.

Band positions (cm^{-1})	Assignments
3444	OH stretching vibrations
2937	Hydrogen bonding
1389	B–O stretching containing planar 6 membered borate groups
1024	B–O bonds stretching vibrations in BO_4 units
907	Progressive change of boron coordination number from BO_3 to BO_4 units
679	BiO_3 and BiO_6 group vibrations

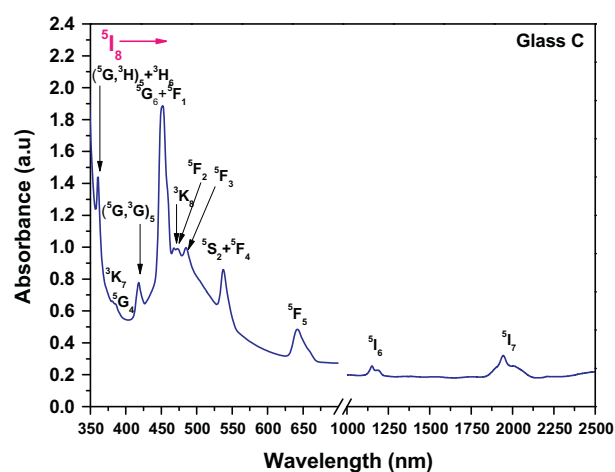


Fig. 3a. Absorption spectrum of 1 mol% of Ho^{3+} doped ZnAlBiB glass (Glass C) in UV–vis–NIR region.

The experimental oscillator strengths (f_{exp}) of the absorption transitions originating from the ground level of a rare earth ions have been evaluated by measuring the integrated areas under the absorption bands from the absorption spectrum by using the expression given in Ref. [14]. The J–O theory [35,36] has been applied to the experimentally evaluated oscillator strengths to find the J–O intensity parameters Ω_λ ($\lambda = 2, 4$ and 6) by the least square fit analysis. The J–O parameters thus evaluated are used to calculate oscillator strength (f_{cal}) and such calculated oscillator strengths are given in Table 4 along with experimental oscillator strengths. The quality of the fit is measured by root mean square deviation (δ_{rms}) between the experimental and calculated oscillator strengths by using the following expression given in Ref. [37]. The values of such r.m.s deviation are given in Table 4. From Table 4 it is observed that, the changes in the oscillator strengths within a relatively small compositional range are the direct evidence of the large and abrupt modification in the local structure surrounding Ho^{3+} ions.

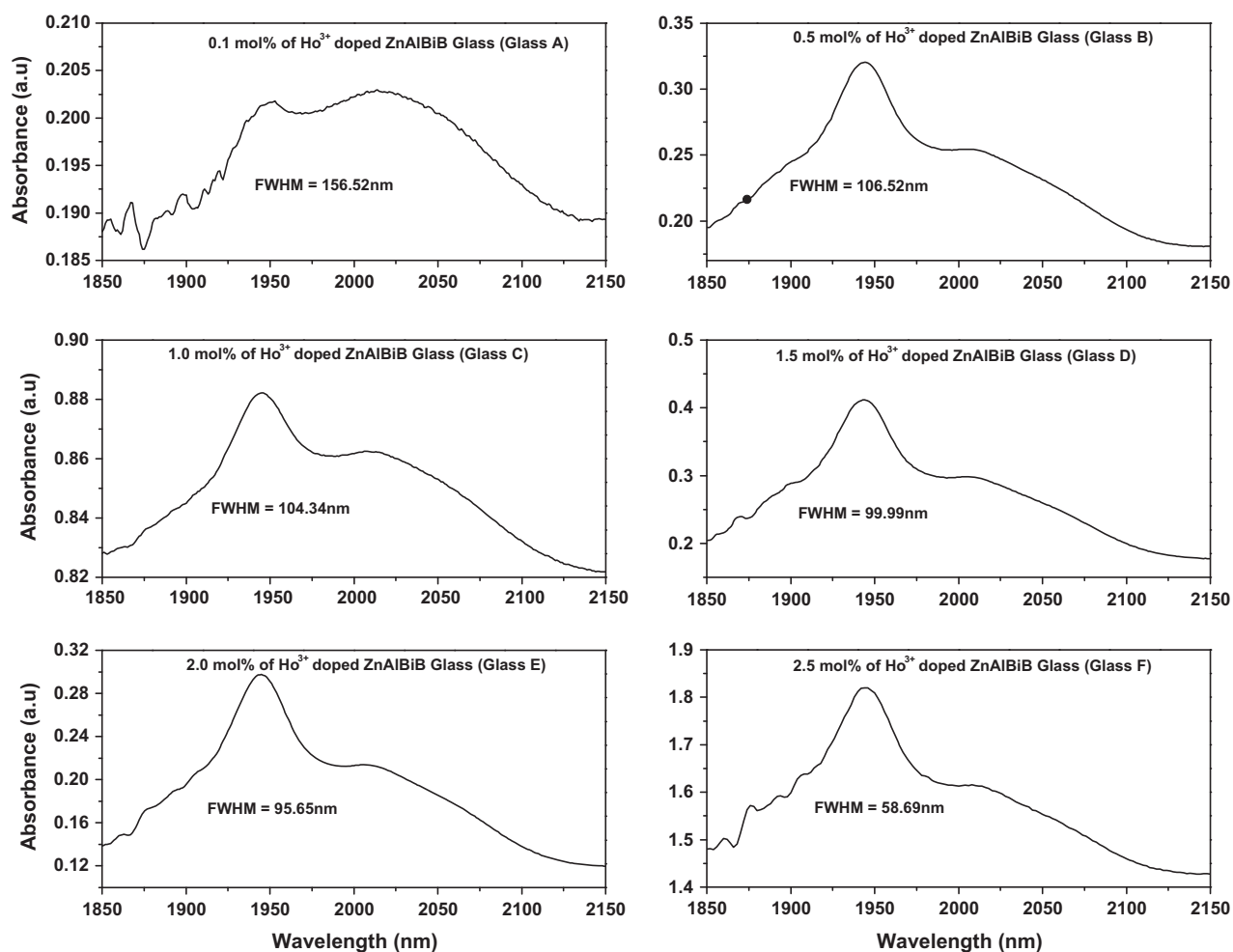
Among the various absorption transitions of the Ho^{3+} ion, the $^5\text{I}_8 \rightarrow ^5\text{G}_6$ (451 nm) is known as hypersensitive transition [38]. They are very sensitive to the host environment of the rare earth ion and these transitions obey the selection rules $|\Delta S| = 0$, $|\Delta L| \leq 2$, $|\Delta J| \leq 2$. Among the electric dipole transitions, the hypersensitive transitions possess large value of $\|U^2\|^2$ matrix element. The hypersensitivity arises mainly due to the inhomogeneity around the rare earth ions and in turn associated with the magnitude of Ω_2 parameter. The higher the Ω_2 parameter the greater will be the hypersensitivity. In the present investigation, the $^5\text{I}_8 \rightarrow ^5\text{G}_6 + ^5\text{F}_1$ transition centered at 451 nm has higher intensity when compared to the other transitions (Table 4).

