Current Awareness Bulletin

of

SCHOLARLY ARTICLES PUBLISHED

by

Faculty, Students and Alumni

~ August 2012 ~

DELHI TECHNOLOGICAL UNIVERSITY CENTRAL LIBRARY
(formerly Delhi College of Engineering, Bawana Road, DELHI)

PREFACE

This is the first Current Awareness Bulletin Service 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 either the full text 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 August 2012. 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 has been provided in this Bulletin.

CONTENTS

14. Realization Of An 8-bit Pipelined Microprocessor in Verilog HDL by ***@Jayant Chowdhary(Corresponding Author), @Vivek Garg*** and ***@ Tushar Negi Shreya Jain***

15. Role of e-Agriculture in Rural Development in Indian Context *by Deka Ganesh Chandra* and ****Dutta Borah Malaya.***

16. Segmented Cladding Fiber Design for Femtosecond Laser Pulse Delivery at 1550-nm and 1064-nm Wavelengths by *Babita, Vipul Rastogi* and ****Ajeet Kumar.***

17. Sentiment Analysis on Twitter by ***@Akshi Kumar*** and ***@Teeja Mary Sebastian.***

18. System modeling and analysis of a combined cycle power plant by *Nikhil Dev, Samsher* and ***@ S. S. Kachhwaha***

***** **Faculty**
@ **Students/Research Scholars**
# **Alumni**

A Defect Prediction Model for Open Source Software

Ruchika Malhotra

***Abstract*- Defect prediction models are significantly beneficial for software systems, where testing experts need to focus their attention and resources on problematic areas in the software under development. In this paper we find the relation between object oriented metrics and fault proneness using logistic regression method. The results are analyzed using open source software. The performance of the predicted models is evaluated using Receiver Operating Characteristic (ROC) analysis. The results show that Area under Curve (calculated by ROC analysis) of the predicted model is 0.829.**

Index Terms— **Object oriented Metrics, Software quality, Empirical validation, Fault prediction, Receiver Operating Characteristics analysis**

I. INTRODUCTION

Application of software quality models early in the software development life cycle contributes to efficient defect removal and results in delivering more reliable software products. Empirical studies for predicting defects have been carried out in past and have stressed on the need to carry out more such studies in order to provide strong evidence in this important area.

Several metrics have been proposed in the literature to capture the OO design and code, constructs for example, (Aggarwal et al. [1]; Briand et al., [2, 3]; Bieman and Kang [4]; Cartwright and Shepperd [5]; Chidamber and Kemerer [6]; Harrison et al. [7]; Henderson-sellers [8]; Hitz and Montazeri [9]; Lake and Cook [10]; Li and Henry [11]; Lee et al. [12] Lorenz and Kidd [13]; Tegarden et al [14]).

These metrics provide ways to assess the quality of software and their use in early phases of software development can help software companies in evaluating large software development quickly and at a reasonable cost [1].

Manuscript received March 22, 2011.

Dr. Ruchika Malhotra (Corresponding Author phone: 91-9910290445) is with Department of Software Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India, (email: ruchikamalhotra2004@yahoo.com)

There have been empirical studies evaluating the impact of OO metrics on faulty classes such as (Aggarwal et al. [15], Singh et al. [16-18]; Basili et al. [19]; Binkley and Schach [20]; Briand et al [21]; Cartwright and Shepperd [5]; El Emam et al. [22]; Gyimothy et al. [23]; Zhou et al. [24]).

The work described in this paper focuses on the use of Object Oriented (OO) metrics in predicting defect prone classes. Our results are based on open source software Ant 1.7 developed using java language [25]. Although the open source software has achieved an acceptable level of quality, but there is more to be done in order to outperform proprietary software [26]. Hence, the defect prediction models can help in improving the quality and reducing faulty classes in the open source software. The validation of the methods is carried out using Receiver Operating Characteristic (ROC) analysis. Hence, the study is divided into the following parts:

1. Descriptive statistics and outlier analysis is performed to extract useful metrics and remove irrelevant data points.
2. Multivariate logistic regression is used for model prediction and validation.
3. Defect prediction model is evaluated using performance measures including ROC analysis.

The model constructed to predict faulty classes may help in focusing testing and inspection resources on the defect prone parts of the design and code in a cost effective manner.

The paper is organized as follows: Section 2 summarizes the OO metrics studied. Section 3 describes sources from which data is collected. The results of the study are given in section 4 and the model is evaluated in section 5. Section 6 presents threats to validity of the models and the conclusions of the research are presented in section 7.

II. METRICS USED

The binary dependent variable in our study is fault proneness. Fault proneness is defined as the probability of fault detection in a class [17]. We use logistic regression, which is based on predicting probabilities. For this study, we predict fault prone classes from object oriented metrics. The metrics are collected by using a tool for calculating Chidamber and Kemerer java metrics (ckjm). The tool is available at http://gromit.iiar.pwr.wroc.pl/p_inf/ckjm/metric.html. The metrics are given in table I [6, 7, 27].

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
METRICS USED IN THE STUDY

Metric	Definition
Coupling between objects (CBO)	CBO for a class is a count of the number of other classes to which it is coupled and vice versa.
Lack of cohesion (LCOM)	Measures the dissimilarity of methods in a class by looking at the instance variable or attributes used by methods.
Number of children (NOC)	The number of immediate subclasses of a class in a hierarchy.
Depth of inheritance (DIT)	The depth of a class within the inheritance hierarchy is 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)	A count of sum of complexities of all methods in a class.
Response for a class (RFC)	A set of methods that can be potentially executed in response to a message received by an object of that class.
Number of public methods (NPM)	The count of number of public methods in a class.
Afferent couplings (Ca)	It counts how many other classes use a given class.
Lack of cohesion in methods (LCOM3).	$LCOM1 = \frac{\frac{1}{N}\sum_{i=1}^{n}\mu(D_i) - m}{1-m}$
Data Access Metric (DAM)	It is defined as number of private methods divided by total number of methods.
Measure of Aggregation (MOA)	It counts abstract data types in a class.
Measure of Functional Abstraction (MFA)	It is defined as number of inherited methods divided by total number of methods accessible by its member functions.
Cohesion Among Methods of Class (CAM)	It is based upon parameters list of a method.
Inheritance Coupling (IC)	It is based upon inheritance based coupling.
Coupling Between Methods (CBM)	It counts the newly added functions with which inherited based methods are coupled.
Average Method Complexity (AMC)	It counts average size of method in a class.
Cyclomatic Complexity (CC)	CC= e-n+P, where e= number of edges in a flow graph, n=number of nodes in a flow graph, p= connected components
Lines of code (LOC)	The count of lines in the text of the source code excluding comment lines

III. EMPIRICAL DATA COLLECTION

In this study Apache Ant 1.7 open source software is used [25] for finding relationship between OO metrics and fault proneness. ANT 1.7 is a command-line tool that allows to build, compile, test and run Java applications. Ant 1.7 is developed using the Java language consisting of 745 classes. The development period of this version was December 2006 to September 2009. The details of the data set are summarized in table II.

TABLE II
DATA USED IN THE STUDY

System	Apache Ant
Language	Java
Total Classes	745
% faulty classes	22.28
Total faults	338

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IV. ANALYSIS RESULTS

In this section, we described the analyses performed to find the relationship between OO metrics and fault proneness of the classes. We employed multivariate logistic regression analysis. The multivariate analysis is used to find the combined effect of OO metrics on fault proneness. The models predicted were applied to Apache Ant 1.7 data set consisting of 745 classes. Descriptive statistics and outlier analysis was performed to find the irrelevant data. The following measures are used to evaluate the performance of each predicted fault proneness model in the above sub sections [17]:

- Sensitivity and Specificity, Completeness, Precision, Receiver Operating Characteristic (ROC) analysis [28].
- In order to predict the accuracy of the model it should be applied to different data sets. We therefore performed k-cross validation of models [29]. The data set is randomly divided into k subsets. Each time one of the k subsets is used as the test set and the other k-1 subsets are used to form a training set. Therefore, we get the fault proneness for all the k classes.

A. Descriptive statistics

To perform research analysis, the data should be preprocessed by removing irrelevant and unnecessary attributes that have less than six data points and eliminating outliers from the data set. Thus, we obtained descriptive statistics and performed outlier analysis on the data set. The descriptive statistics of the data set are shown in tablle III.

B. Research Methodology

Logistic Regression (LR) is used to predict the dependent variable (fault proneness) from a set of independent variables (OO metrics) to determine the percent of variance in the dependent variable explained by the independent variable (a detailed description is given by [30]. LR is of two types [17]: a) Univariate LR b) Multivariate LR.

C. Multivariate LR Results

In this section, we summarize the results obtained from multivariate fault prediction model using LR method. The multivariate analysis is used to find the combined effect of OO metrics (explained above) on fault proneness. We attempted to use the backward elimination method. However, the results of the model obtained were poorer (i.e. the values of R^2 statistic were low) than the model obtained from the forward stepwise procedure. We therefore used the forward stepwise procedure in this study. The conditional number is below 30 for the model predicted. This implies that the multicollinearity of the predicted model is tolerable. Table IV provide the coefficient (B), standard error (SE), statistical significance (sig), odds ratio (exp(B)) for OO metrics included in the model. Table IV shows that two metrics, RFC and CC, are included in the predicted model. The results of model accuracy are summarized in table V. The AUC is 0.834, hence the accuracy of the model predicted is very high.

TABLE III

MODEL STATISTICS

Metrics	Min.	Max.	Mean	Std. Deviation
WMC	0	120	11.07114	11.97596
DIT	1	7	2.522148	1.398869
NOC	0	102	0.731544	4.800357
CBO	0	499	11.04698	26.34315
RFC	0	288	34.36242	36.02497
LCOM	0	6692	89.14765	349.9376
CA	0	498	5.655034	25.81422
CE	0	37	5.746309	5.653176
NPM	0	103	8.365101	9.331319
LCOM3	0	2	1.013342	0.619015
LOC	0	4541	280.0711	411.8721
DAM	0	1	0.644855	0.438138
MOA	0	11	0.726174	1.426581
MFA	0	1	0.509968	0.398696
CAM	0	1	0.474685	0.259931
IC	0	5	0.720805	0.938948
CBM	0	19	1.312752	2.332602
AMC	0	2052	23.64087	76.98608
CC	0	53	4.669799	6.276853

TABLE IV

MODEL STATISTICS

Variable	B	S.E.	Sig.	Exp(B)
RFC	0.032	0.004	0.000	1.032
CC	0.075	0.022	0.001	1.077
Constant	-2.926	0.188	0.000	0.054

TABLE V

RESULTS OF MODEL PREDICTED

Measures	Results
Cutoff	0.17
Sensitivity	75.90
Specificity	75.50
Precision	75.06
AUC	0.834
SE	0.018

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. MODEL EVALUATION USING ROC ANALYSIS

In this section, we present the results of model evaluation.

D. Model Evaluation

The accuracy of the models predicted is somewhat optimistic since the models are applied on same data set from which they are derived. To predict accuracy of the model it should be applied on different data sets thus we performed 10-cross validation of the models. For the 10-cross validation, the classes were randomly divided into 10 parts of approximately equal data points. We summarized the results of cross validation of predicted models via the LR approach in Table VI. Table VI shows that AUC calculated by ROC analysis is very high 0.829. Thus, the accuracy of model obtained from validation data is similar to the accuracy of model obtained from training data. The values of sensitivity, specificity and precision are also high. The researchers and software practitioners can use the model predicted in early phase of software development. This will reduce testing effort and resources.

TABLE VI

RESULTS OF 10-CROSS VALIDATION OF MODEL

Measures	Results
Cutoff	0.18
Sensitivity	75.30
Specificity	75.50
Precision	75.04
AUC	0.829
SE	0.018

The ROC curve for the model given in table VI is shown in fig. 1.

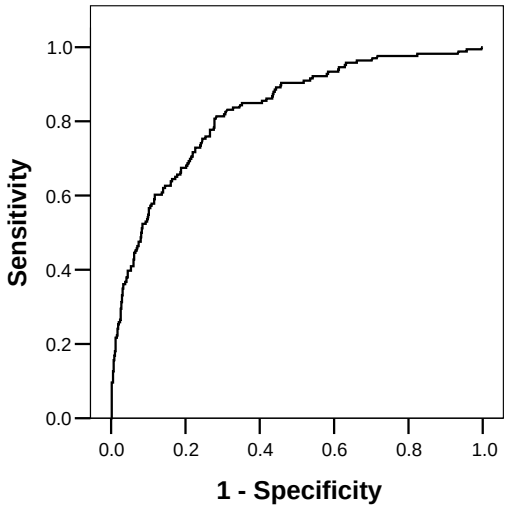


Fig. 1: ROC curve of LR Model

VI. THREATS TO VALIDITY

The study has many limitations that are common with most of the empirical studies in literature. However, it is necessary to repeat them here.

The usefulness of OO metrics for predicting fault proneness models also depends on the programming language (e.g. C++, Java) being used. Thus, similar studies with different data sets are required to be carried out in order to establish the acceptability of the model.

Our conclusions are pertinent to only dependent variable fault proneness, as it seems to be most popular dependent variable in empirical studies. We do not claim about the validity of the chosen OO metrics in this study when the dependent variable changes like maintainability or effort [17].

VII. CONCLUSION

The goal of this work is to find the effect of OO metrics on fault proneness. We also empirically analyze the performance of logistic regression method in order to predict faulty classes.

Based on the results obtained from open source software Apache Ant 1.7 data set we analyzed the performance of predicted defect model using the ROC analysis. We analyzed the OO metrics including metrics given by Chidamber and Kemerer and McCabe. Our main results are summarized as follows:

Two metrics RFC and CC were included in the model. The model predicted has higher accuracy with AUC 0.829. Thus, the model predicted shows that the model predicts faulty classes of open source software with good accuracy.

This study confirms that construction of model using logistic regression method will be effective, adaptable, and useful in predicting fault prone classes.

We plan to replicate our study to predict models based on machine learning algorithms such as support vector machines and genetic algorithms. We may carry out cost benefit analysis of models that will help to determine whether a given fault proneness model would be economically viable.

REFERENCES

[1]. K.K.Aggarwal, Yogesh Singh, Arvinder Kaur, Ruchika Malhotra, “Software Reuse Metrics for Object-Oriented Systems”, Third ACIS Int'l Conference on Software Engineering Research, Management and Applications (SERA'05), IEEE Computer Society, pp. 48-55, 2005.

[2]. L..Briand , W.Daly and J. Wust, “Unified Framework for Cohesion Measurement in Object-Oriented Systems”, Empirical Software Engineering, vol. 3, pp.65-117, 1998.

[3]. L.Briand , W.Daly and J. Wust, “A Unified Framework for Coupling Measurement in Object-Oriented Systems. IEEE Transactions on software Engineering”, Vol. 25, pp.91-121, 1999.