The Ω_λ parameters thus evaluated for different concentrations of Ho^{3+} doped ZnAlBiB glasses (Glass A to Glass F) are compared with other Ho^{3+} doped hosts in Table 5 [39–46,19]. The J–O intensity parameters Ω_λ ($\lambda = 2, 4$ and 6) for the present glasses follow trend $\Omega_2 > \Omega_4 > \Omega_6$ which is similar to the trends observed in lead phosphate [39], Nb_2O_5 , La_2O_3 , Ta_2O_3 [40], phosphate (Na based) [41], oxy fluoride tellurite [42], oxy fluoride (Si based) [43], Ba–Al–Ge [44], $\text{Na}_4\text{AlZnP}_3\text{O}_{12}$ [45] glasses where as the dissimilar trends have been noticed in other Ho^{3+} systems such as fluorophosphate [46] and $\text{PbO–Al}_2\text{O}_3\text{–B}_2\text{O}_3$ [19].

It is well known fact that the magnitude of J–O intensity parameters is related to the physical and chemical properties such as viscosity and covalent character of the chemical bonds. In general Ω_2

Table 3Absorption band positions (cm^{-1}) and bonding parameters ($\bar{\beta}$ and δ) of Ho^{3+} doped ZnAlBiB glasses.

Transition $^5\text{I}_8 \rightarrow$	Glass A	Glass B	Glass C	Glass D	Glass E	Glass F
$^5\text{I}_7$	5136	5144	5144	5144	5144	5144
$^5\text{I}_6$	8710	8688	8695	8688	8688	8688
$^5\text{F}_5$	15,625	15,600	15,576	15,600	15,576	15,600
$^5\text{S}_2 + ^5\text{F}_4$	18,656	18,656	18,621	18,621	18,587	18,621
$^5\text{F}_3$	20,661	20,618	20,661	20,618	20,618	20,618
$^5\text{F}_2$	21,186	21,186	21,141	21,186	21,141	21,186
$^3\text{K}_8$	21,413	21,367	21,413	21,367	21,413	21,367
$^5\text{G}_6 + ^5\text{F}_1$	22,123	22,123	22,172	22,123	22,172	22,123
$(^5\text{G}, ^3\text{G})_5$	23,980	23,980	23,980	23,980	23,980	23,980
$^5\text{G}_4$	–	25,906	25,839	25,906	25,906	25,839
$^3\text{K}_7$	–	26,246	26,246	26,246	26,246	26,246
$(^5\text{G}, ^3\text{H})_5 + ^3\text{H}_6$	–	27,777	27,777	27,777	27,777	27,777
$\bar{\beta}$	1.0049	1.0038	1.0037	1.0036	1.0035	1.0034
δ	–0.4912	–0.3841	–0.3770	–0.3686	–0.3583	–0.3468

**Fig. 3b.** The ground state absorption transition $^5\text{I}_8 \rightarrow ^5\text{I}_7$ (1944 nm) in different concentrations of Ho^{3+} doped ZnAlBiB glasses.

values for the rare earth ions in glasses are intermediate between crystalline oxides and chelating ligands [47]. In glasses, the rare earth ions are randomly distributed over nonequivalent sites with a wide distribution in the crystal fields. In such cases, a distribution of large number of rare earth ions occupying sites with non-centro-metric potential will contribute significant changes in the Ω_2 value [48]. The large Ω_2 parameter values are indicators of the covalent character of the chemical bonds among the glass matrix and rare earth ions. The evaluated Ω_2 parameters for the present glasses are found to be comparable with the other glasses [39–46,19].

From Table 5 it is observed that Ω_2 parameter is high for Glass C among all the glasses and found to be higher than those reported for different glasses [39–46,19]. The Ω_2 parameter depends mainly on the asymmetry of the sites in the neighborhood of rare earth ion. The higher the Ω_2 parameter the higher is the degree of asymmetry around the rare earth ion and stronger the covalency of rare earth ion-oxygen bond. From Table 5, it is concluded that the sites occupied by the 1 mol% of Ho^{3+} ions in ZnAlBiB glass (Glass C) possess higher asymmetry and greater degree of covalency of $\text{Ho}^{3+}-\text{O}^{2-}$ bond than those of earlier reports [39–46,19]. The

Table 4
Measured ($f_{exp} \times 10^{-6}$) and calculated ($f_{cal} \times 10^{-6}$) oscillator strengths and rms deviation for various absorption transitions of Ho³⁺ doped ZnAlBiB glasses.