[4]. J.Bieman, B.Kang, “Cohesion and Reuse in an Object-Oriented System”, Proc. ACM Symp. Software Reusability (SSR’94), pp.259-262, 1995.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[5]. M.Cartwright, M.Shepperd, "An Empirical Investigation of an Object-Oriented Software System", IEEE Transactions of Software Engineering, 1999.
[6]. S.Chidamber and C.F.Kemerer, "A metrics Suite for Object-Oriented Design", IEEE Trans. Software Engineering, vol. SE-20, no.6, 476-493, 1994.
[7]. B.Henderson-sellers, "Object-Oriented Metrics, Measures of Complexity", Prentice Hall, 1996.
[8]. R.Harrison, S.J.Counsell, and R.V.Nithi, "An Evaluation of MOOD set of Object-Oriented Software Metrics", IEEE Trans. Software Engineering, vol. SE-24, no.6, pp. 491-496, June 1998.
[9]. B.Henderson-sellers, "Object-Oriented Metrics, Measures of Complexity", Prentice Hall, 1996.
[10]. M.Hitz, B. Montazeri, "Measuring Coupling and Cohesion in Object-Oriented Systems", Proc. Int. Symposium on Applied Corporate Computing, Monterrey, Mexico, 1995.
[11]. A.Lake, C.Cook, "Use of factor analysis to develop OOP software complexity metrics". Proc. 6th Annual Oregon Workshop on Software Metrics, Silver Falls, Oregon, 1994.
[12]. W.Li, S.Henry, "Object-Oriented Metrics that Predict Maintainability", Journal of Systems and Software, vol 23 no.2, pp.111-122, 1993.
[13]. Y.Lee, B.Liang, S.Wu and F.Wang, "Measuring the Coupling and Cohesion of an Object-Oriented program based on Information flow", 1995.
[14]. M.Lorenz, and J.Kidd, "Object-Oriented Software Metrics", Prentice-Hall, 1994.
[15]. D.Tegarden, S. Sheetz, D.Monarchi, "A Software Complexity Model of Object-Oriented Systems. Decision Support Systems", vol. 13, pp.241-262.
[16]. K.K. Aggarwal, Y. Singh, A. Kaur, R. Malhotra, "Empirical Analysis for Investigating the Effect of Object-Oriented Metrics on Fault Proneness: A Replicated Case Study", *Software Process Improvement and Practice*, John Wiley & Sons, vol. 16, no. 1, pp. 39-62, 2009.
[17]. Y. Singh, A. Kaur, and R. Malhotra, "Empirical validation of object-oriented metrics for predicting fault proneness models", *Software Quality Journal*, vol. 18, pp.3-35, Jan 2010.
[18]. Y. Singh, A. Kaur, and R. Malhotra, "Predicting Software Fault Proneness Model Using Neural Network", *Product-Focused Software Process Improvement*, *Lecture Notes in Computer Science*, pp. 204-214, 2008.
[19]. Y. Singh, A. Kaur, and R. Malhotra, "Application of Decision Trees for Predicting Fault Proneness", *International Conference on Information Systems, Technology and Management-Information Technology*, Ghaziabad, India, 2009.
[20]. V.Basili, L.Briand, W.Melo, "A Validation of Object-Oriented Design Metrics as Quality Indicators", IEEE Transactions on Software Engineering, vol. 22 no.10, pp. 751-761, 1996.
[21]. L. Briand, W. Daly, and J. Wust, "Exploring the relationships between design measures and software quality", *Journal of Systems and Software*,, vol. 51, no. 3, pp. 245-273, 2000.
[22]. K. El Emam, S. Benlarbi, N. Goel, and S. Rai, "A Validation of Object-Oriented Metrics", *Technical Report ERB-1063*, NRC, 1999.
[23]. T.Gyimothy , R.Ferenc , I.Siket , "Empirical validation of object-oriented metrics on open source software for fault prediction", IEEE Trans. Software Engineering, vol. 31, Issue 10, pp.897 – 910, Oct. 2005.
[24]. Y. Zhou, and H. Leung, "Empirical analysis of Object-Oriented Design Metrics for predicting high severity faults", *IEEE Transactions on Software Engineering*, vol. 32, no. 10, pp. 771-784, 2006.
[25]. promise. Available at http://promisedata.org/repository/
[26]. I. Samoladas, and I. Stamelos, "Assessing Free/Open Source Software Quality", Department of Informatics, Aristotle University of Thessaloniki, Greece.
[27]. T. McCabe, "A Complexity Measure", *IEEE Transactions on Software Engineering*, vol. 2, no. 4, pp. 308-320, 1976.
[28]. J. Hanley, BJ. McNeil, "The meaning and use of the area under a Receiver Operating Characteristic ROC curve", *Radiology*, vol. 143, pp.29-36, 1982.
[29]. M. Stone, "Cross-validatory choice and assessment of statistical predictions", *J. Royal Stat. Soc.*, vol. 36, pp. 111-147, 1974.
[30]. D. Hosmer, S. Lemeshow, Applied Logistic regression, John Wiley and Sons 1989.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Wireless Pers Commun
DOI 10.1007/s11277-012-0720-5

Adaptive Congestion Controller for ABR Traffic in ATM Network

R. L. Ujjwal · C. S. Rai · Nupur Prakash

Abstract One of the imperatives for improving the quality-of-service (QoS) of high-speed networks of next generation is to reduce congestion. Congestion occurs when the resource demand exceeds the capacity of the network. The purpose of this paper is to analyze the performance of Robust Adaptive Congestion Control algorithm for available bit rate traffic in asynchronous transfer mode networks and suggest improvements of the algorithm for better performance and fairness. The results of simulation presented here corroborates the fact that the improved algorithm reduces settling time and cell loss ratio and thus maximizes utilization of the network.

Keywords ATM network · Congestion · Settling time · Bandwidth

Abbreviations

ATM	Asynchronous transfer mode
ABR	Available bit rate
RM	Resource management
PCR	Peak cell rate
MCR	Minimum cell rate
ACR	Allowed cell rate
ICR	Initial cell Rate
ER	Explicit rate
NI	No increment
CI	Congestion indication
N_{rm}	Number of cells between FRM cells
RIF	Rate increase factor
RDF	Rate decrease factor
VC	Virtual circuit

R. L. Ujjwal (✉) · C. S. Rai · N. Prakash
University School of Information and Communication Technology,
GGS Indraprastha University, Delhi 110075, India
e-mail: ujjwal_rl@rediffmail.com

Published online: 01 August 2012

1 Introduction to ATM Network

In future high speed networks are needed to offer emerging applications such as high definition TV and real-time video. The Asynchronous Transfer Mode (ATM) plays an important role in which the information is transmitted using short fixed-size cells consisting of 53 bytes long [1,2]. The fixed size of the cells reduces the variance of delay making the networks suitable for integrated traffic. En efficient congestion control mechanism in ATM network needed to provide quality-of-service to the users. ATM is a packet-switching technology enabling Broadband-ISDN. There are different types of communication services defined by ATM forum [3–5].

i. CBR (constant bit rate): The cell rate is constant through-out the connection.
ii. VBR (variable bit rate): It allows users to send at a variable rate.
iii. ABR (available bit rate): Designed for normal data traffic such as file transfer and email.

2 Management Model for ABR Traffic

Traffic management is concerned with ensuring that users get their desired quality of service. It is difficult during heavy traffic in the network. In ATM network, the ABR service is served only if there is some bandwidth left by CBR service or/and VBR service which get the higher priority over ABR [6,7]. In the ABR service, the source adapts its rate to changing network conditions. Information about the state of the network such as bandwidth availability, state of congestion, and impending congestion is conveyed to the source through special control cells called Resource Management (RM) cells which contains some special fields including the minimum cell data (MCR), the peak cell rate (PCR), and the explicit rate (ER). The RM cells travel to the destination and returned to the source in the same path. [11,12].

The RM cells are generated by source and send in forward direction to the destination after every N_{rm} data cells. In ATM network, the value of N_{rm} is 32 [3,8]. The travel path of RM cells and data cells are same but purpose of RM cell is to provide the information about network and basic of these information sources update the data rates. The basic management model for ABR traffic is shown in Fig. 2. When the switch receives the RM cell it treated specially and may perform the following operations: (a) update the explicit rate (ER) field of RM cells directly. (b) Indirectly inform the source about congestion by setting the explicit

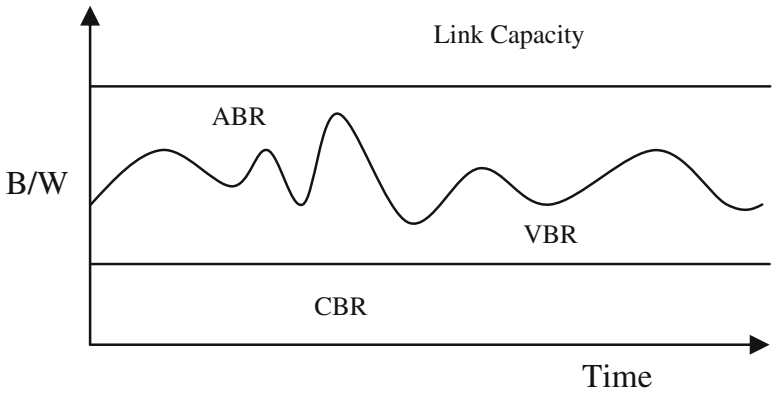


Fig. 1 Available ABR bandwidth

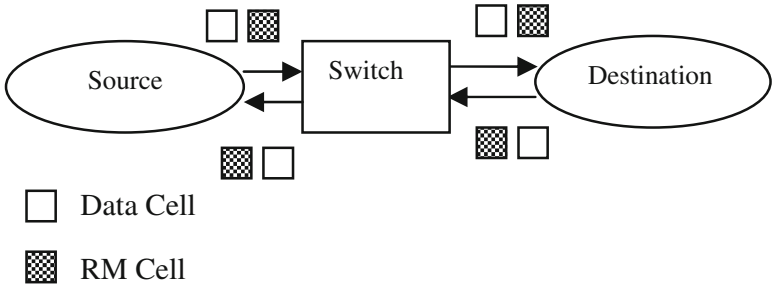


Fig. 2 Management model for ABR traffic

forward congestion indication bit in the data cell header, and rely on the destination to convey congestion information back to the source by marking the CI bit in the backward RM cells it generates. (c) Set or reset the congestion indication (CI) or no increase (NI) bit in the RM cells according to the status of network. (d) Generate backward RM cells and send to the source spontaneously [11,12].

When an ABR connection established, the source specifies to the network both a maximum required bandwidth and a minimum usable bandwidth. These are designated as peak cell rate (PCR) and the minimum cell rate (MCR), respectively. Sources should always transmit at a rate equal to or below their computed allowed cell rate (ACR). The ACR cannot exceed PCR and need not go below MCR. It means source rate $\leq$ACR and MCR $\leq$ ACR $\leq$ PCR. The bandwidth available from the network may vary, but should not become less than MCR. The ER field of RM cell is set by source at the beginning and sends out in forward direction. The RM cells passes through the many switches and fields of RM cell may be updated according network state such as ER field and CI and NI bits [9–11]. When the source receives the RM cell back, it takes one of the following actions: (a) if the CI bit is set, it decreases its ACR by a quantity RDF $\times$ PCR, where RDF is the rate decrease factor. (b) If both CI and NI bits are not set it means no congestion in the network and ACR can be increased by a quantity RIF $\times$ PCR, where RIF is the rate increase factor but not above the explicit rate specified in the ER field. (c) Finally, if only the NI bit set, it sets its rate to the minimum of ACR and ER [3,6,12].

3 ATM Switch Model

3.1 Switch Functions

In the ATM switching system, two functions have been implemented. The first function is the space switching function. The space switching is the one, which allows the connection between every input and every output. An important aspect of space switching is the internal routing. This means how the information is routed internally in the switch. The internal structure of the switch must allow connections between every input to every output. The second function is time switching. Since ATM is working in an asynchronous mode, cells that had arrived in various time slots from the different inputs can be delivered from different outputs in different time slots. Since there is no pre-assigned time slot connection, a connection problem arises if more than two logical channels are connected to the same output at the same time slot. Implementing a queuing function in the ATM switch system solves the problem in the ATM switch [6,12].

3.2 RM Cell Description

The RM cell is a special cell which using for traffic management in ATM network. It provides state information of network to the sources. The RM cell has the regular ATM Header of five bytes. The payload type indicator field is set to 110 to indicate that the cell is an RM cell. The protocol id field, which is one byte long, is set to one for ABR connections. It has following fields: (a) Direction (D) bit indicates which direction of data flow is associated with the RM. (b) Backward Notification (BN) bit indicates whether the RM cell is a Backward. (c) Congestion Indication (CI) bit when set asks the source to decrease its ACR immediately. (d) No Increase (NI) bit when set asks the source not to increase its ACR. (e) Explicit Cell Rate (ECR) field carries the rate feedback from the switches. (f) Minimum Cell Rate (MCR field carries the connection's negotiated Minimum Cell Rate. (g) Current Cell Rate (CCR) field is set by the source. (h) CRC: Cyclic redundancy checks [7–9].

3.3 Modeling the Switch

An ATM switch contains a set of input and output ports, which are utilized to interconnect users, other switches, and other network elements. The main function of an ATM switch is to buffer cells and quickly relay them without cell loss. It also performs control and management functions in order to support both asynchronous and synchronous traffic as well as connectionless and connection-oriented traffic. The control function deals with call establishment and release of virtual path/virtual channel connections and the Management function ensure correct and efficient network operation [11]. In the ATM network, each node has an output buffer to prevent cell loss. The queue length of cells should be limited to avoid the overflow because of fixed buffer size. Let q_n denote the queue length at time n and q_{n+1} queue length at time n+1. The linear equation of queue length is following.

$$q_{n+1} = q_n + r_n - \mu_n. \tag{1}$$

where r_n denote the total number of cells that arrive in one of the output buffers of an ATM switch in the interval $[n, n+1]$, and let μ_n denote the number of cells that depart from this buffer in the same time interval [12].

Suppose there are N connections and the number of cells that arrive from m different sources during time-slot [n, n+1] be denoted by r_{mn}, clearly

$$r_n = \sum_{m=1}^{N} r_{mn} \tag{2}$$

Where r_{mn}have two components, first is number of cells receiving from controlled sources and second is number of cells receiving from uncontrolled sources denoted by r_n^u. Let r_{mn}^c the number of cells that arrive from controlled source m in the time interval [n, n+1] and total number of controlled source is M. Then we can find r_n by [6,12]

$$r_n = \sum_{m=1}^{M} r_{mn}^c + r_n^u \tag{3}$$

The fairness among sources is necessary in the ATM networks. The ABR congestion control may provide this with the minimum cell loss ratio and maximum utilization of network resources. The bandwidth may be fairly divided if the switch knowing number of controlled

sources that are being-served at a given time. The fair share of each VC is computed as follows [3,6].

$$Fairshare = \frac{ABR\_Capacity}{Number\_of\_active\_sources} \tag{4}$$

In the case of congestion at any ATM switch, the time is taken by sources to handle the situation because it takes time from the moment the ER decision is made by the switch and take action by sources. Thus, the cell rate of source m at time n, r_{mn}, is actually an outcome of an action taken d_m time units earlier, where d_m represents the action delay (the time from the moment control information is sent to a source, until an action is taken by it, and until subsequently that action affects the state of the switch that initiated that command) for source m and is taken to be independent of time n. Thus, we have $r^c_{mn} = C_n - d_m$ Where C_n denotes the command issued by the switch at time n and $m = 1, 2, \ldots, M$. If the maximum round-trip delay is d_M (from source to destination and destination to source) then minimum and maximum limits for these delays is: $0 \leq d_1 \leq d_2 \leq \ldots. \leq d_m \leq d_M$ [12].

In any high speed network, delay is an important factor to design of rate flow controllers. The end-to-end round-trip delay may be computed at the time of VC but there is variability in the delays due to queues in the virtual circuits. The RM cell is generated after every 32 data cells and hence feedback is not instantaneous. The calculation of M (actual number of controlled sources) is also a tough task. Hence, the switch must have a method of estimating this parameter.

After the observation of these we have realized the need to develop a controller that is adaptive and robust. The values of μ_n and r^u_n varies with the time and both may be represented by μ and r^u respectively. With these assumptions, the Eq. (1) becomes as below [3,12].

$$q_{n+1} = q_n + \sum_{m=1}^{M} {r_{mn}}^c + r^u - \mu \tag{5}$$

where m_i is number of controlled sources.