Transitions from $^5I_8 \rightarrow$	Glass A		Glass B		Glass C		Glass D		Glass E		Glass F	
	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}	f_{exp}	f_{cal}
5I_7	0.110	0.549	0.276	0.801	0.571	1.218	0.257	0.724	0.156	0.475	0.096	0.350
5I_6	0.198	0.367	0.300	0.502	0.343	0.805	0.250	0.443	0.200	0.278	0.102	0.204
5F_5	3.749	2.288	4.581	2.986	5.196	3.662	4.173	2.863	3.468	2.337	2.768	1.617
$^5S_2 + ^5F_4$	3.263	1.026	3.832	1.260	4.607	1.734	3.354	1.157	2.836	0.880	1.938	0.608
5F_3	1.265	0.321	2.236	0.318	5.144	0.627	3.056	0.244	1.081	0.118	0.981	0.081
5F_2	1.750	0.183	2.281	0.181	3.300	0.356	1.314	0.139	0.376	0.067	0.237	0.046
3K_8	0.905	0.358	1.812	1.056	2.331	1.403	1.422	1.045	0.227	0.697	0.113	0.556
$^5G_6 + ^5F_1$	2.870	2.819	27.84	27.44	34.76	34.29	28.81	28.39	18.68	18.40	15.83	15.54
$(^5G, ^3G)_5$	1.678	3.640	3.063	5.001	3.871	5.578	3.427	4.941	2.787	4.235	1.583	2.931
$\delta_{rms} (\times 10^{-6})$	± 1.285		± 1.613		± 2.170		± 1.550		± 1.016		± 0.937	

Table 5
Comparison of Judd–Ofelt parameters ($\Omega_i \times 10^{-20} \text{ cm}^2$), Spectroscopic quality factor ($\chi = \frac{\Omega_4}{\Omega_6}$) and their trend for Ho³⁺ doped ZnAlBiB glasses with different hosts.

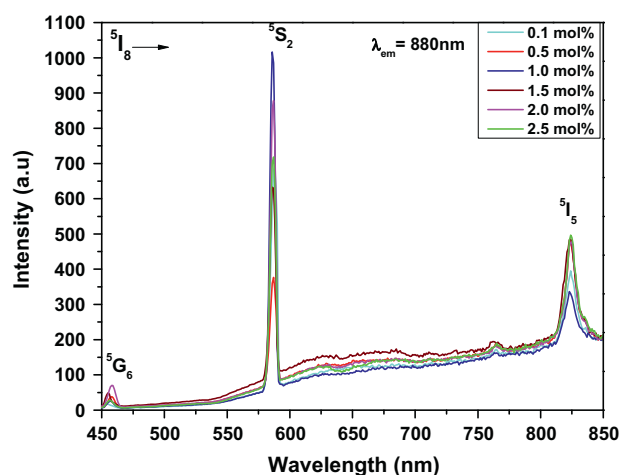
Glass system	Ω_2	Ω_4	Ω_6	$\sum \Omega_i$	Trend	References
Glass A	5.58 ± 0.19	2.59 ± 0.01	0.40 ± 0.25	8.57	$\Omega_2 > \Omega_4 > \Omega_6$	Present work
Glass B	13.01 ± 0.07	3.63 ± 0.44	0.41 ± 0.20	17.05	$\Omega_2 > \Omega_4 > \Omega_6$	Present work
Glass C	16.47 ± 0.59	4.06 ± 1.40	0.81 ± 0.11	21.34	$\Omega_2 > \Omega_4 > \Omega_6$	Present work
Glass D	13.61 ± 0.07	3.60 ± 0.46	0.32 ± 0.18	17.53	$\Omega_2 > \Omega_4 > \Omega_6$	Present work
Glass E	8.40 ± 0.20	3.07 ± 0.08	0.15 ± 0.31	11.62	$\Omega_2 > \Omega_4 > \Omega_6$	Present work
Glass F	7.39 ± 0.27	2.15 ± 0.44	0.10 ± 0.11	9.64	$\Omega_2 > \Omega_4 > \Omega_6$	Present work
Lead phosphate	8.68	3.04	2.37	14.09	$\Omega_2 > \Omega_4 > \Omega_6$	[39]
Nb ₂ O ₅	13.51	3.34	2.98	19.83	$\Omega_2 > \Omega_4 > \Omega_6$	[40]
La ₂ O ₃	6.52	2.22	2.53	11.27	$\Omega_2 > \Omega_4 > \Omega_6$	[40]
Ta ₂ O ₃	10.43	2.02	2.57	15.02	$\Omega_2 > \Omega_4 > \Omega_6$	[40]
Glass A (Li)	11.00	5.62	3.73	20.35	$\Omega_2 > \Omega_4 > \Omega_6$	[41]
Glass C (K)	13.06	7.72	4.60	25.38	$\Omega_2 > \Omega_4 > \Omega_6$	[41]
Glass E (Li–K)	10.41	6.07	3.68	20.16	$\Omega_2 > \Omega_4 > \Omega_6$	[41]
Phosphate (Na based)	8.58	4.31	2.88	15.77	$\Omega_2 > \Omega_4 > \Omega_6$	[41]
Oxyfluoride tellurite	4.20	2.80	1.10	8.1	$\Omega_2 > \Omega_4 > \Omega_6$	[42]
Oxyfluoride (Si based)	5.84	2.38	1.75	9.97	$\Omega_2 > \Omega_4 > \Omega_6$	[43]
Ba–Al–Ge	1.96	0.64	0.11	2.71	$\Omega_2 > \Omega_4 > \Omega_6$	[44]
Na ₄ AlZnP ₃ O ₁₂ glass	4.65	2.02	1.34	8.01	$\Omega_2 > \Omega_4 > \Omega_6$	[45]
Fluorophosphate	2.08	3.11	1.50	6.69	$\Omega_2 > \Omega_4 > \Omega_6$	[46]
PbO–Al ₂ O ₃ –B ₂ O ₃	5.83	17.15	26.50	49.48	$\Omega_2 > \Omega_4 > \Omega_6$	[19]

intensity parameters Ω_4 and Ω_6 are related to the bulk properties such as viscosity and dielectric of the media and are also affected by the vibronic transitions of the rare earth ions bound to the ligand atoms [49] and less sensitive to the medium in which the ions are situated.

3.5. Emission spectra and radiative properties

Rare earth ions doped materials exhibit sharp emission and excitation bands. The excitation spectra of different concentrations of Ho³⁺ doped ZnAlBiB glasses (Glass A to Glass F) at 880 nm emission wavelength is shown in Fig. 4. The excitation spectra has three excitation bands centered at 459, 586 and 824 nm corresponding to the transitions $^5I_8 \rightarrow ^5G_6$, $^5I_8 \rightarrow ^5S_2$ and $^5I_8 \rightarrow ^5I_5$ respectively. Among these three excitation bands the peak at 586 nm is sharp and intense and it has been used to investigate the luminescence characteristics of different concentrations of Ho³⁺ doped ZnAlBiB glasses. To investigate the photoluminescence characteristics in the mid infrared region, the present glasses were excited with 980 nm laser. The photoluminescence spectra of different concentrations of Ho³⁺ doped ZnAlBiB glasses excited at 586 nm and 980 nm obtained in the spectral region 600–900 nm and 1880–1940 nm respectively are shown in Figs. 5a and 5b. The luminescence spectra have three emission bands corresponding to the transitions $^5F_5 \rightarrow ^5I_8$ (623 nm), $^5F_5 \rightarrow ^5I_7$ (880 nm) and $^5I_7 \rightarrow ^5I_8$ (1.91 μm). These three transitions are sharp and intense and exhibit the bright red, near infrared and mid infrared emissions at 623, 880 nm and 1.91 μm respectively. From Figs. 5a and 5b, it is ob-

served that the luminescence intensity of Ho³⁺ ions increases with the increase of Ho³⁺ ion concentration, reaches a maximum value at 1.0 mol% and then decreases i.e., the concentration quenching has been occurred at 1 mol% for all the peaks. This is mainly due to the Ho³⁺–Ho³⁺ interactions at higher concentrations. When the Ho³⁺ ions are excited to 5S_2 and 5I_6 levels radiative and non-radiative (NR) relaxations take place due to the closely spaced higher

**Fig. 4.** Excitation spectra of various concentrations of Ho³⁺ doped ZnAlBiB glass monitored at the emission wavelength 880 nm.

energy levels. The inset in Fig. 5a shows the variation of $\sum \Omega_i$ with the intensity of the transition $^5F_5 \rightarrow ^5I_8$. This figure shows the linear relationship between $\sum \Omega_i$ with spectral intensities of the emission transition ($^5F_5 \rightarrow ^5I_8$). From this figure, it can be observed that, $\sum \Omega_i$ is used to represent the covalency of the R–O bond which have strong influence on the enhancement of the intensity of the emission transitions. The partial energy level diagram shown in Fig. 6 clearly describes the visible, near infrared and mid infrared emission mechanisms of 1 mol% of Ho^{3+} doped ZnAl–BiB glass (Glass C).

In order to predict the emission performance of Ho^{3+} doped ZnAlBiB glasses, the radiative parameters such as spontaneous transition probabilities (A_R), total transition probabilities (A_T), radiative lifetimes (τ_R) and luminescence branching ratios (β_R) for all the emission transitions have been evaluated by using the J–O parameters obtained from the absorption spectral measurements and are given in Table 6. In treating the fluorescence process, the radiative transition probabilities (A_R), total transition probability (A_T), radiative lifetime (τ_R) and branching ratios (β_R) have been evaluated using the relevant expressions given in our previous paper [50]. The magnitude of fluorescence branching ratios (β_R), which defines the luminescence efficiency of transition involving different levels are given in Table 6. In general branching ratio is a critical parameter to the laser designer, because it characterizes the possibility of attaining stimulated emission from any specific transition. The fluorescence branching ratio which characterizes and makes the transition as potential for laser emission, if its value $\beta_R \geq 0.5$. From Table 6, it is observed that among all the glasses Glass C possess highest branching ratio for all the transitions. In the present study the calculated branching ratios (β_R) for the transitions $^5F_5 \rightarrow ^5I_7$, $^5F_5 \rightarrow ^5I_8$ and $^5I_7 \rightarrow ^5I_8$ are found to be 0.285, 0.683 and 1.00 respectively. The experimental branching ratios are determined (β_{exp}) from the emission spectra by measuring the area of the corresponding emission and dividing it by the total integrated area of all the emission bands and are given in Table 6. The experimental and calculated branching ratios (β_{exp} and β_R) listed in Table 6 are in good agreement with each other. Hence Glass C is quite suitable to design red, near infrared and mid infrared laser.

The stimulated emission cross-section (σ_{se}) is an important parameter and its value signifies the rate of energy extraction from the optical material. From the observed emission bands σ_{se} can be estimated by the following expression [50]:

$$\sigma_{se}(a_j, b_j) = \frac{\lambda_p^4}{8\pi c n^2 \Delta\lambda_p} A_R(a_j, b_j) \quad (1)$$

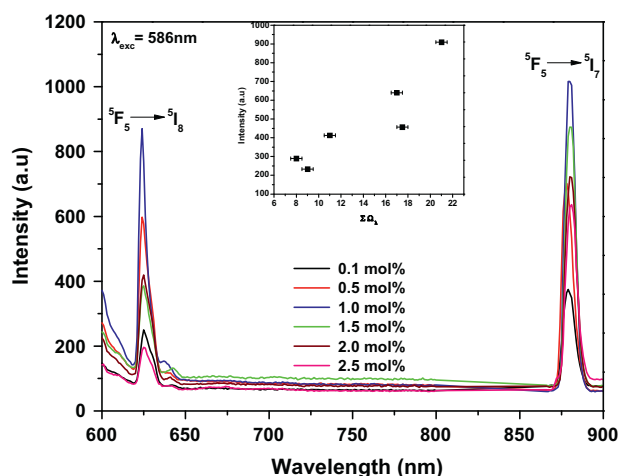


Fig. 5a. Emission spectra of various concentrations of Ho^{3+} doped ZnAlBiB glasses at an excitation wavelength 586 nm. The inset shows variation of $\sum \Omega_i$ with the intensity of the transition $^5F_5 \rightarrow ^5I_8$.