4 Literature Reviews

An efficient explicit-rate switch algorithm with max-min fairness for ABR service class in ATM networks has been designed [6] by H. Oshsaki, M. Murata, and H. Miyahara. In this algorithm authors focuses on explicit-rate marking switch and enhance to control the queue length of the switch for preventing cell loss and achieving link utilization. The performance has been evaluated of proposed algorithm that is achieves better efficiency and stability. A rate-based congestion control algorithm is a closed-loop control method suitable for data transfer applications. In the rate-based congestion control algorithm, cell transmission rates of source end systems are regulated according to congestion information returned by the network. The source end system periodically sends a forward RM cell per N_{rm} data cells, and the destination end system sends it back to the corresponding source end system as a backward RM cell. The switch notifies its congestion to source end systems by marking an EFCI (Explicit Forward Congestion Indication) bit of data cells or a CI bit of RM cells. Since it uses one-bit information, the switch utilizing the EFCI bit or the CI bit is often referred to as a binary-mode switch. In the standard, the switch is allowed to explicitly designate the cell transmission rate by modifying an ER value of

the RM cell. This sort of switch is called as explicit-rate switch. In paper [3], author design an adaptive and roust congestion controller for ATM networks. The controller minimize the rate of stochastic inputs from uncontrollable high priority sources and avoids the computations needs for pole placement design of the minimum variance controller, and utilize an online resources algorithm in direct tuning of the controller parameters. It is provide fair allocation of resources, minimum cell less ratio, and efficient use of networks.

In paper [12], the author Imer, baser, and srikant also proposed a robust adaptive algorithm for ABR congestion control in ATM network that is performs well under various criteria, such as max- min fairness, PCR and MCR constraints. This algorithm does not suffer from computational complexity.

5 Modification

In the existing algorithm the controller updates sources every $(d_M + 1)$ time units, where d_M is maximum possible round-trip delay and its value is ten time units [12]. Suppose the congestion occurs at a switch, it will send the information about congestion to all the sources in $(d_M + 1)$ time units. After $(d_M + 1)$ time units it will update sources and then will do necessary computations. In the modified algorithm the switch replaces $(d_M + 1)$ by $(d_1 + 1)$ whenever the congestion occurs. d_1 is the time during which all sources receive information about congestion sent by the switch through RM cells and d_1 is always less than d_M. The value of d_1 is assumed seven time units. So the switch will update $(d_M - d_1)$ times units earlier. As a result all the sources will change their rates $(d_M - d_1)$ times units earlier in each step. So the congestion state at the switch will change $(d_M - d_1)$ times units earlier in every step. Therefore the cell rates will reach their stable value much earlier. Let a RM cell starts traveling form source towards its destination. As soon as this RM cell will pass through the congested switch, the switch will set the CI bit in this cell and also send the value of d_1 to all the switches, which will be on the path towards the destination. To send this information the switch may spontaneously generate the RM cell and send that cell towards the destination or it may put the value of d_1 in the RM cell generated by source.

6 Design and Implementation

6.1 Simulation Model

We considered a high speed ATM network shown in Fig. 3, which has four ATM switches (nodes) SW0, SW1, SW2, and SW3, four sources S0, S1, S2, and S3, and also four destinations D0, D1, D2, and D3. The distance between two adjacent nodes is constant and equal to 1,000 Km. The link speed is the speed of light and the service offered by link is 1 Gbps. Time unit is taken to be the time to serve 5,000 cells (i.e. $5{,}000 \times 53 \times 8$ bits) [12]. The one-way propagation delay from one node to another is 1.6 time units. This model is used to evaluate the fairness of the method when there is only one bottleneck node. In this network ABR sources with different delays are bottlenecked at the SW2 node. The value of d_M is equal to 10, which is largest possible network delay in this network [12].

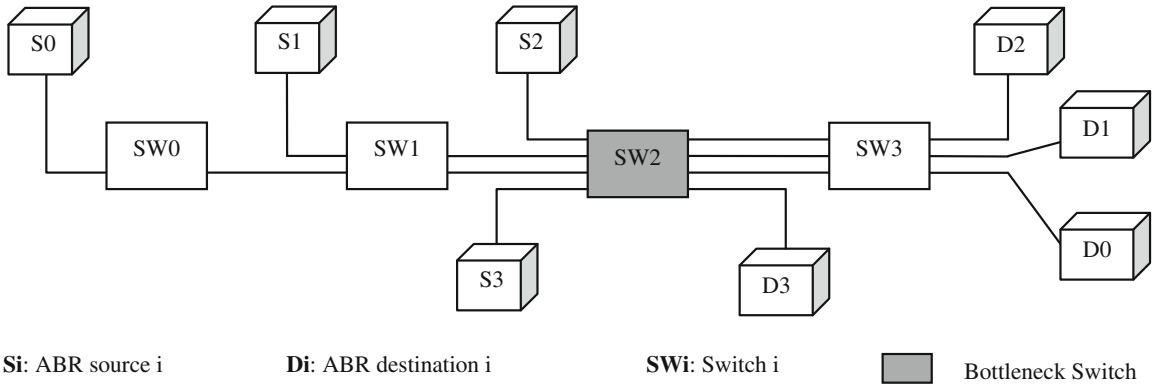
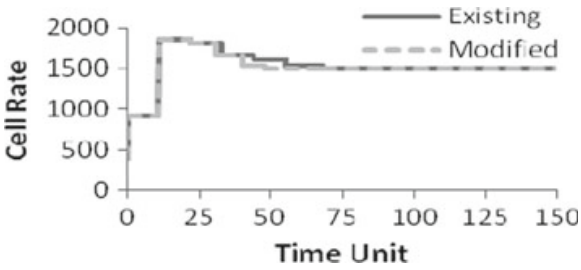


Fig. 3 Fairness model

Fig. 4 Cell rate of source1



6.2 Simulation Parameters

The following parameters are used to simulate above model:-

i. Initial cell rate (ICR) = 400 cells.
ii. Peak cell rate (PCR) = 4,500 cells.
iii. Minimum cell rate (MCR) = 0
iv. Time unit: time required to serves 5,000 cells.
v. Target queue length Q = 700 cells.
vi. Maximum network delay (d_M) = 10 time units.
vii. Time duration d_1 = 7 time units

The propagation delay between two adjust node is 1.6 time units, but the actual delay have variation since the cells go through nodes buffers, which leads to additional queuing delays.

6.3 Simulation Results

The simulation results in Figs. 4 and 5 showing cell rate of the source1 and source 2 respectively for both existing and modified methods. Initially the cell rate for every source is 400 cells per unit time then it is increased by RIF*PCR gradually until congestion has not occurred. But when congestion occurs at any switch then its information send to all sources to reduce their cell rate by RDF*PCR immediately. In the graph of Figs. 4 and 5 we observed that the modified method takes less time compared to existing method to reduce cell rate if congestion has occurred in ATM network. It is possible due to delay d_1 that replaced to d_M where $d_1 \leq d_M$. The source0 and source3 also take less time in modified method to settle cell rate. The all sources getting almost equal bandwidth so that we can say the system is fair.

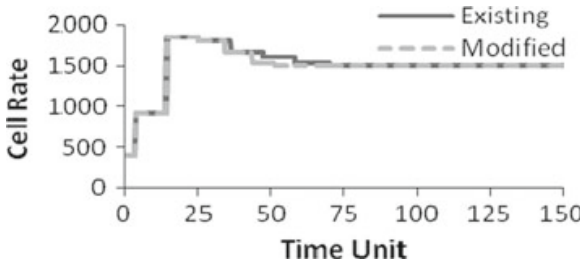


Fig. 5 Cell rate of source2

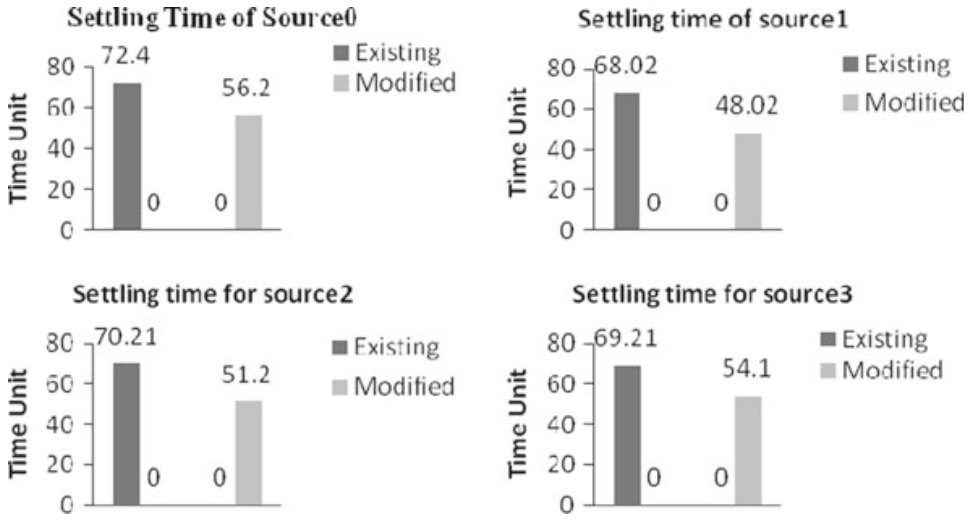


Fig. 6 settling time of all four sources

The settling time of existing method and modified method are shown in figure 6. In this, we can also observe that the data rate settling time of sources in modified method are $20 \approx 25\%$ less compared to the existing method. The less settling time may avoid congestion in ATM network.

7 Conclusions

The effect of the modification of the Adaptive Congestion Control Algorithm is that if congestion occurs at any switch, the switch generates a RM cell and sends it to all sources immediately. The sources are informed to reduce cell rate and maximum round-trip delay which is replaced with d_1 in RM cell so that the source may update their cell rate earlier. The method presented for ABR congestion control has shown that it performs well under various criteria such as basic fairness while achieving high bandwidth utilization. The network delay is explicitly taken into account, which may be useful in modeling network behavior over links with large delays. It is shown that the modified method has much smaller settling time in comparison to the robust adaptive congestion control method and it can avoid congestion in ATM networks.

References

1. Tanenbaum, A. S. (1999). *Computer networks*. India: Prentice Hall of India Pvt. Ltd.
2. Bertsekas, D., & Gallarger, R. (1999). *Data networks*. India: Prentice-Hall of India Pvt.
3. Riazi, A., Habibipour, F., & Galily, M. (2005). Design of congestion controller for ATM networks via adaptive control law. *IEEE Proceedings*.

4. Jing, Y., Ren, T., Zhou, Y., & Zheng, X. (2007). ABR Traf_c control over ATM networks with time-varying multiple time-delays using fuzzy-immune controller. In *IEEE Proceedings on Decision and Control, New Orleans, LA, USA,December 12–14*.
5. Altman, E., Basar, T., & Srikant, R. (1999). congestion control as a stochastic control problem with action delays. *In Automatica, 35*, 1937–1950.
6. Ohsaki, H., Murata, M., & Miyahara, H. (1997). Designing efficient explicit-rate switch algorithm with max- min fairness for ABR service class in ATM networks. *IEEE Proceedings*.
7. Ohsaki, H., Murata, M., Suzuki, H., Ikeda, C., & Miyahara, H. (1995). Rate-based congestion control for ATM networks. In *ACM SIGCOMM* Computer Communication Review.
8. Law, L. M., & Kelton, W. D. (1991). *Simulation modelling and analysis*. New York: Mc Graw Hill international edition.
9. Hander, R., Huber, M. N., & Schroder, S. (2001). *ATM networks concepts, protocol, application*. Singapore: Addison Wesley Longman Pte.Ltd.
10. Kalyanaranman, S., Jain, R., Fahmy, S., Goyal, R., & Vandalore, B. (2000). The ERICA switch algorithm for ABR traffic management in ATM networks. *IEEE/ACM Transactions on Networking, 8*(1), 87–89.
11. Imer, O. C., Basar, T., Compans, S., & Srikant, R. (2001). Available bit rate congestion control in ATM networks: Developing explicit rate control algorithms. In *IEEE Control System Magazine*. pp. 38–56.
12. Imer, O. C., Basar, T., & Srikant, R. (2000). A robust adaptive algorithm for ABR congestion control in ATM networks. In *Proceedings IEEE ICCCN, Las Vegas, NV*, pp. 48–53.

Author Biographies

R. L. Ujjwal is working as Asstt. Professor in University School of Information Technology since 1st September 2006. He obtained his B.E. (Computer Science and Engineering) from M.B.M. Engineering College, JNV University Jodhpur in 1999 and M.E. (Computer Technology and Application) from Delhi College of Engineering, University of Delhi. His Major areas of interests include Wireless Communications, Computer Network.

C. S. Rai is a Professor with the University School of Information Technology. He obtained his M.E. degree in Computer Engineering from SGS Institute of Technology & Science, Indore. He completed Ph.D. in area of Neural Network from Guru Gobind Singh Indraprastha University in 2003. His teaching and research interests include: Artificial Neural Systems, Computer Networks, Signal Processing.

Nupur Prakash is working as a Principal, Indira Gandhi Institute of Technology, a constituent college of GGS Indraprastha University and a girl's engineering college started by Govt. of Delhi. She has also worked as Additional Director (IT) at Institute of Chartered Accountants of India, Delhi. She was holding the charge of Dean in the University School of Information Technology at GGS Indraprastha University, Delhi (2006-2008) prior to this assignment. She received her B.E. (Electronics & Communication Engineering) and M.E. (Computer Science & Technology) degree from University of Roorkee (now I.I.T. Roorkee) in 1981 and 1986 respectively. She was awarded with the Chancellor's Medal for being the best outgoing student of 1981. She has worked as a scientist B at CSIO, Chandigarh (1983-85) on microprocessor based projects. She completed her Ph.D. (Computer Engineering & Technology) from Punjab University, Chandigarh in 1998 in the area of Natural language Processing using Artificial Neural Networks.

Available ONLINE www.vsrdjournals.com

VSRD International Journal of Biotechnology & Pharmaceutical Sciences

ISSN No. 2278-9197

VSRD-IJBPS, Vol. 1 (1), 2012, 6-10

RESEARCH ARTICLE

Analysis of Nanobiotechnological Applications and Study of Their Potential Implementation on Bacterial Growth and Reduced Toxicity

[1]Isha Bharti*

ABSTRACT

Nanobiotechnological applications involve microbial technology, cycloning, flotation, micro flotation, electro flotation and filtration processes. These are some of the chemical methods for removing the toxicity in the effluents to make them available for further practical uses. Continuous flow and fixed film bioreactors are also used for the reduction of chromium compounds. Despite the toxicity of the effluents, the microbial flora of tannery wastes is relatively rich with the organisms such as bacteria, yeast, algae, protozoa and fungi causing irritation and corrosion of the skin and respiratory tract. Nanomagnets in the form of iron coating molecules can be used to improve the contrast between healthy and diseased tissue in MRI. The obtained results highlight the electrostatic interactions between nanoparticles and cell surface as the primary step towards nanotoxicity, followed by cell morphological changes, increase in membrane permeability and their accumulation in the cytoplasm.

Keywords : Nanotoxicity, Nanobiotechnology, Biomolecules.