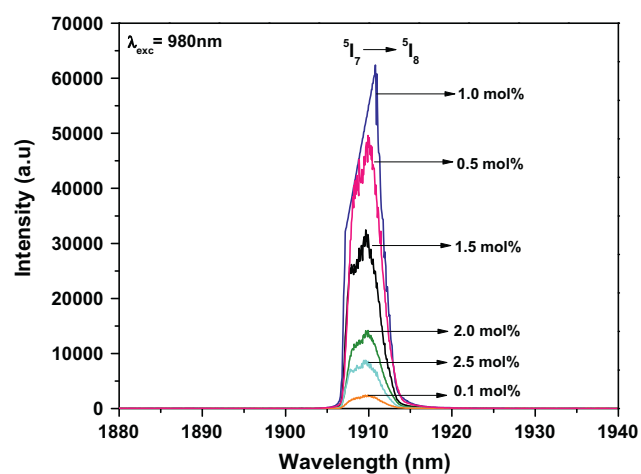


Fig. 5b. Emission spectra of various concentrations of Ho^{3+} doped ZnAlBiB glasses at an excitation wavelength 980 nm.

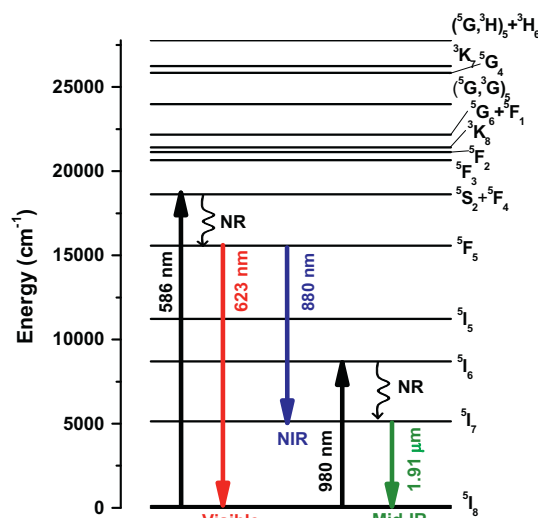


Fig. 6. Typical energy level diagram depicting the various radiative and non-radiative (NR) transitions of Ho^{3+} ion in ZnAlBiB glass.

Table 6

Transition probabilities (A_R) (s^{-1}), calculated and experimental branching ratios (β_R & β_{exp}) total transition probability (A_T) (s^{-1}) and radiative lifetimes (τ_R) (μs) for the emission transitions of Ho^{3+} ions in ZnAlBiB glasses.

Transition		Glass A	Glass B	Glass C	Glass D	Glass E	Glass F
$^5F_5 \rightarrow ^5I_7$	A_R	353.76	550.51	775.65	532.86	414.73	288.11
	β_R	0.221	0.250	0.285	0.253	0.245	0.244
	β_{exp}	0.203	0.298	0.346	0.222	0.223	0.203
	A_T	1600.9	2197.6	2711.7	2103.5	1686.3	1180.2
	τ_R	624	455	368	475	592	847
$^5F_5 \rightarrow ^5I_8$	A_R	1062.4	1492.3	1863.6	1411.4	1124.2	756.41
	β_R	0.663	0.679	0.683	0.670	0.666	0.641
	β_{exp}	0.593	0.606	0.635	0.577	0.568	0.543
	A_T	1600.9	2197.6	2711.7	2103.5	1686.3	1180.2
	τ_R	624	455	368	475	592	847
$^5I_7 \rightarrow ^5I_8$	A_R	24.25	29.19	38.37	27.40	22.39	19.23
	β_R	1.00	1.00	1.00	1.00	1.00	1.00
	β_{exp}	1.00	1.00	1.00	1.00	1.00	1.00
	A_T	24.25	29.19	38.37	27.40	22.39	19.23
	τ_R	4123	3430	2610	3650	4470	5200

where λ_p is the emission peak wavelength and $\Delta\lambda_p$ is the full width at half maximum which can be calculated by integrating the intensity at the peak wavelength [51].

Table 7
Emission peak wavelength (λ_p) (nm), effective band widths ($\Delta\lambda_p$) (nm), stimulated emission cross-sections (σ_{se}) ($\times 10^{-20}$ cm²), gain band width ($\sigma_{se} \times \Delta\lambda_p$) ($\times 10^{-26}$ cm³), optical gain ($\sigma_{se} \times \tau_R$) ($\times 10^{-23}$ cm² s) and CIE co-ordinates (X, Y) for the emission transitions of Ho³⁺ doped ZnAlBiB glasses.

Transition	Spectral parameters	Glass A	Glass B	Glass C	Glass D	Glass E	Glass F
⁵ F ₅ → ⁵ I ₈ λ_p –623 nm	$\Delta\lambda_p$	7.81	6.25	4.69	6.25	7.81	9.38
	σ_{se}	8.22	14.5	24.1	13.7	8.77	4.93
	$\sigma_{se} \times \Delta\lambda_p$	0.64	0.90	1.13	0.85	0.68	0.46
	$\sigma_{se} \times \tau_R$	5.13	6.58	8.88	6.52	5.19	4.17
⁵ F ₅ → ⁵ I ₇ λ_p –880 nm	$\Delta\lambda_p$	9.38	7.81	6.25	7.81	7.81	9.38
	σ_{se}	9.08	17.0	30.0	16.5	12.9	7.47
	$\sigma_{se} \times \Delta\lambda_p$	0.85	1.33	1.87	1.29	1.01	0.70
	$\sigma_{se} \times \tau_R$	5.66	7.73	11.0	7.84	7.63	6.33
⁵ I ₇ → ⁵ I ₈ λ_p –1.91 μm	$\Delta\lambda_p$	13.1	11.6	11.1	12.6	13.1	13.6
	σ_{se}	0.99	1.34	1.85	1.11	0.92	0.76
	$\sigma_{se} \times \Delta\lambda_p$	1.29	1.56	2.06	1.47	1.21	1.04
	$\sigma_{se} \times \tau_R$	4.08	4.06	4.81	4.27	4.12	3.98
CIE co-ordinates (X, Y) for the transition ⁵ F ₅ → ⁵ I ₈		(0.59, 0.35)	(0.60, 0.34)	(0.62, 0.33)	(0.60, 0.30)	(0.60, 0.29)	(0.59, 0.28)

Table 8
Comparison of stimulated emission cross-section of Glass C with different hosts for the visible (623 nm) and mid infrared (1.91 μm) transitions.

Sample	⁵ F ₅ → ⁵ I ₈ λ_p –623 nm		⁵ I ₇ → ⁵ I ₈ λ_p –1.91 μm		References
	$\Delta\lambda_p$	σ_{se}	$\Delta\lambda_p$	σ_{se}	
Glass C	4.69	24.1	11.1	1.85	Present Work
LBTAHFO10	56.76	0.32	–	–	[24]
Glass A (Li)	–	2.31	–	–	[41]
Glass B (Na)	–	0.89	–	–	[41]
Glass C (K)	–	3.06	–	–	[41]
Glass D (Li–Na)	–	1.22	–	–	[41]
Glass E (Li–K)	–	2.21	–	–	[41]
Glass F (Na–K)	–	1.85	–	–	[41]
Glass G (Ca)	–	1.17	–	–	[41]
AH1	9	0.34	–	–	[52]
AH2	13	0.27	–	–	[52]
AH3	7	0.51	–	–	[52]
AH4	5	0.36	–	–	[52]
Lead phosphate	–	–	–	1.110	[39]
Fluoro zirco aluminate	–	–	–	0.398	[53]
Oxyfluoride tellrite	–	–	–	0.880	[44]
Germanate glass	–	–	–	0.469	[54]
Tellurite glass	–	–	–	0.880	[55]
PG	–	–	–	0.469	[56]
GCA	–	–	–	0.518	[56]
GCB	–	–	–	0.590	[56]
GCC	–	–	–	0.664	[56]
PbO–Bi ₂ O ₃ –Ga ₂ O ₃	–	–	–	0.544	[57,58]
Fluorophosphate	–	–	–	0.406	[59]
Oxyfluoride	–	–	–	0.520	[60]
Tellurite	–	–	–	0.915	[61]
Silicate	–	–	–	0.710	[62]

The large value of stimulated emission cross-section (σ_{se}) is an attractive feature for low-threshold and high gain applications to obtain continuous wave laser action and it is also used to identify the potential laser transitions of rare earth ions in any host matrix. The stimulated emission cross-section evaluated for all the ZnAlBiB glasses using Eq. (1) are given in Table 7 along with the emission peak wavelength (λ_p), effective band width ($\Delta\lambda_p$), gain band width ($\sigma_{se} \times \Delta\lambda_p$) and optical gain ($\sigma_{se} \times \tau_R$). From Table 7 it is observed that Glass C possesses lowest effective band width ($\Delta\lambda_p$) for all the transitions. The effective band width ($\Delta\lambda_p$) of the emission spectra is mainly caused by splitting of the levels of transitions and the site to site variation of the ligand field around Ho³⁺ ions in the glass i.e., inhomogeneous broadening. The decrease in effective band width implied that the asymmetry of the ligand field becomes weaker. Among all the glasses, Glass C possess highest stimulated emission cross-section (σ_{se}) for all the transitions ⁵F₅ → ⁵I₈ (623 nm), ⁵F₅ → ⁵I₇ (880 nm) and ⁵I₇ → ⁵I₈ (1.91 μm). Hence Glass C is suitable to emit red, near infrared and mid infrared lasers. Table 8 shows comparison of stimulated emission cross-section (σ_{se}) of Glass C for the transitions ⁵F₅ → ⁵I₈ (623 nm) and ⁵I₇ → ⁵I₈ (1.91 μm) along with the reported values

[24,39,41,52,53,44,54–62]. The gain band width ($\sigma_{se} \times \Delta\lambda_p$) and optical gain ($\sigma_{se} \times \tau_R$) parameters are critical to predict the amplification of the medium in which the rare earth ions are situated. A good optical amplifier should have larger stimulated emission cross-section (σ_{se}), gain band width ($\sigma_{se} \times \Delta\lambda_p$) and optical gain ($\sigma_{se} \times \tau_R$) values. From Table 7 it is evident that among all the glasses, Glass C possesses higher values of stimulated emission cross-section (σ_{se}), gain band width ($\sigma_{se} \times \Delta\lambda_p$) and optical gain ($\sigma_{se} \times \tau_R$). Reasonably higher values of these parameters for the transitions ⁵F₅ → ⁵I₈, ⁵F₅ → ⁵I₇ and ⁵I₇ → ⁵I₈ suggest that Glass C with 1 mol% of Ho³⁺ is suitable candidate for red, near infrared and mid infrared laser and also for optical amplification.

3.6. CIE chromaticity co-ordinates

The color coordinates for the emission spectra excited by 586 nm are calculated by using the CIE 1931 color matching functions [63,64] and are given in Table 7. Fig. 7 represents the CIE plot with color co-ordinates calculated from the emission spectra for 1 mol% of Ho³⁺ in ZnAlBiB glass (Glass C) excited at 586 nm wavelength. The emission spectra contain two parts. The first part is in

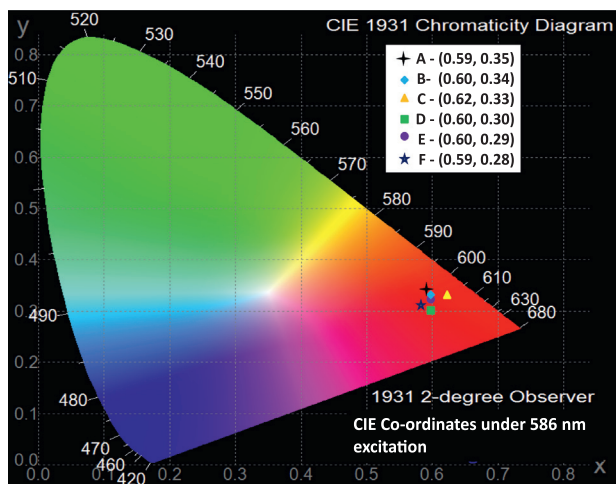


Fig. 7. CIE 1931 chromaticity diagram for different concentrations of Ho^{3+} doped ZnAlBiB glasses.

the bright red region and the second part is in the near infrared (NIR) region. The evaluated color coordinates for all the ZnAlBiB glasses falls in the red color region. Among all the glass samples, particularly Glass C color co-ordinates ($x = 0.6647$, $y = 0.3297$) falls in the bright red color region corresponding to $^5\text{F}_5 \rightarrow ^5\text{I}_8$ transition. Hence Glass C is best host to give bright red color laser at 623 nm.