1. INTRODUCTION

There has been a significant increase in the use and application of nanoparticles in wide realms of life. Miniaturization of materials to nano range dimensions alters their activity and properties. Increased presence of nanoparticles in environment necessitates a basic understanding about their interactions with biomolecules and biological systems. Toxic effects of nanoparticles, known as ‘‘nanotoxicity” is increasingly observed. Recent studies on animals and cell culture have amply shown loss of cell viability, tissue damage and inflammatory reactions. To reduce the toxicity of soil, water, air and the atmosphere as a whole caused due to the discharge of

[1]Research Scholar, Department of Applied Physics, Delhi Technological University, Delhi, INDIA. *Correspondence : dtu.isha@gmail.com

heavy metals and contaminants from effect of tannery effluent on water and soil profile, plant growth and human health the industries, various treatment procedures are widely followed. These methods aim at reducing the toxicity of the metals before letting into the land. NPs may enter the human body via the lungs and the intestines (if ingested); penetration via the skin is less evident but it is possible that some NPs can penetrate deep into the dermis. NP internalization depends on the particle size, surface properties, and functionalization. After internalization, the NP distribution in the body is a strong function of the NP's surface characteristics. The use of plants for extraction of contaminants from the environment or for lowering their toxicity defined as phytoremediation is yet another method for toxicity removal. The toxicity of commonly used nanoparticles on bacterial cells has been investigated. Although antibacterial activity of nanoparticles is known since long, yet the precise interactions of nanoparticles with prokaryotic cells are yet to be explored to a larger extent. The study of the nanotoxicity in microbial systems is important for two reasons; (i) the discharge of nanoparticles in water and soil might affect the ecological diversity of the area (ii) antimicrobial activity of nanoparticles could be usefully exploited for appropriate biotechnological applications. Nanoparticles have been shown to inhibit growth of Escherichia. coli, Pseudomonas. aeruginosa and few other microorganisms. The nanotoxicity is more pronounced on Gram negative bacteria. Nanoparticles reduce the growth of bacterial cells by 50-80%. Nanotoxicity towards Gram positive cells is significantly less, possibly due to presence of thicker layer. The interactions between the bacterium nanoparticles were probed by electron microscopy and energy dispersive X-ray analysis.

2. BACTERIAL SYNTHESIS OF NANOPARTICLES

Control over physical and chemical properties such as shape and crystal anisotropy, size distribution, sensitivity to magnetic fields, existence of a single magnetic domain, and biocompatible surface functionalization for biological and biomedical purposes is essential for envisioning potential applications of magnetic nanoparticles. While different methods of biological and chemical synthesis of biological nanoparticles do exist, they do not allow for precise control over one or more of the aforementioned properties. Moreover, physical synthesis is required to be carried out under harsh conditions (e.g., high temperature and pressure) and results in nanoparticles that can be transiently synthesized at best. Magnetotactic bacteria are aquatic prokaryotes, with microbiological properties of being gram negative and obligate microaerophilic in nature. They synthesize nanomagnets of magnetite or greigite by consuming iron salts, with a very robust control over physicochemical properties. Biofriendly synthesis of nanoparticles by these bacteria occurs at very mild conditions (at room temperature and atmospheric pressure) compared to chemical methods. Figure 1 shows species of Magnetotactic bacteria taken as a SEM image.

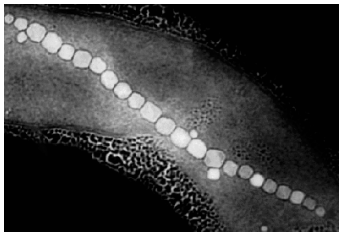


Fig 1 : SEM Image of Magnetotactic Bacteria

3. APPLICATION OF NANOMAGNETS IN BIOTECHNOLOGICAL SYNTHESIS OF NANOPARTICLES

Nanomagnets can be used to enhance signal from magnetic resonance imaging (MRI). Iron-oxide particles (dubbed magnetic nanobeads) can be coated with a suitable neutral chemical and injected into bloodstream. Depending upon their size, chemical coating, they travel to different organs of the body. By selecting particles of particular sizes, study of specific parts of the body is done. Nanomagnets in the form of iron coating molecules can be used to improve the contrast between healthy and diseased tissue in MRI. This test image shows what happens when nanomagnets are used to alter the nuclear properties of hydrogen in water, increasing brightness compared to deionized water. One such mechanism of alteration is shown below in figure 2.

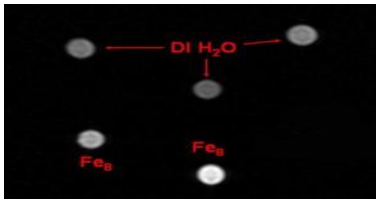


Fig. 2 : Iron oxide particles altering the nuclear properties of hydrogen in water, thereby increasing the brightness (bright spots below left and center)

4. NANOMAGNETS IN DRUG-DELIVERY

Nanomagnets can be used for drug-delivery. Nanobeads are first laced with drug molecules. Then they are steered by external magnetic-field gradients through which they reach the desired parts of the human body. This targeted drug delivery technique limits the exposure of healthy tissue to the drug. It is one of the most active areas in cancer research. It is currently the subject of clinical trials. It has been reported that cancer cells are more susceptible to high temperatures than normal cells. By increasing the temperature of tissue to more than 42°C, the cells could be selectively destroyed. To achieve this, a dose of magnetic nanoparticles could be injected into a region of malignant tissue. Then alternating magnetic field is applied. With sufficiently strong field and of optimum frequency, the particles absorb energy and heat surrounding tissue, affecting only infected cells.

5. RESEARCH METHODOLOGY

Methodology in synthesis of nanomagnets involves the intensive literature review to have an elaborated idea of various experimental conditions required for synthesis of nanomagnets with high yield succeeded by the isolation of species with the high production capability. Metal-reducing bacteria were grown through the oxidation of hydrogen or organic electron donors, coupled to the reduction of oxidized metals such as Fe(III)-bearing minerals that resulted in the formation of magnetite via the extracellular reduction of amorphous Fe(III)-oxyhydroxides, releasing soluble Fe(II) and completely recrystallizing the amorphous mineral into a new phase. Thereafter characterization of the particles formed is done.Nanoparticles formed were analyzed through X-ray diffraction and scanning electron microscopy that showed that the material is nanocrystalline

6. DATA ANALYSIS

Various observations are tabulated and graph is plotted between time and optical density for detailed data analysis. With time as growth of the bacteria increases, Optical density increases and therefore the nonmagnetic particle production also increases. Graphical representation of various events show that increase in optical density is directly related with nanomagnets production and therefore directly proportional. Further various characterization techniques were used to calculate the size of particle , morphology and magnetizing power.

7. RESULT AND DISCUSSION

Effect of nanomolecules was monitored on growth and viability of the Enterobacter sp. (Gram negative), B. subtilis (Gram positive) and halophilic Marinobacter sp. (Gram negative) and halophilic bacterium sp. EMB4 (Gram positive) cells. The cells were grown in the nutrient media containing nanoparticles of ZnO (5 mM; concentration below this did not show any significant effect) or Ag (2 mM; concentrations above this was highly toxic). In another set of experiment, similar concentrations of bulk ZnO and Ag were used instead of nanoparticles. A control was simultaneously run in absence of bulk or nanoparticles. The growth curves, viable counts, scanning electron micrographs/transmission electron micrographs and energy dispersive X-ray profiles were recorded. The growth curves of Enterobacter sp. are shown in Fig. 3 a and b. It is evident that bulk ZnO did not affect the growth rate and viable counts but the ZnO nanoparticles caused substantial decrease in them. Similar trend was observed in case of Ag nanoparticles. The reduction in growth and viability caused by silver nanoparticles was lesser than nanoparticulate ZnO. Above results clearly indicate that nano size of the metal exerts nanotoxic effect on cell growth and viability. Scanning electron micrographs of the nano ZnO and Ag treated Enterobacter sp. cells revealed changes in cell morphology and reduction in cell size. These observations further support the nanotoxic effect of metal nanoparticles.

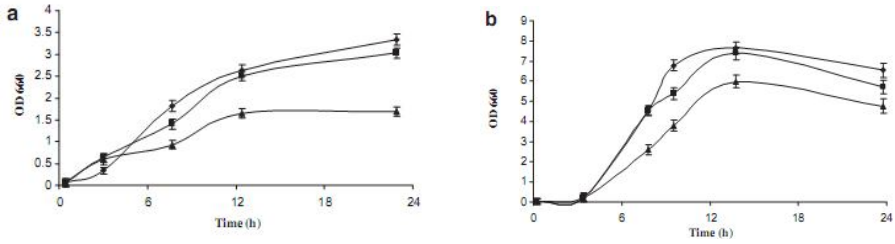


Fig. 3: Effect of nanoparticle on growth and viability of Gram negative Enterobacter sp.
(a) cells grown in the culture media containing bulk and nanoparticles,
(b) Effect of nanoparticles on growth profile.

8. CONCLUSION

The results validated that the toxic effect of nanoparticles depends primarily on the nature of membrane and its interaction with them. At present, there is a very limited understanding of the toxicity of nanoparticles on the microbial systems. The discharge of the used nanoparticles in the environment needs to be looked carefully from the viewpoint of their adverse effect on microbial biodiversity so that the toxicity level can be confined to an acceptable level.

9. FUTURE SCOPE

High yielding nano magnet particle production by bacteria and industrial level scale up of the processes may have brighter aspects and some promising applications in various areas. Nanomedicine in nanotechnology is rapidly growing with various prospects such as gene targeting, drug delivery, diagnosis, and treatment of various diseases and ailments. Several advanced fabrication techniques for the synthesis of multifunctional nanomagnets with novel physico-chemical properties are already known with various applications such as tissue engineering, gene therapy and site directed delivery. However the major difficulty in utilization of these applications lies in the magnetization of naturally produced magnets since these are not strong enough for particular biomedical uses and are not strong enough to magnetize the magnetic nanoparticles to counterbalance and therefore effective accumulation of nanomagnets at the target site becomes difficult. The development of high-quality magnetic nanoparticles with a strong magnetization and powerful external magnetic field that is essential in potential magnetic drug/gene delivery applications is yet another challenge for future studies. It would also result in better studies of cellular interactions and controlled release of therapeutic agents in a controlled manner.

10. REFERENCES

[1] Vanessa, H., Katrina, E., Eric, W., Alison, B., 2009. Siderophores of Marinobacter aquaeolei: petrobactin and its sulfonated derivatives. Biometals 22, 565–571.

[2] Singh, M., Singh, S., Prasad, S., Gambhir, I.S., 2008. Nanotechnology in medicine. Dig.J. Nanomater. Bios. 3, 115–122.

[3] Characterization of enhanced antibacterial effects of novel silver nanoparticles.Nanotechnology 18, 225103–225112.

[4] Reddy, K.M., Feris, K., Bell, J., Wingett, D.G., Hanley, C., Punnoose, A., 2007. Selective toxicity of zinc oxide nanoparticles to prokaryotic and eukaryotic systems. Appl. Phys. Lett. 90, 213902–213903.

[5] Oberdorster, G., Oberdorster, E., Oberdorster, J., 2005. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environ. Health Perspect.113, 823–829.

[6] Neal, A.L., 2008. What can be inferred from bacterium–nanoparticle interactions about the potential consequences of environmental exposure to nanoparticles. Nanotoxicology 17, 362–371.

❑❑

Automated Tool to Generate Parallel CUDA code from a Serial C Code

Akhil Jindal
Department of Computer Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi - 110042, India.

Nikhil Jindal
Department of Computer Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi - 110042, India.

Divyashikha Sethia
Department of Software Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi - 110042, India.

ABSTRACT
With the introduction of GPGPUs, parallel programming has become simple and affordable. APIs such as NVIDIA's CUDA have attracted many programmers to port their applications to GPGPUs. But writing CUDA codes still remains a challenging task. Moreover, the vast repositories of legacy serial C codes, which are still in wide use in the industry, are unable to take any advantage of this extra computing power available. Lot of attempts have thus been made at developing auto-parallelization techniques to convert a serial C code to a corresponding parallel CUDA code. Some parallelizes, allow programmers to add "hints" to their serial programs, while another approach has been to build an interactive system between programmers and parallelizing tools/compilers. But none of these are really automatic techniques, since the programmer is fully involved in the process. In this paper, we present an automatic parallelization tool that completely relieves the programmer of any involvement in the parallelization process. Preliminary results with a basic set of usual C codes show that the tool is able to provide a significant speedup of ~10 times.

General Terms
Auto parallelization, parallelization, C, CUDA, hiCUDA, GPU.

Keywords
Auto parallelization, parallelization, C, CUDA, hiCUDA, GPU.

1. INTRODUCTION
In the last decades, there have been great advancements in the field of Parallel Computing. With the introduction of General Purpose Graphical Processing Units (GPGPUs), attaining parallel processing capability has become simple and affordable. A typical GPU is a multi-core architecture with each core capable of running thousands of threads simultaneously. Hence, an application with a large amount of parallelism can use GPUs to realize significant performance benefits. SDKs and APIs such as Nvidia's CUDA [1], AMD's FireStream and Khronos Group's Open CL [2] have simplified the task of programming GPUs. Some of the areas where GPUs have been used extensively for General Purpose computing are: scientific computing [3][4], image processing [5][6][7], animation and simulation [8][9] and cryptography [10].

But the vast repositories of legacy serial C codes, which are still in use, are unable to exploit this extra computing power available to them. Manually updating all such codes is tedious and error-prone. Parallelizing even a single C code is not a trivial task. The programmer needs to have a complete knowledge of the code being parallelized and should be comfortable with the target parallel architecture. Also, even though APIs, such as those of CUDA, have attracted many non-graphics programmers to port their applications to GPGPUs, the process still remains very challenging. In particular, CUDA places on the programmer the burden of packaging GPU code in separate functions, of explicitly managing data transfer between the host memory and various GPU memories, and of manually optimizing the utilization of the GPU memory [11].

Due to the reasons mentioned above, we have undertaken the task to develop "Automated Tool to Generate Parallel CUDA code from a Serial C Code". The tool is aimed at enabling easy portability of existing serial softwares to parallel architectures. This should be possible without the user having any knowledge whatsoever of the algorithm and the architecture.

Though the quality of automatic parallelization has improved in the past several decades, fully automatic parallelization of sequential programs by compilers remains a grand challenge due to its need for complex program analysis and the unknown factors (such as input data range) during compilation. Attempts have been made at simplifying the process of manual parallelization by allowing programmers to add "hints" to their programs to guide compiler parallelization, such as High Performance Fortran (HPF) [12] for distributed memory systems and OpenMP [13] for shared memory systems. Another approach has been to build an interactive system between programmers and parallelizing tools/compilers. Notable examples are Vector Fabrics' vfAnalyst [14], SUIF Explorer [15] (The Stanford University Intermediate Format compiler), the Polaris compiler [16], and ParaWise (formally CAPTools) [17].

There exist some directive based auto-parallelization tools for CUDA such as PGI Accelerator, CAPS HMPP, Goose, NOAA F2C Fortran/C to CUDA C auto-parallelization compiler and hiCUDA [18]. But the drawback in all these tools is that the programmer has to understand and learn the specific compiler directive syntax. Our proposed tool goes a step further than these tools in simplifying the process for the user by automatically generating the parallel code from input serial code without any additional input from the user.

The tool works in two phases. In the first phase, the input serial code is parsed to identify independent portions of code which can be executed in parallel. Identifiers are then automatically inserted at appropriate positions to mark these

parallelizable portions. In the second phase, an equivalent CUDA code is generated which parallelizes the portions identified in Phase 1.

The parallel code obtained might not be as efficient as hand-tuned programs but can still lead to tremendous speedups with a quick production phase.

The paper is organized as follows. Section 2 presents the system architecture while the major issues in identifying and then parallelizing portions of serial code are discussed in section 3. Section 4 analyses the results and Section 5 ponders over some future work that can lead to better results.

2. SYSTEM ARCHITECTURE

The proposed tool takes a serial C code as input and generates an equivalent parallel code as output, as illustrated by the flowchart in Figure 1. The generated output code can be compiled and executed on any machine with a CUDA enabled graphics card.

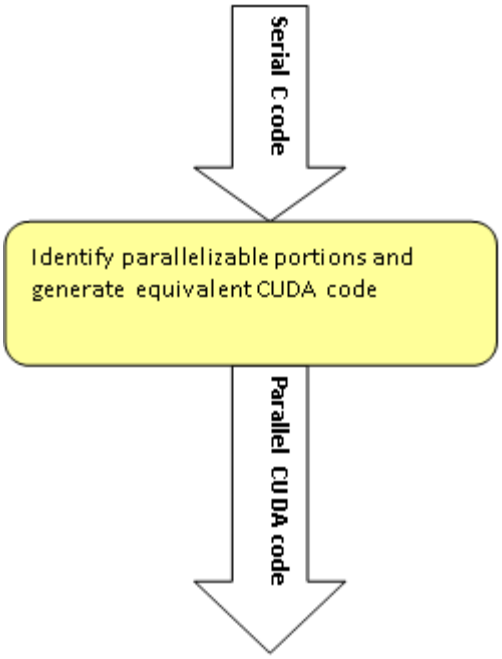


Fig 1: Flowchart showing the input and output with the tool as a black-box

Internally, the tool works in two phases (Figure 2). hiCUDA is used as the intermediate language between them:

- In the first phase, the input serial code is parsed using a Perl script to identify independent portions of code which can be executed in parallel. Identifiers (hiCUDA pragmas) are then automatically inserted at appropriate positions to highlight these parallelizable portions.
- In the second phase, the hiCUDA compiler is used to generate an equivalent CUDA code using the hiCUDA pragmas inserted in Phase 1.

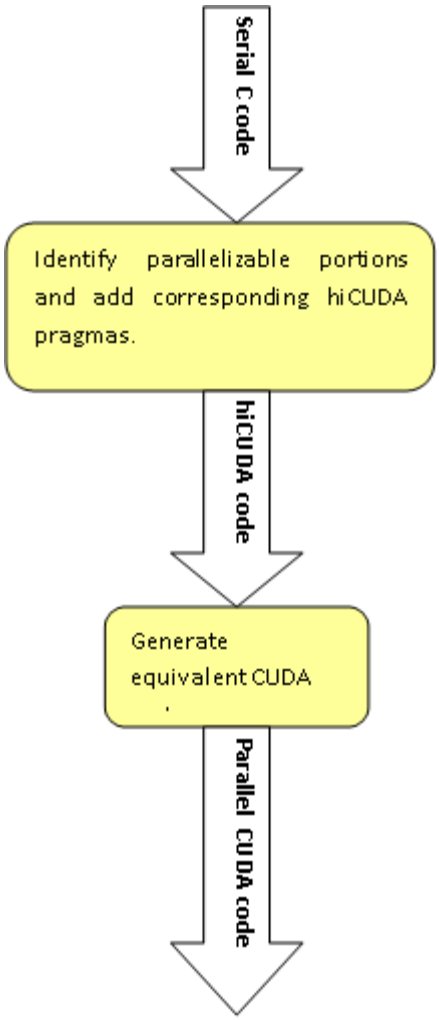


Fig 2: Representation of the internal phases

3. IDENTIFY AND PARALLELIZE PARALLELIZABLE PORTIONS

As most of the execution time of a program takes place inside some form of loop, we intend to focus most on them and will try to split each loop so that each of its iteration can be executed on separate processors concurrently.

Also, since GPUs are optimized for executing SIMD type instructions, they are guaranteed to give maximum gain.

For example, consider the following code snippet:

```
for( i = 0; i < n; i++) {
        a[i] = b[i] * c[i];
}
```

Code 1. Single for loop

Since instructions in each iteration are independent of each other, this code can be easily parallelized by using n threads running in parallel, each operating on one element of vector a.

3.1 When to parallelize

Not all for loops can be parallelized. Hence, we need to have a set of rules defining when a for loop is parallelized and when not. For our purpose, we do not parallelize in the following cases:

- Presence of an I/O instruction in a loop.
- Presence of a break/return/goto statement in a loop.
- A scalar is being Written after Read (WAR).
- Same element of an array being written in each iteration of the loop.
- 2 different elements of same array being accessed in each iteration (at least one being written).

3.2 Handling Nested Loops

In case of nested loops, a separate read and write list is created for each loop independently and the analysis as explained above is done to determine whether the loop is parallelizable.

To handle nested loops, we create a GPU kernel for each bunch of nested loops. Each loop is analyzed independently and a kernel is created if at least one of the nested loops can be parallelized.

For example, consider the following snippet:

```
for( i = 0; i < n; i++) {
        for( j  = 0; j < n; j++) {
                sum[i] = sum[i] + a[i][j];
        }
}
```

Code 2. Nested for loop

In this case, the outer loop will be parallelized while the inner loop will not be, which is indicated by the hiCUDA pragma just before the outer loop only:

```
#pragma hicuda loop_partition over_tblock over_thread
for( i = 0; i < n; i++) {
        for( j  = 0; j < n; j++) {
                sum[i] = sum[i] + a[i][j];
        }
}
```

.Code 3. Nested for loop with a hiCUDA pragma for outer loop

3.3 Determining number of threads (block size and number of blocks)

The number of threads required for the parallel execution of a loop are determined by the number of iterations of each loop.

For example, consider the following code snippet:

```
for( i = 0; i < n; i++) {
        sum = sum + a[i];
}
```

Code 4. Code for computing the sum of all elements of an array

Here, the number of threads required are (n - 0). Hence, for a block_size = 512, number_of_blocks = (n – 0)/512.

For nested loops, the dimensionality of block_size and number_of_blocks changes accordingly.

For example, consider the following code snippet:

```
for( i = 0; i < n; i++) {
        for( j = 0; j < m; j++) {
                a[i][j] = i * j;
}
}
```

Code 5. Sample code for initializing a 2-D array

Here, the number of threads required are (n - 0) * (m – 0). Hence, for a block_size = (16, 16), number_of_blocks = (((n – 0)/16), ((m – 0)/16)), which is indicated by the hicuda pragma in the following code:

```
#pramga hicuda kernel kernel_name tblock(((n – 0)/16), ((m – 0)/16)) thread(16, 16)
#pramga hicuda loop_partition over_tblock over_thread
for( i = 0; i < n; i++) {
#pramga hicuda loop_partition over_tblock over_thread
        for( j = 0; j < m; j++) {
                a[i][j] = i * j;
}
}
```

Code 6. Nested for loops with hiCUDA pragmas

3.4 Memory allocation/de-allocation on GPU

Whenever a kernel is created, memory needs to be allocated (and then de-allocated) on the GPU, for all data variables which are accessed (read/write) inside the kernel. Hence additional information about the dimensionality of each array is maintained.

All variables accessed inside the loop instructions are categorized into two lists: the read list (if the variable is read from) and the write list (if the variable is written onto). These lists are then utilized to determine which variables will be "copyin" and which variables will be "copyout" from GPU memory.

For example, consider the following serial code:

```
for( i = 0; i < n; i++) {
        b[i] = a[i] * a[i];
}
```

Code 7. Sample code for calculating square of each element of array

Here a[] will only be "copyin" to the GPU memory while b[] will be both "copyin" and "copyout" from it, which is indicated by the following pragmas:

```
#pragma hicuda global alloc a[*] copyin
#pragma hicuda global alloc b[*] copyin
#pramga hicuda kernel kernel_name tblock((n – 0)/16) thread(16)
```

```
#pramga hicuda loop_partition over_tblock over_thread
for( i = 0; i < n; i++) {
        b[i] = a[i] * a[i];
}
#pragma hicuda kernel_end
#pragma hicuda global copyout b[*]
#pragma hicuda global free a
#pragma hicuda global free b
```

Code 8. The complete hiCUDA code

4. RESULTS AND ANALYSIS

To test the effectiveness of this tool, it is important to quantify the following:

1. Performance of the generated parallel code v/s original serial code.
2. Performance of the automatic parallel code v/s a hand written best optimized parallel code.

The codes are run on an Intel Core2Duo 1.6 Ghz processor with 2GB RAM. NVIDIA's GeForce 8400 GS graphics card is used for GPU's.

4.1 Measuring speedup obtained by parallelization

To measure the speedup obtained by parallelizing serial codes, the execution time of various input serial C codes are compared with the execution time of the corresponding auto generated parallel CUDA codes.

4.1.1 Matrix Multiplication

First of all, the standard problem of matrix multiplication is considered. Two matrices, a and b are initialised as follows:

$a[i][j] = (i + j) \wedge N$
$b[i][j] = (i - j) \wedge N$
where, a and b are of size N*N.

Then, the resultant c matrix is obtained by multiplying matrices a and b.

$$c[i][j] = \sum_{k=1}^{N} a[i][k] * b[k][j]$$

Figure 3 and Table 1 show the speedup obtained by parallelizing a serial code for this problem. For a moderate matrix size, N = 1024, the speedup obtained is 5-6 times.

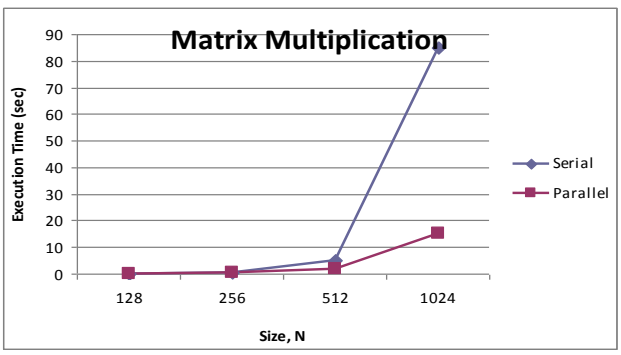


Fig 3: Performance comparison of serial and parallel matrix multiplication codes

Table 1. Execution times for serial and parallel matrix multiplication codes

Size, N	Serial	Parallel
128	0.08	0.15
256	0.6	0.4
512	5	2
1024	85	15

4.1.2 Compute Power Array

The problem of computing the power vector is considered next. Each element of vector a is calculated by using the following formula:

$a[i] = i \wedge N$
where, N is the size of the vector a.

Since, each element is independent of others, this task can easily be parallelized. The proposed tool is used to parallelize the serial code and the speedup obtained is indicated in Figure 4 and Table 2.

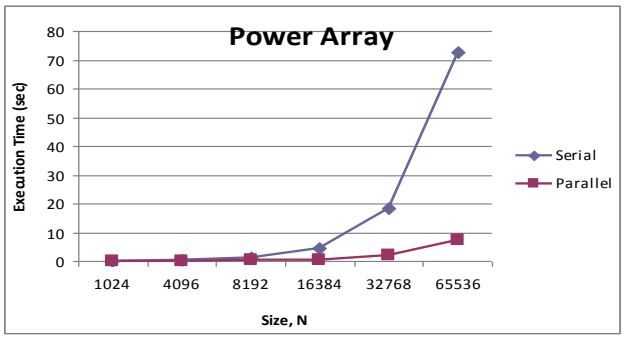


Fig 4: Performance comparison of serial and parallel power array codes

Table 2. Execution times for serial and parallel power array codes

Size, N	Serial	Parallel
1024	0.02	0.141
4096	0.287	0.172
8192	1.142	0.243
16384	4.552	0.584
32768	18.201	1.942
65536	72.8	7.347

In this case, the speedup obtained is ~10 times.

The speedup, in this case, is greater than that for matrix multiplication, which is as expected. A high cost is paid in transferring data from CPU to GPU (and vice-versa), so the computation for each GPU thread should be long enough to justify the overhead transfer costs. In matrix multiplication, for the second kernel (computing c[i]), each thread performs only a single operation of multiplying a single element of

matrix a[] with that of b[], which is unable to compensate for the cost of threads and kernel creation.

4.1.3 Calculate the Prime divisors

Next, we consider a modular code, where each task is divided amongst various functions. We consider the problem of finding the prime divisors of a given number n. Different functions are responsible for finding whether an integer 1 < i < n is a prime number and whether it completely divides the number n. This tool successfully parallelises this code and the speedup obtained is shown by Figure 5 and Table 3.

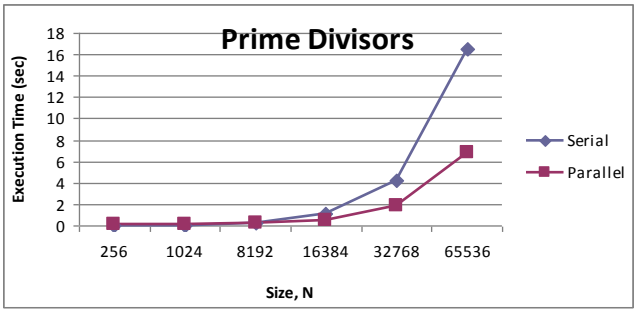


Fig 5: Performance comparison of serial and parallel prime divisors codes

Table 3. Execution times for serial and parallel prime divisors codes

Size, N	Serial	Parallel
256	0.002	0.101
1024	0.006	0.128
8192	0.278	0.212
16384	1.056	0.554
32768	4.162	1.815
65536	16.55	6.82

4.2 Comparing generated parallel code with hand-written code

To get hand written CUDA code, we use the hiCUDA compiler to generate CUDA code from a hand written hiCUDA code.

A hand written parallel code is expected to outperform the automatically generated parallel code in a few cases, but here we document by how much they outperform and the reasons for the same.

4.2.1 Matrix Multiplication

The first code we consider in this section is the matrix multiplication code (which is the same code as in section 4.1.1). When we write the code manually for this specific problem, we can gain some advantage by utilizing the shared memory on GPU's, which is ignored by the generic tool. The data needed by all threads in a thread block can be loaded into the shared memory before they are used, reducing access latency to memory. This can be done by using the following hiCUDA pragma's:

```
#pragma hicuda shared alloc A[*][*] copyin
#pragma hicuda shared remove A
```

Since the amount of data is too large to fit in the shared memory at once, it must be loaded and processed in batches. For this problem, we store 32 elements of each matrix in the shared matrix at a moment. The difference in the automatically generated and the hand-written code is shown in Codes 9 and 10.

```
for(kk=0; kk<N2; kk+=32) {
#pragma hicuda shared alloc A[i][kk:kk+32] copyin
#pragma hicuda shared alloc B[kk:kk+32][j] copyin
#pragma hicuda barrier
        for (k = 0; k < 32; ++k) {
                sum = sum + A[i][kk+k] * B[kk+k][j];
                sum = sum % 100000000;
}
#pragma hicuda barrier
#pragma hicuda shared remove A B
}
```

Code 9. The hand-written code

```
for(k=0; k<N2; k++) {
        sum = sum + A[i][k] * B[k][j];
        sum = sum % 100000000;
}
```

Code 10. The automatically generated code

Figure 6 and Table 4 show a comparison of the running times of both the codes.

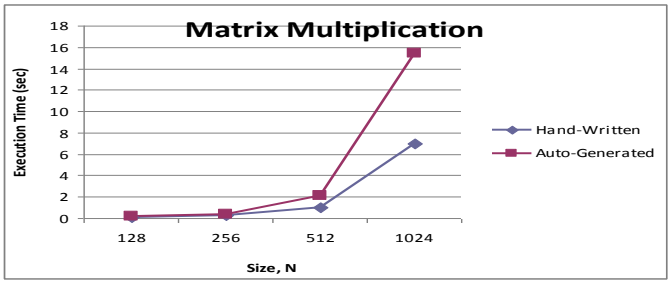


Fig 6: Performance comparison of hand-written and auto-generated matrix multiplication codes

Table 4. Execution times for hand-written and auto-generated matrix multiplication codes

Size, N	Hand-Written	Auto-Generated
128	0.122	0.177
256	0.252	0.344
512	1.002	2.073
1024	6.984	15.47

4.2.2 Vector multiplication

Another area where a hand written code can beat the automatically generated code is by combining multiple kernels together. We use the standard vector multiplication algorithm in this section.

The first kernel, initializes a[] and b[] vectors(or arrays) as power arrays:

a[i] = i ^ N;
b[i] = (N - i) ^ N;

The next kernel calculates c vector as follows:

*c[i] = a[i] * b[i];*

While these tasks are performed by different kernels in the automatically generated parallel code, they are combined into a single kernel in the hand written code. Performance gains are obtained by eliminating the memory transfer instructions. In the automatically generated code, a[] and b[] vectors are first copied from GPU memory into the CPU memory after the execution of first kernel. These are then successively copied back to the GPU memory for the beginning of the second kernel. A comparison of both these codes reveals that the performance gain is not very significant, as shown in Figure 7 and Table 5.