4. Conclusions

The present study yields a complete picture of the spectroscopic characteristics such as J–O intensity parameters, radiative transition probabilities, radiative lifetimes, branching ratios, stimulated emission cross-sections and color co-ordinates of Ho^{3+} ions in ZnAlBiB glasses. From the magnitude of J–O intensity parameter (Ω_2) it is concluded that, among all the glasses Glass C possess highest Ω_2 parameter which indicates the higher covalence of Ho–O bond in Glass C. The luminescence spectra recorded at an excitation wavelength 586 nm exhibited two strong emission bands at 623 nm and 880 nm corresponding to the transitions $^5\text{F}_5 \rightarrow ^5\text{I}_8$ and $^5\text{F}_5 \rightarrow ^5\text{I}_7$ respectively. These two transitions $^5\text{F}_5 \rightarrow ^5\text{I}_8$ and $^5\text{I}_5 \rightarrow ^5\text{I}_7$ observed in red and near infrared regions are found to be more intense. The luminescence spectra recorded at an excitation wavelength 980 nm exhibited a strong emission band at 1.91 μm corresponding to the transition $^5\text{I}_7 \rightarrow ^5\text{I}_8$ falling in mid infrared region. For all the emission peaks observed the experimental branching ratios, stimulated emission cross-sections are measured and are found to be maximum for Glass C with 1 mol% of Ho^{3+} ions. Hence it is concluded that Glass C is suitable to emit laser in bright red region (623 nm), near infrared region (880 nm) and mid infrared region (1.91 μm). The CIE chromaticity co-ordinates were also evaluated from the emission spectra for all the glasses by exciting them at 586 nm to understand the feasibility of using these materials for Red emission. From the measured emission cross-sections and CIE chromaticity co-ordinates, it was found that 1 mol% of Ho^{3+} (Glass C) is quite suitable for the development of visible red, near infrared and mid infrared lasers.

Acknowledgements

This work is supported by High-impact Research initiative of IIT Delhi, UK-India Education Research Initiative (UKIERI). Two of the authors Mahamuda Shaik (File No: SR/WOS-A/PS-53/2011) and Swapna Koneru (File No: SR/WOS-A/PS-35/2011) are very much thankful to Department of Science and Technology (DST), Govern-

ment of India, New Delhi for awarding them with a Women Scientist's scheme under DST-WOS (A) programme.