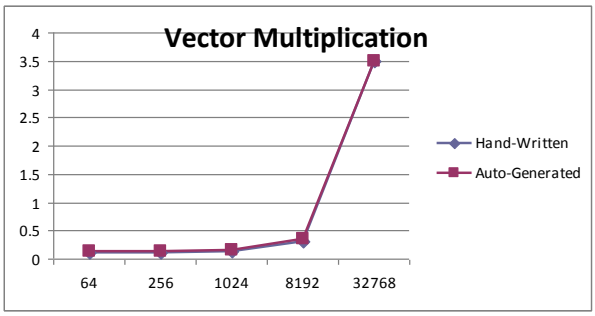


Fig 11: Performance comparison of hand-written and auto-generated vector multiplication codes

Table 5. Execution times for hand-written and auto-generated vector multiplication codes

Size, N	Hand-Written	Auto-Generated
64	0.105	0.139
256	0.103	0.136
1024	0.124	0.143
8192	0.313	0.344
32768	3.488	3.499

5. CONCLUSION AND FUTURE WORK

A working end-to-end tool has successfully been developed and tested over a wide range of codes. The performance results obtained are very satisfying with speedup gains obtained of up to 10 times.

The *automatic* parallelization, we believe, is a very significant step forward and would help the industrial community immensely. It being a *generic* tool capable of handling most kinds of C codes increases its worth.

This good performance only motivates us to improve it further. The ultimate aim is that the tool should be able to parse all legacy C codes. The parsing technique needs to be improved for this. We can also look at some dedicated parsing tools available for C code such as Elsa [19].

Also, the tool uses static analysis to detect data independency, that is, it reads the code as a simple text. On the other hand, if it used dynamic analysis, wherein the code is actually executed and the runtime memory accesses monitored, it would have enabled handling pointers too. But that would have also resulted in the execution time of the serial code being a bottleneck in the parallelization process.

Presently, each set of nested for loops in the C code are combined together to form an independent kernel for the GPU. In cases, such as in matrix multiplication, where we have two consecutive kernels one directly after the other and when the output of one is an input for the next, we can combine the two kernels and remove the unnecessary memory transfer instructions in between them. This would require another pass over an intermediate hiCUDA code.

To obtain maximum speed up, the tool will have to deal with the shared and textured memory of GPU as well. It, at present, deals with only the global GPU memory

6. ACKNOWLEDGMENT

The authors would like to thank Mr. Tarun Beri (IIT Delhi) for sharing his invaluable knowledge of the field and discussing the ideas of the authors and even giving some of his own.

7. REFERENCES

[1] NVIDIA, NVIDIA CUDA Compute Unified Device Architecture-Programming Guide, Version 3, 2010.

[2] Stone, J.E., Gohara, D., Guochun Shi, "OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems", Computing in Science and Engineering, Vol. 12, Issue 3, pp. 66-73, May 2010

[3] E. Alerstam, T. Svensson and S. Andersson-Engels, "Parallel computing with graphics processing units for high speed Monte Carlo simulation of photon migration" , J. Biomedical Optics **13**, 060504 (2008).

[4] Larsen E. S., Mcallister D., "Fast matrix multiplies using graphics hardware", Proceedings of the 2001 ACM/IEEE Conference on Supercomputing, Nov. 2001, pp. 55.

[5] Vladimir Glavtchev, Pinar Muyan-Ozcelik, Jeffrey M. Ota, John D. Owens, "Feature-Based Speed Limit Sign Detection Using a Graphics Processing Unit", IEEE Intelligent Vehicles, 2011.

[6] Woetzel J., Koch R., "Multi-camera realtime depth estimation with discontinuity handling on PC graphics hardware", Proceedings of the 17th International Conference on Pattern Recognition (Aug. 2004), pp. 741–744.

[7] Rumpf M., Strzodka R., "Level set segmentation in graphics hardware", Proceedings of the IEEE International Conference on Image Processing (ICIP '01), Oct. 2001, vol. 3, pp. 1103–1106.

[8] Purcell T. J., Buck I., Mark W. R., Hanrahan P., "Ray tracing on programmable graphics hardware", ACM Transactions on Graphics 21, 3 (July 2002), pp 703–712.

[9] Knott D., Pai D. K., "CInDeR: Collision and interference detection in real-time using graphics hardware", Proceedings of the 2003 Conference on Graphics Interface, June 2003, pp. 73–80.

[10] Svetlin A. Manavski, "Cuda compatible GPU as an efficient hardware accelerator for AES cryptography" Proc. IEEE International Conference on Signal Processing and Communication, ICSPC 2007, (Dubai, United Arab Emirates), November 2007, pp.65-68.

[11] T. D. Han and T. S. Abdelrahman, "hiCUDA: High-Level GPGPU Programming", IEEE Transactions on Parallel and Distributed Systems, Jan. 2011, vol. 22, no. 1, pp. 78-90.

[12] David B. Loveman, "High Performance Fortran", IEEE Parallel & Distributed Technology: Systems & Technology, February 1993, v.1 n.1, pp 25-42.

[13] Leonardo Dagum and Ramesh Menon, "OpenMP: An industry-standard API for shared-memory programming", IEEE Computational Science and Engineering, 5(1):46–55, January–March 1998.

[14] VectorFabrics. vfAnalyst: Analyze your sequential C code to create an optimized parallel implementation. http://www.vectorfabrics.com/.

[15] M. Hall, J. Anderson, S. Amarasinghe, B. Murphy, S.-W. Liao, E. Bugnion, and M. Lam, "Maximizing multiprocessor performance with the SUIF compiler", IEEE Comput. 29, 12, Dec. 1996, pp 84–89.

[16] W. Blume, R. Doallo, R. Eigenmann, J. Grout, J. Hoeflinger, T. Lawrence, J. Lee, D. Padua, Y. Paek, B. Pottenger, L. Rauchwerger, and P. Tu. "Advanced Program Restructuring for High-Performance Computers with Polaris", IEEE Computer, December 1996, Vol. 29, No. 12, pages 78- 82.

[17] Johnson, S.P., Evans, E., Jin, H., Ierotheou, C.S., "The ParaWise Expert Assistant—Widening accessibility to efficient and scalable tool generated OpenMP code", WOMPAT, pp. 67–82 (2004).

[18] T.D. Han, "Directive-Based General-Purpose GPU Programming", master's thesis, Univ. of Toronto, Sept. 2009.

[19] Elsa: The Elkhound-based C/C++ Parser. http://www.scottmcpeak.com/elkhound/sources/elsa/

Design of Circular/Triangular Patch Microstrip Antennas using a Single Neural Model

Taimoor Khan and Asok De
Department of Electronics and Communication Engineering
Delhi Technological University (Formerly Delhi College of Engineering)
Shahbad Daulatpur, Bawana Road, Delhi-110 042, India
ktaimoor@gmail.com

Abstract- **An efficient and quick approach based on artificial neural networks (ANN) is being applied on different patches of microstrip antennas since last one decade. Different scientists have proposed different neural models for analyzing the different types of microstrip patches like rectangular, triangular, square, circular etc. whereas few models have also been proposed for designing rectangular, triangular and square patches. But in the available literature of microstrip antennas with neural networks no single model has been proposed till date firstly for calculating the radius of the circular patch microstrip antenna (CPMSA) and secondly for calculating more than one parameter like radius of the CPMSA and side-length of the equilateral triangular patch microstrip antenna (ETMSA) simultaneously. In this paper authors have proposed a multi-layered perceptron feed forward neural network model for calculating the radius/side-length of the circular patch/triangular patch microstrip antennas simultaneously. The model has also been validated on some mathematically generated datasets that are not included in training or testing of the model. The results obtained by the proposed model are in conformity and very good in agreement with the experimental results given in the literature.**

Index Term- Multi-layered perceptron feed forward neural networks, microstrip antennas, radius, circular patch, side-length and triangular patch.

I. INTRODUCTION

Applications of artificial neural networks in microstrip antennas are more than one decade old. In last one decade, different neural approaches have been proposed for calculating a single performance parameter (physical dimensions or resonant frequency) of the microstrip patches [1-7]. Mishra and Patnaik [1-2] have proposed two different neural approaches; one is based on neural network-based CAD model for designing the square patch microstrip antennas [1] and another based on neuro-spectral method for designing the rectangular patch microstrip antennas [2]. The designing of ETMSA using artificial neural networks has been done by Gopalakrishnan and Gunasekaran [3] whereas the analysis and design of rectangular patch microstrip antennas using neural approach has been done by Turker et al [4]. Guney et al [5] have proposed a generalized neural method for calculating the resonant frequencies of rectangular, circular and triangular patches using equivalent area concept. They have used three different algorithms, backpropagation, delta-bar-delta, and extended delta-bar-delta for training purpose. The average absolute errors in their neural model for circular patch are calculated as 0.02480 GHz, 0.06760 GHz, and 0.05170 GHz for these three algorithms respectively. It means the least error in their model is 0.02480 GHz. Ouchar et al [6] have used multi-layered perceptron artificial neural networks model with backpropagation training algorithm for calculating the resonant frequencies of the circular patch and the average absolute error from this model is calculated as 0.03461 GHz. Sagiroglu et al [7] have also calculated resonant frequencies of the circular patch microstrip antennas using neural approach. They have used standard backpropagation algorithm with learning coefficient of 0.08000 and the momentum coefficient of 0.10000. The average absolute error from this model is calculated as 0.00185 GHz. Guney and Sarikaya [8-9] have used two different neural approaches; one is based on ANFIS (artificial neural network and fuzzy interference system) method [8] and another based on CNFS (concurrent neuro-fuzzy system) method [9] for calculating the resonant frequencies of rectangular, triangular and circular microstrip antennas, simultaneously. From these two models, the least average absolute errors for circular patch are calculated as 0.00460 GHz and 0.00580 GHz respectively.

The previously reported works [1-3] and [6-7] have been used for calculating one parameter and Turker et al [4] have calculated two different parameters (resonant frequency and physical dimensions) of the rectangular patch microstrip antennas whereas Guney et al [5] and Guney and Sarikaya [8-9] have calculated only one parameter i.e. resonant frequency of different types of patches, simultaneously using equivalent area concept. However, no single model has been proposed till date firstly for calculating the radius of CPMSA and secondly calculating more than one parameter like radius of the CPMSA and side-length of the ETMSA simultaneously using neural approach. The training and testing data sets in the previously reported neural models [1-9] have been taken directly from the literature and no one has checked the validity of their neural models for their own generated data sets.

This paper suggests a common feed forward neural model based on multi-layered perceptron for calculating the radius of a CPMSA and side-length of ETMSA simultaneously. The parameter selection (radius or side-length) is done by considering a parameter say M where M = 1 corresponds to radius and M = 2 corresponds to side-length calculation. The model is trained by Levenberg-Marquardt training algorithm for both radius and side-length calculation simultaneously.

978-1-4577-1099-5/11/$26.00 ©2011 IEEE

Some random numbers have been selected as initial weights and biases for the model. Finally, the calculated radiuses of CPMSA and side-lengths of ETMSA from the proposed model are also compared with their corresponding experimental counterparts.

II. GEOMETRY OF MICROSTRIP ANTENNAS

A microstrip antenna, in its simplest configuration, consists of a radiating conductive patch on one side of a dielectric substrate having a ground plane on the other side. The CPMSA is one for which the patch conductor has a circular shape whereas the ETMSA has an equilateral triangle shape of radiating patch. A circular patch of radius, 'r', in CPMSA or an equilateral triangular patch of side-length, 'L' in ETMSA, is mounted on a substrate of relative permittivity, 'ε_r', and at height, 'h', from the ground plane as shown in Figure 1(a) and Figure 1(b) respectively.

It is clear from the literature [10-16] and [17-21] that the resonant frequency, 'f' of the CPMSA (or of the ETMSA) is the function of radius, 'r' of the CPMSA (or side-length, 'L' of ETMSA), relative permittivity, 'ε_r', height of the dielectric material, 'h' and the mode of propagation, 'integers m and n. Similarly, if the resonant frequency, 'f', thickness of the dielectric, '*h*', relative permittivity, 'ε_r' and mode of propagation, 'm and n' are given, then the radius, 'r' of CPMSA (or side-length, 'L' of the ETMSA) can be calculated easily. Keeping this concept in mind some datasets for calculating the radius of CPMSA and side-length of ETMSA has been arranged from the literature [10-16] and [17-21] respectively and are given in Table 1. These datasets are used in training and testing of the proposed neural model.

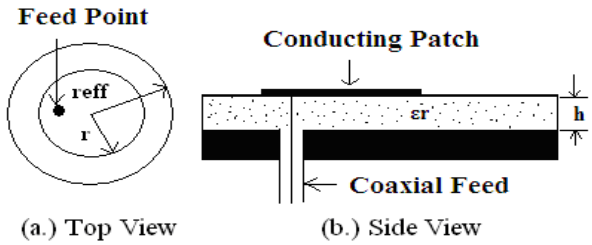


Figure 1(a): Geometry of CPMSA

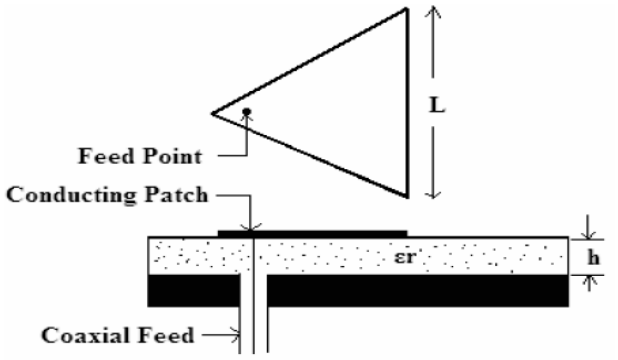


Figure 1(b): Geometry of ETMSA

III. PROPOSED NEURAL APPROACH

In this proposed work a two hidden layer feed forward neural model is suggested for calculating the radius (r_n) and side-length (L_n) simultaneously and is shown in Figure 2.

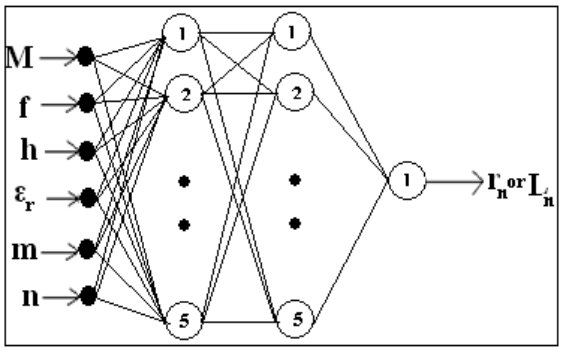


Figure 2: Neural model for radius/side-length calculation

After many trials it is found that five neurons in the first and second hidden layer are sufficient for calculating the radius (r_n) and side-length (L_n) simultaneously. Logsigmoidal is used as the activation function for both the hidden layers. The activation function in the output layer is pure linear whereas for the input layer no activation function is used and the training algorithm is based on Levenberg-Marquardt back propagation [22].

IV. CALCULATED RESULTS

(i) During Training and Testing

The radiuses (r_n) and side-lengths (L_n) calculated during training and testing of the proposed neural model are tabulated in Table 1. It is clear from this table that the results obtained from this common neural model during training and testing are closer to their corresponding experimental values.

(ii) During Validation

To check the validity of the proposed model a design procedure for CPMSA [23] is given in (1)-(2):

$$r = \frac{A}{\left\{1 + \frac{2h}{\pi\varepsilon_r A}\left[\ln\frac{\pi A}{2h} + 1.7726\right]\right\}^{1/2}} \tag{1}$$

Where

$$A = \frac{8.791\times10^{9}}{f\sqrt{\varepsilon_r}} \tag{2}$$

In equations (1)-(2) the resonant frequency, 'f' is given in Hz, height of the dielectric material, 'h' in cm, and radius, 'r' in cm. Using (1)-(2) total four datasets for radius have been calculated corresponding to four randomly selected combination of 'f', 'h', 'ε_r', 'm' and 'n' and are also tabulated in the bottom side of Table 1.

V. COMPARISON AND CONCLUSION

Table 1 describes the comparison of experimental and calculated results. This table shows that the average absolute error for radius calculation is 0.00030 cm whereas for side-length calculation it is 0.0009 cm during training ant testing of the model whereas for during validation of CPMSA it is 0.4333cm that is very less for all the datasets taken from the literature [10-16] and [17-21] and generated by the design procedure [23]. After getting training successfully, one can calculate the radius, 'r_n' of CPMSA or side-length, 'L_n' of ETMSA in microseconds within the acceptable range of 'f', 'ε_r' and 'h' without having the knowledge of the microstrip antennas.

Thus the proposed model is used firstly for calculating the radius of CPMSA and secondly for calculating the radius of CPMSA/ side-length of ETMSA simultaneously in very good agreement to the experimental results as well as to the generated results. All the models available in the literature [1-9] have taken training and testing datasets from the existing experimental results [10-21]. No one has validated their proposed neural model by taking some other datasets that are not included in the training or testing of the model. Here the beauty of the proposed model is that the model is not only verifying the experimental results [10-21] but it is also validating the datasets generated by the procedure available in the literature [23-24].

REFERENCES

[1] R.K. Mishra and A. Patnaik, "Neural network-based CAD model for the design of square-patch antennas", IEEE Trans on Antennas and Propag, vol. 46, no. 12, pp. 1890-1891, 1998.

[2] Rabindra K. Mishra, and Amalendu Patnaik, "Designing rectangular patch antenna using the neurospectral method" IEEE Trans on Antennas and Propag, vol. 51, no. 8, 2003.

[3] R. Gopalakrishnan and N. Gunasekaran, "Design of equilateral triangular microstrip antenna using artificial neural networks", IEEE International Workshop on Antenna Technology: Small Antennas and Novel Metamaterials, 2005.

[4] Nurhan Turker, Filiz Gunes and Tulay Yildirim, "Artificial neural design of microstrip antennas" Turk J Elec Engin, vol.14, no.3, 2006.

[5] Karim Guney, Seref Sagiroglu and Mehmet Erler, "Generalized neural method to determine resonant frequencies of various microstrip antennas", International Journal of RF and Microw Computer Aided Engineering, vol. 12, pp. 131-139, 2002.

[6] A. Qucher, R. Aksas and H. Baudrand, "Artificial neural network for computing the resonant frequency of circular patch antennas", Microw and Optical Technol Lett, vol. 47, pp. 564-566, 2005.

[7] Seref Sagiroglu, Karim Guney and Mehmet Erler, "Resonant frequency calculation for circular microstrip antennas using artificial neural networks", International Journal of RF, Microw and Computer Aided Engineering, vol. 8, pp. 270-277, 1998.

[8] K. Guney and N. Sarikaya, "A hybrid method based on combining artificial neural network and fuzzy interference system for simultaneous computation of resonant frequencies of rectangular, circular, and triangular microstrip antennas" IEEE Trans on Antennas and Propag, vol. 55, no 3, 2007.

[9] K. Guney and N. Sarikaya, "Concurrent neuro-fuzzy systems for resonant frequency computation of rectangular, circular, and triangular microstrip antennas" *PIER,* vol. 84, pp. 253-277, 2008.

[10] J. S. Dahele and K. F. Lee, "Effect of substrate thickness on the performance of a circular-disk microstrip antenna", IEEE Trans Antenna and Propagat, Vol. 31, No. 2,pp. 358-364, 1983.

[11] J. S. Dahele and K. F. Lee, "Theory and experiment on microstrip antennas with air-gaps", IEE Proc, vol. 132, No.7, pp. 455-460, 1985.

[12] K. R. Carver, "Practical analytical techniques for the microstrip antenna", Proc Workshop on Printed Circuit Antennas, New Mexico State University, pp. 7.1-7.20, 1979.

[13] K. Antoszkiewicz and L. Shafai, "Impedance characteristics of circular microstrip patches", IEEE Trans Antenna and Propag, Vol. 38, No. 6, pp. 942-946, 1990.

[14] J. Q. Howell, "Microstrip antennas", IEEE Trans Antenna and Propagat, Vol. 23, pp. 90-93, 1975.

[15] F. Abboud, J. P. Damiano, and A. Papiernik, "New determination of resonant frequency of circular disc microstrip antenna: application to thick substrate", Electron Lett, vol. 24, no. 17, pp. 1104-1106, 1988.

[16] T. Itoh and R. Mittra, "Analysis of a microstrip disk resonator", Arch Electron Ubertrugungs, vol. 27, no. 11, pp. 456-458, 1973.

[17] J. S. Dahele and K. F. Lee, "On the resonant frequencies of the triangular patch antenna,'' IEEE Trans on Antenna & Propag, vol. AP-35, no 1, pp. 100-101, 1987.

[18] W. Chen, K. F. Lee, and J.S. Dahele, "Theoretical and experimental studies of the resonant frequencies of the equilateral triangular microstrip antenna,'' IEEE Trans on Antennas & Propag, vol. 40, no 10, pp. 1253-1256, 1992.

[19] R. Garg, and S. A. Long, "An improved formula for the resonant frequency of the triangular microstrip patch antennas," *IEEE Trans. on Antennas and Propag*, 36, 570, 1998.

[20] J. Helszajn, and D. S. James, "Planer triangular resonators with magnetic walls", IEEE Trans. on Microwave Theory and Techniques, vol. 26 no 2, pp. 95-100, 1978.

[21] D. Guha and J. Y. Siddiqui, "Resonant frequency of equilateral triangular microstrip antenna with and without air gap," *IEEE Trans. on Antennas and Propag* vol. 52, no 8, pp. 2174-2177, 2004.

[22] M. T. Hagan, and M. Menhaj, "Training feed forward networks with the Marquardt algorithms," IEEE Trans. on Neural Networks, Vol. 5, pp. 989-993, 1994.

[23] C.A. Balanis, Antenna Theory-Analysis and Design, Second Edition, John Wiley & Sons, Inc. 1982.

TABLE 1: COMPARISON OF MEASURED AND CALCULATED RADIUS/ SIDE-LENGTH OF CPMSA/ETMSA

Designing of CPMSA									
S. No.	**M**	**f(GHz)**	**h (cm)**	ε_r	**m**	**n**	**r_e/L_e (cm)**	**r_n/L_n (cm)**	**Error**
1	1.00000	0.83500	0.08000	2.32000	1.00000	1.00000	6.80000	6.80000	0.00000
2	1.00000	0.82900	0.15900	2.32000	1.00000	1.00000	**6.80000***	6.80000	0.00000
3	1.00000	0.81500	0.31800	2.32000	1.00000	1.00000	6.80000	6.80010	0.00010
4	1.00000	1.12800	0.15900	2.32000	1.00000	1.00000	5.00000	5.00000	0.00000
5	1.00000	1.44300	0.15240	2.49000	1.00000	1.00000	3.80000	3.80000	0.00000
6	1.00000	1.09900	0.31800	2.52000	1.00000	1.00000	4.85000	4.85000	0.00000
7	1.00000	1.57000	0.15880	2.50000	1.00000	1.00000	3.49300	3.49310	0.00010
8	1.00000	4.07000	0.07940	2.59000	1.00000	1.00000	1.27000	1.26920	0.00080
9	1.00000	1.51000	0.31750	2.50000	1.00000	1.00000	3.49300	3.49300	0.00000
10	1.00000	0.82500	0.23500	4.55000	1.00000	1.00000	4.95000	4.94990	0.00010
11	1.00000	1.03000	0.23500	4.55000	1.00000	1.00000	3.97500	3.97500	0.00000
12	1.00000	1.36000	0.23500	4.55000	1.00000	1.00000	**2.99000***	2.99000	0.00000
13	1.00000	2.00300	0.23500	4.55000	1.00000	1.00000	2.00000	2.00010	0.00010
14	1.00000	3.75000	0.23500	4.55000	1.00000	1.00000	1.04000	1.04040	0.00040
15	1.00000	4.94500	0.23500	4.55000	1.00000	1.00000	0.77000	0.77020	0.00020
16	1.00000	4.42500	0.15875	2.65000	1.00000	1.00000	1.15000	1.15200	0.00200
17	1.00000	4.72300	0.15875	2.65000	1.00000	1.00000	1.07000	1.06820	0.00180
18	1.00000	5.22400	0.15875	2.65000	1.00000	1.00000	**0.96000***	0.96020	0.00020
19	1.00000	6.63400	0.15875	2.65000	1.00000	1.00000	0.74000	0.74000	0.00000
20	1.00000	6.07400	0.15875	2.65000	1.00000	1.00000	0.82000	0.81980	0.00020
Average Absolute Error in Training and Testing of CPMSA (cm) ⟶									0.00030
Designing of ETMSA									
21	2.00000	1.51900	0.07000	10.50000	1.00000	0.00000	**4.10000***	4.10010	0.00010
22	2.00000	2.63700	0.07000	10.50000	1.00000	1.00000	4.10000	4.10010	0.00010
23	2.00000	2.99500	0.07000	10.50000	2.00000	0.00000	4.10000	4.10000	0.00000
24	2.00000	3.97300	0.07000	10.50000	2.00000	1.00000	4.10000	4.09980	0.00020
25	2.00000	4.43900	0.07000	10.50000	3.00000	0.00000	4.10000	4.10000	0.00000
26	2.00000	1.28000	0.15900	2.32000	1.00000	0.00000	10.00000	9.99970	0.00030
27	2.00000	2.24200	0.15900	2.32000	1.00000	1.00000	10.00000	10.00010	0.00010
28	2.00000	2.55000	0.15900	2.32000	2.00000	0.00000	10.00000	10.00040	0.00040
29	2.00000	3.40000	0.15900	2.32000	2.00000	1.00000	10.00000	9.99990	0.00010
30	2.00000	3.82400	0.15900	2.32000	3.00000	0.00000	**10.00000***	10.00000	0.00000
31	2.00000	1.48900	0.07800	2.32000	1.00000	0.00000	**8.70000***	8.70010	0.00010
32	2.00000	2.59600	0.07800	2.32000	1.00000	1.00000	8.70000	8.70000	0.00000
33	2.00000	2.96900	0.07800	2.32000	2.00000	0.00000	8.70000	8.70000	0.00000
34	2.00000	3.96800	0.07800	2.32000	2.00000	1.00000	8.70000	8.70000	0.00000
35	2.00000	4.44300	0.07800	2.32000	3.00000	0.00000	8.70000	8.70000	0.00000
Average Absolute Error in Training and Testing of ETMSA (cm) ⟶									0.00009
Validation of CPMSA									
S. No.	**M**	**f(GHz)**	**h (cm)**	ε_r	**m**	**n**	**r (cm)**	**r_n (cm)**	**Error**
36	1.00000	1.17500	0.15900	2.32000	1.00000	1.00000	4.79340	4.80860	0.0152
37	1.00000	1.44300	0.14300	2.49000	1.00000	1.00000	3.80820	3.81030	0.0021
38	1.00000	1.09900	0.29800	2.52000	1.00000	1.00000	4.86510	4.77930	0.0858
39	1.00000	1.61000	0.26750	2.50000	1.00000	1.00000	3.32080	3.23500	0.0858
Average Absolute Error in Validation of CPMSA (cm) ⟶									0.4333

Datasets at Sr. No.1-3 used by Dahele and Lee [10], at Sr. No 4 by Dahele and Lee [11],at Sr. No 5 by Carver [12], at Sr. No.6 by Antoszkiewicz and Shafai [13] at Sr. No. 7-9 by Howell [14], at Sr. No.10-15 by Abboud et al [15], at Sr. No. 16-20 by Itoh and Mittra [16], at Sr. No 21-25 and 31-35 by Dahele and Lee[18], at Sr. No.26-30 by Dahele and Lee[17] at Sr. No.36-39 by equations (1)-(2). ***** represent testing datasets.

Domain specific priority based implementation of mobile services- an agile way

Dr. Daya Gupta[1], **Rinky Dwivedi** [2] **and Sinjan Kumar** [3]
[1, 2, 3] Computer Engineering Department, Delhi Technological University
, Delhi, India
[1] daya_gupta2005@yahoo.co.in
[2] rinkydwivedi@yahoo.co.in
[3] sinjan.dtu@gmail.com

Abstract - *In this era of globalization mobile companies are spreading their business across the world. This has resulted in a wide range of customers needs and mobile service providers are facing new challenges every day. We have examined the issues involved and found that a dynamic requirement from diverse domains is the most challenging task to be implemented while developing mobile communication software. In order to address this issue we are extending the notion of configurability in mobile communication. We will use traditional methodology for development of basic services because that is stable requirement and will act as the foundation for the further development. We will configure the on-demand services, which we will develop using agile methodology, on this basic component. In this paper we are proposing a Priority based Domain specific Mobile service Implementation (PDMI) process which helps an effective and fast mobile service implementation with high customer's satisfaction.*

Keywords: Agile software methods, CASE tool, Domain specific software engineering, Requirement based Mobile software development.

1 Introduction

First generation mobile systems were introduced in the early 1980's. These systems were designed to carry narrow–band circuit switched voice services and were based on analog frequency modulation. The mobile services gained a huge popularity and to fulfill the increasing customer demand second generation (2G) digital mobile systems were developed [7]. The 2G digital mobile systems were introduced in early 90's and are still occupying the market. These systems are based either on GSM (Global System for Mobile Communication) or CDMA (Code Division Multiple Access) technologies. The third generation (3G) mobile systems were introduced after the year 2000 that allows simultaneously use of speech and data services with higher data exchange rates. Presently 4G systems are under development, mostly based on multicarrier modulation such as Orthogonal Frequency Division Multiple Accesses (OFDMA) or Single Carrier Frequency Division Multiple Access (SC-FDMA). The application of mobile devices certainly improves and supports the lives of all age groups, but in our research we found that there exists a difference in the usage of mobile services for users of different age groups. The PDMI process proposed in this paper helps an effective, quick and customer's requirement based mobile service implementation.

Mobile domains have challenges of addressing dynamic requirements like gaming services, entertainment and lifestyle, etc. The agile methods have dynamic characteristics that help to facilitate the task of mobile software developers and to fulfill a wide range of customer's requirements in real quick time. We use light weight methods for satisfying volatile requirements for implementation purpose. The agile methods are an example of light weight methods which we have used for implementation. In addition to this we propose metric based calculation to establish that agile methodologies are suitable for dynamic requirement in mobile domain. We extend the configuration in method engineering to integrate agile method with traditional method which we are using for basic services.

In the next section we discuss popular agile methods. In section 3 we formulate our proposal for PDMI process. In section 4, we present a set of metrics for analysis of project characteristics that guide us to choose suitable methodology. In section 5, we propose to configure agile methodology with traditional method to develop services required in mobile domain. In section 6, the proposed PDMI process is implemented using view point oriented approach through agile methodology.

2 Agile Methodologies

In recent past agile methodologies are gaining popularity in project development [5, 13]. In this section we review some popular agile methodology and highlighted their underlying process, pros and cons and the guidelines to use them.
Agile software development is capable of producing quality software quickly and cheaply and is suited to project with changing requirement without excessive rework.

They do not require long term planning and use the strategy of incremental development through the team work where customer is also involved. As customer is directly involved in the project development so, it minimizes overall risk and adapt to changing requirement quickly.

2.1 **Scrum**

The scrum method is a general agile method that focuses on managing iterative development and does not adapt specific agile practices.