References

- [1] A.A. Kaminskii, L.K. Aminov, V.L. Ermolaev, E.B. Sveshnikova, *Fizika I Spektroskopia Lazernykh Kristallov*, Physics and Spectroscopy of Laser Crystals, Moskva Nauka, 1986 (in Russian).
- [2] B. Henderson, R.H. Bartram, *Crystal-Field Engineering of Solid State Lasers Materials*, Cambridge University, 2000.
- [3] H.T. Yuen, J.C. Daniel, A.K. Terence, *Opt. Commun.* 231 (2004) 357–364.
- [4] G.J. Koch, M. Petros, J. Yu, U.N. Singh, *Appl. Opt.* 41 (2002) 1718–1721.
- [5] K. Scholle, E. Heumann, G. Huber, *Laser Phys. Lett.* 1 (2004) 285–290.
- [6] S.B. Rai, *Spectrochim. Acta A* 58 (2002) 1559–1566.
- [7] Y.G. Choi, J. Heo, J. Non-Cryst. Solids 217 (1997) 199–207.
- [8] N. Soga, K. Haurio, M. Yoshimoto, H. Yamamoto, *J. Appl. Phys.* 63 (1988) 4451–4454.
- [9] S.M. Kaczmarek, *Opt. Mater.* 19 (2002) 189–194.
- [10] W.H. Dumbaugh, J.C. Lapp, *J. Am. Ceram. Soc.* 75 (1992) 2315–2326.
- [11] D.W. Hall, M.A. New House, N.F. Borelli, W.H. Dumbaugh, D.L. Weidman, *Appl. Phys. Lett.* 54 (1989) 1293–1295.
- [12] J.E. Shielby, *J. Am. Ceram. Soc.* 71 (1998) C-254–C-256.
- [13] A.A. Kharlamov, R.M. Almeida, J. Heo, *J. Non-Cryst. Solids* 202 (1996) 233–240.
- [14] Sk. Mahamuda, K. Swapna, A. Srinivasa Rao, M. Jayasimhadri, T. Sasikala, K. Pavan, L. Rama Moorthy, *J. Phys. Chem. Solids* 74 (2013) 1308–1315.
- [15] C.K. Jayasankar, A.R. Devi, *Opt. Mater.* 6 (1996) 185–201.
- [16] C.H. Qi, X.R. Zhang, H.F. Hu, *Acta Opt. Sin.* 14 (1994) 583–593.
- [17] W. Koehner, *Solid State Laser Engineering*, Springer-Verlag, 1999.
- [18] T. Hoon Lee, J. Heo, *Phys. Rev. B* 73 (2006) 144201–1–144201–9.
- [19] M.R. Reddy, S.B. Raju, N. Veeraiah, *J. Phys. Chem. Solids* 61 (2000) 1567–1571.
- [20] S.B. Rai, A.K. Sing, S.K. Singh, *Spectrochim. Acta A* 59 (2003) 3221–3226.
- [21] A. Erbit, H.P. Jenssen, *Appl. Opt.* 19 (1980) 1729–1886.
- [22] A.S. Rao, Y.N. Ahammed, R.R. Reddy, T.V.R. Rao, *Opt. Mater.* 10 (1998) 245–252.
- [23] H. Ebendorff-Heidepriem, W. Seeber, D. Ehr, *J. Non-Cryst. Solids* 163 (1) (1993) 74–80.
- [24] T. Suhasini, B.C. Jamalaih, T. Chengaiah, J. Suresh Kumar, L. Rama Moorthy, *Physica B* 407 (2012) 523–527.
- [25] B.B. Kale, A. Jha, S.K. Apte, P.V. Adhyapak, D.P. Amalnerkar, *Mater. Chem. Phys.* 78 (2002) 330–336.
- [26] T.I. Suratwala, R.A. Steele, G.D. Wilke, J.H. Campbell, K. Takeuchi, *J. Non-Cryst. Solids* 263–264 (2000) 213–227.
- [27] I. ArulRayappan, K. Selvaraju, K. Marimuthu, *Physica B* 406 (2011) 548–555.
- [28] Viender Kundu, R.L. Dhilman, A.S. Maan, D.R. Goyal, *Adv. Condens. Mat. Phys.* 2008 (2008) 1–7.
- [29] Dinesh K. Kanchan, Harshvadan R. Panchal, *Turk. J. Phys.* 22 (10) (1998) 989–996.
- [30] V.C. Veeranna Gowda, C. Narayana Reddy, K.C. Radha, R.V. Anavekar, J. Etourneau, K.J. Rao, *J. Non-Cryst. Solids* 353 (2007) 1150–1163.
- [31] Petru Pasuta, Gheorghe Borodi, Engen Culea, *J. Mater. Sci. Mater. Electron.* 23 (2008) (10854–008-9734).
- [32] W.T. Carnall, P.R. Fields, K. Rajnak, *J. Chem. Phys.* 49 (1968) 4424–4442.
- [33] W.J. Chung, J. Heo, *J. Lumin.* 99 (2002) 73–77.
- [34] A. Jha, S. Shen, M. Naftaly, *Phys. Rev. B* 62 (2000) 6215–6227.
- [35] B.R. Judd, *Phys. Rev.* 127 (1962) 750–761.
- [36] G.S. Ofelt, *J. Chem. Phys.* 37 (1962) 511–520.
- [37] Sk. Mahamuda, K. Swapna, A. Srinivasa Rao, T. Sasikala, L. Rama Moorthy, *Physica B* 428 (2013) 36–42.
- [38] R.D. Peacock, *Struct. Bond.* Berlin 22 (1975) 83–122.
- [39] Ch. Srinivasa Rao, K. Upendra Kumar, P. Babu, C.K. Jayasankar, *Opt. Mater.* 35 (2012) 102–107.
- [40] T. Srikumar, M.G. Brik, Ch. Srinivasa Rao, N. Venkatramiah, Y. Gandhi, N. Veeraiah, *Physica B* 406 (2011) 3592–3598.
- [41] M. Seshadri, Y.C. Ratnakaram, D. Thirupathi Naidu, K. Venkata Rao, *Opt. Mater.* 32 (2010) 535–542.
- [42] G. Gao, G. Wanga, C. Yu, J. Zhang, L. Hu, *J. Lumin.* 129 (2009) 1042–1047.
- [43] L. Feng, J. Wang, Q. Tang, L.F. Liang, H.B. Liang, Q. Su, *J. Lumin.* 124 (2007) 187–194.
- [44] S. Nishibu, S. Yonezawa, M. Takashima, *J. Non-Cryst. Solids* 351 (2005) 1239–1245.
- [45] G. Vijaya Prakash, *Mater. Lett.* 46 (2000) 15–20.
- [46] Meng Wang, Lixia Yi, Guonian Wang, Hu Lili, Junjie Zhang, *Solid State Commun.* 149 (2009) 1216–1220.
- [47] A.F. Kirby, R.A. Palmer, *Inorg. Chem.* 20 (1981) 4219–4222.
- [48] K.B. Yalimiskii, N.K. Davidenko, *Coord. Chem. Rev.* 27 (1979) 223–273.
- [49] W.F. Krupke, *Phys. Rev.* 145 (1966) 325–337.
- [50] K. Swapna, Sk. Mahamuda, A. Srinivasa Rao, M. Jayasimhadri, T. Sasikala, L. Rama Moorthy, *Ceram. Int.* 39 (2013) 8459–8465.
- [51] B. Karthikeyan, S. Mohan, M.L. Baesso, *Physica B* 337 (2003) 249–254.
- [52] T. Satyanarayana, T. Kalpana, V. Ravi Kumar, N. Veeraiah, *J. Lumin.* 130 (2010) 498–506.
- [53] X. Zou, H. Toratani, *J. Non-Cryst. Solids* 195 (1996) 113–124.
- [54] Q.F. Lin, H.P. Xia, Y.P. Zhang, J.H. Wang, *J. Rare Earths* 27 (2009) 76–82.
- [55] G.J. Gao, G.N. Wang, C.L. Yu, J.J. Zhang, L.L. Hu, *J. Lumin.* 129 (2009) 1042–1047.
- [56] Jiajia Pan, Xu Rongrong, Ying Tian, Kefeng Li, Hu Lili, Junjie Zhang, *Opt. Mater.* 32 (2010) 1451–1455.

- [57] Y.B. Shin, J.N. Jang, J. Heo, *Opt. Quant. Electron.* 27 (1995) 379–386.
- [58] Y.B. Shin, H.T. Lim, Y.G. Choi, Y.S. Kim, J. Heo, *J. Am. Ceram. Soc.* 83 (2000) 787–791.
- [59] Ying Tian, Liyan Zhang, Suyu Feng, Xu Rongrong, Hu Lili, Junjie Zhang, *Opt. Mater.* 32 (2010) 1508–1513.
- [60] V.K. Tikhomirov, J. Mendez-Ramos, V.D. Rodriguez, D. Furniss, A.B. Sedden, *J. Alloys. Compd.* 436 (2007) 216–220.
- [61] G.X. Chen, Q.Y. Zhang, G.F. Yang, Z.H. Jiang, *J. Fluoresc.* 17 (2007) 301–307.
- [62] B. Peng, T. Izumitani, *J. Opt. Mater.* 4 (1995) 797–810.
- [63] W.M. Yen, S. Shionoya, *Phosphor Hand Book*, CRC Press, New York, 1999, p. 805.
- [64] F.W. Billmeyer, M. Saltzman Jr., *Principles of color technology*, Wiley, New York, 1981.