It has three phases, they are discussed below:-

1. Outline planning phase for designing software architecture derived from general objective of the project. This is the management phase where first general objectives are elicited these are then used to design software architecture.
2. Sprint cycle consisting of a product backlog which is converted into sprint backlog from where product is developed and delivered in increment. The cycle is repeated until the backlog is emptied.
3. Project closure phase which terminates the project and prepares the termination report and user manual.

Once a shippable module is delivered, the product backlog is analyzed for the next priority module. Module reprioritization can also be done if required.

Pros: Scrum focuses on management activities which makes the system more adaptive. It encourages team work enabling the large and complicated project to be developed by multiple teams.

Cons: Scrum projects have poor documentation which may result in dirty programming.

Use: Best suited for large and complex product which requires lot of management activity, cooperation between product developer and product manager.

2.2 Extreme-Programming (XP)

Extreme Programming (XP) is very famous agile methodologies. It takes extreme approach to iterative development and delivers working software frequently [14].

In this, involvement of customer during development is very essential. The development team works in very close environment in presence of customer. There are six phases in XP that are discussed below.

1. Requirement Phase: - In this phase users give requirements as story that are recorded on story cards then these story cards are prioritized.
2. Task gathering: - Here stories are broken into tasks.
3. Plan release phase: - In this phase the most prioritized stories are selected and planned for early release.
4. Development Phase: - The design of the only most prioritized task, which is to be developed, is done in this phase.
5. Release Phase: - The above designed task is developed and released for the use.
6. Evaluation of the system: - working of the released system is evaluated and the next cycle is started.

Pros: It allows developer to focus on coding resulting in fast software development that satisfies the user, does not require highly technical team. Management can deliver the working software for less cost and reduces the risk.

Cons: It requires dedicated effort from client/ user side they should have good knowledge of training on XP for project to be successful. Little documentation constraints the usability of the project.

Use: It is best suitable for project which require fine programming practices and for which testing before development is essential.

2.3 Feature-Driven Development (FDD)

FDD is model driven methodology that releases software in form of features in short iterations. It is suitable for large team and consists of very short phases and delivers specific features in each phase.

Some of the Feature Sets for our project are "Incoming Call," "Outgoing Call," "Messaging," and this comes under "Basic Services" Subject Area. In FDD, we do planning, designing and building of a feature under consideration. It consists of five specific processes in specified order, which is discussed below.

1. Develop an overall model: - Requirements are gathered in top down approach where all subject areas are designed. Subject areas are aggregation of feature set. Feature set are combination of feature. Each feature is task to be performed.
2. Build a list of feature: - Features gathered are compiled to form feature list.
3. Plan by feature: - Planning is done to build a feature.
4. Design by feature: - Proper designing is done for the planned feature.
5. Build by feature: - Actual implementation of the feature is done.

Pros: It focuses on design and code inspection which results in high quality software. It provides excellent documentation and support object oriented programming which are popular among programmers.

Cons: Requires highly skilled team. Delay in release due to design and modeling activity and cost may go up.

Use: Best suited for the development of the software in which focus should be on all the features in detail such as banking, insurance and finance.

3 Priority based Domain specific Mobile service Implementation (PDMI)

We divide mobile services in various groups that act as an input to our PDMI process.

Table1: Services provided by mobile service providers

Basic Services	Incoming and outgoing calls. Messaging (send, receive) a

	message service.
Entertainment and lifestyle	Movie reviews and movie tickets. Music & videos, Jokes. Chat Services. Caller tunes. News.
Gaming Applications	Online Games. Games download.
Business Applications	Stock market news. Mobile banking. Finance News.
Traditional Applications	Astrology. Vaastu. Spiritual listening

The basic services are considered as essential for mobile communication, which comes as a default to the customers. The other services are considered as optional to the customers whom they choose as per their requirements. A survey was conducted on 400 people of different age groups wherein the respondents were asked to set a priority for the optional services according to their requirements. The summary of the results are shown in the table 2.

Table 2: Priority of mobile services for various domains

Domain	Mobile Service	Priority
5-15 yrs (Group-1)	Entertainment and lifestyle. Gaming Applications. Business Applications. Traditional Applications	2 1 3 4
16-27 yrs (Group-2)	Entertainment and lifestyle. Gaming Applications. Business Applications. Traditional Applications	1 2 3 4
28-45 yrs (Group-3)	Entertainment and lifestyle. Gaming Applications. Business Applications. Traditional Applications	2 3 1 4
46-62 yrs (Group-4)	Entertainment and lifestyle. Gaming Applications. Business Applications.	3 4 1
	Traditional Applications	2
63 yrs. Onwards (Group-5)	Entertainment and lifestyle. Gaming Applications. Business Applications. Traditional Applications	3 4 2 1

The mobile service provider will provide basic services to all mobiles. Based on users detail such as age it will provide the optional services in the order of priority. Based on the survey and the present market demand the mobile provider should provide different type of mobiles for different domain (age group). The highly prioritized services will be in-built and the lower priority services will be configured on-demand. Users are free to choose the optional services and can also omit the low priority services. PDMI process will be used for maintaining the priority of services for different age groups.

4 Analysis of project characteristics for choosing suitable methodology

We have chosen 14 project metrics; some of these metrics are taken from COCOMO model [15]. We have divided weight between all the 14 metrics. Weights are assigned in such manner so that the metrics having more support for traditional has given more weight and the metrics having more support for agile has given less weight. We have divided input parameters in three categories. Table below describes the numerical equivalent value for different categories.

Table 3:-Numerical equivalent value for different category.

Category	Low	Medium	High
Value	1	5	9

4.1 Description of different metrics

Complexity:-This metric is the measure of effort required to develop the software. More complex project should be developed using traditional methodology. So, the more weight is given to this metric.

Time to Market: - How fast product is required in the market? If it is required quickly then we should follow agile methodology.

Risk Involved: - Impact of failure of software. If risk is high then we should follow traditional methodology.

Flexibility: - Flexibility is the effort required to modify an operational program. High flexibility has more support for agile methodology.

Modularization of Task: - Is it possible to divide the product into different modules? If its value is high that is, if modularization is easy then we can follow agile methodology and deliver different modules in increments.

Volatility of requirements: - At later stage of development, how much requirements may change? The more value of this metric has more support for agile methodology.
Expendability: - The ease with which changes can be made to the software at later stages. The more value of this metric has more support for agile methodology.
Coupling: - The degree of interdependence between classes. Coupling increases complexity and hence more value of this metric has more support for traditional methodology.
Cohesion: - The larger the similarity between methods, the more cohesive is the class. Low cohesive class is more complex.
Tool Experience: - How much year of work experience the developer has on the tool to be used?
Platform volatility: - How frequently the platform (Operating system for which product is being developed) is changing.
Application Experience: - what is the work experience of the developer on the desired application (application may be java or c etc.).
Programmer's capability: - How much capable the programmer is; for development of the project?

Table 4:- Weight, Input values and their sum of product for decision making

Name of Metrics (Project Characteristics)	Weight	Value of different metrics for Mobile App.	Product of weight and values
Complexity	0.15	Low (1)	0.15
Time to Market	0.08	High (9)	0.72
Risk Involved	0.15	Low (1)	0.15
Flexibility	0.05	High (9)	0.45
Modularization of Task	0.04	High (9)	0.36
Volatility of requirements	0.05	Medium (5)	0.25
Expandability	0.04	Medium (5)	0.20
Coupling	0.15	Low (1)	0.15
Cohesion	0.05	Medium (5)	0.25
Tool Experience	0.04	Medium (5)	0.20
Platform volatility	0.05	High (9)	0.45
Platform Experience	0.06	Low (1)	0.06
Application Experience	0.04	Medium (5)	0.20
Programmers Capability	0.05	Medium (5)	0.25
Total (Sum of Product)			**3.84**

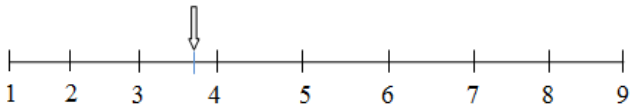
Calculation:-

$$\textbf{Total (sum of product)} = \sum_{i=1}^{14} \textbf{(weight * Input value)}$$

Total = (0.15*1) + (0.08*9) + (0.15*1) + (0.05*9) + (0.04*9) + (0.05*5) + (0.04*5) + (0.15*1) + (0.05*5) + (0.04*5) + (0.05*9) + (0.06*1) + (0.04*5) + (0.05*5)

Total=3.84
Output (Total sum of product) will range from 1 to 9. If this value comes between 1 and 4 then it indicates that we should follow agile methodology. If it comes between 6 and 9 then it indicates that we should follow traditional methodology. If value comes between 4 and 6 then we can consider any one or hybrid methodology. This is shown on the scale given below.

Fig. 1:-Scale for selection of appropriate methodology



For our project the final outcome is 3.84 so we have chosen agile methodology.

5 Configurability in mobile communication domain

The IEEE glossary meaning of the configuration is "The arrangement of a computer system or component, defined by the number, nature, and interconnection of its constituent part" [9, 10]. It has been used to model business processes that are similar to one another in many ways yet differ in some other way. Configurability uses the notion of commonality (commonality that is uniform across a given set of objects) and variability (how members of a family may differ from one another) [11, 12]. We have extended the notion of configurability in method engineering and presented a meta model for configuring situation specific method [16]. It uses the notion of method essentiality and method variability. We apply this notion to configure a suitable method for mobile domain.
According to our proposal the basic services are the essential services and without them, mobile communication will lose its identity. All the other services can be considered as variable services and can be selected as per the requirement of the user. For our PDMI process, method essentiality constitutes a method that provides basic services. We conjecture that traditional methodology will be suitable for it. As evaluated in the previous section for variable mobile services agile methodology is suitable, hence using the method configurability a suitable method will be configured. In this paper we focus on selection of suitable agile methodology and the detail of the method configuration will be dealt in next paper.

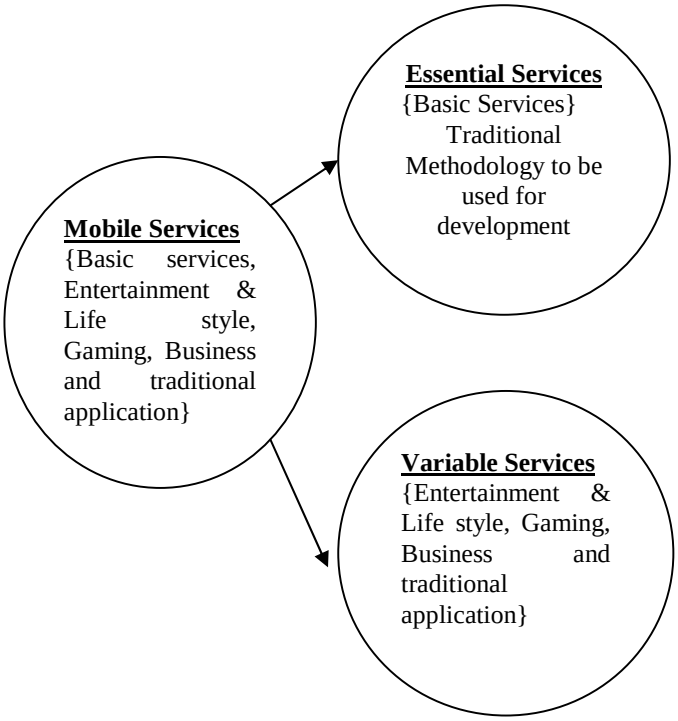


Fig. 2: Configuration for Mobile Services domain.

6 Implementation of PDMI Process through Agile Methodology

As described in section 2 agile methodology comprises of many agile methods like XP, SCRUM, DSDM, FDD, AUP etc. Each agile method [2, 3, 6] has its unique features and is good for one or other task**.** On the basis of this study we found that Scrum is best suitable methodology for our project, because we are going to develop our PDMI tool which is more lenient to project management.

For our project volatility of requirement is high. That means the requirement may change frequently which require reprioritizing the sprint backlog. Scrum supports reprioritization of backlog. On the other hand, as we have discussed that prioritization is different for different domain (age group) so we have come to a decision that mobile provider should target the entire domain and they should deliver various versions of the same product on the basis of demand in market. It requires a lot of management activity and they have to maintain the different backlog for different domain, so scrum is best suitable methodology for our project.

Scrum supports software development team to a great extent and also gives them [1, 4] full independence to perform. Two major stakeholders in a Scrum project are - Scrum master and Product owner. Scrum master helps the team members use the scrum framework and Product owner guides the team towards building the right product. Scrum is a process skeleton that contains sets of practices and predefined roles. The main actors in Scrum projects are:

Scrum Master- Scrum master manages the whole project, it gets project requirements from actors, decompose the requirements into modules in association with the project manager.

Product Owner-Product owner represents the stakeholders of the project.

Team– A group of about 7 people develops the modules and subsequently the project.

Our PDMI process is based on well known process of View Point Oriented Approach [8]. For each associated actor we create a different view, which describes the functionality of that particular stakeholder.

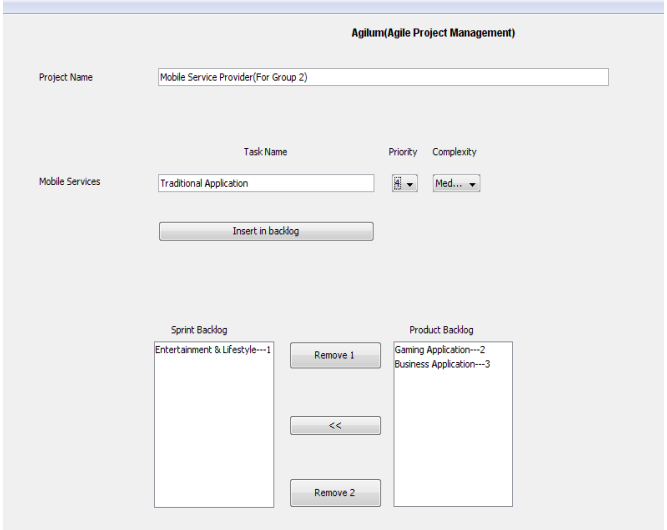
1. Scrum Master View
2. Sprint Planning Meeting View
3. Development Team View

Table 5: Functionalities of associated actors

View Points	Functionalities
Scrum Master View	1. Identify project domain. 2. Enter project requirements and their priority in association with product owner. 3. Decomposition of project into modules.
Sprint Planning Meeting View	1. Decompose highest priority module into the task and prioritization of tasks. 2. Select tasks to perform in current sprint. 3. Assign the task to the appropriate team on the basis of task complexity and team expertise. 4. Remove task from sprint on its completion.
Team View	1. Pickup the task to perform. 2. Make entry for already burned-down task. 3. Submit the executables.

The idea behind using the agile methodology in our PDMI is, agile methodology believes in releasing functional product at the end of every iteration. The whole process starts with the requirements gathering phase, by the scrum master from the actors, the prioritization of requirements is done by actors in association with the scrum master. The highest priority tasks are forwarded to project development teams for an early release. As the software launches in the market, it starts generating revenues that helps the service providers to develop and implement other priority services.

We have developed a prototype Case tool that supports scrum development adapted to our proposal. Based on the priority of the services it distributes the task between product backlog and scrum backlog.



7 Conclusion

Priority based Domain specific Mobile service Implementation process helps an effective and fast mobile service implementation with high customer's satisfaction. Method configuration process can be used to integrate traditional method for basic services and agile methodology for variable services. This process is an advantageous solution for both service providers and customers. Service providers may start their services with the most basic ones and keep on increasing the add-ons as per the priorities of the customer whereas customers may have effective, quick and requirement based mobile service.

8 References

[1] K. Schwaber, M. Beedle, *Agile Software Development with SCRUM*, Prentice Hall, 2002.

[2] K. Petersen and C. Wohlin, *A Comparison of Issues and Advantages in Agile and Incremental Development between State of the Art and an Industrial Case*, Journal of Systems and Software, Vol. 82, No. 9, pp. 1479-1490, 2009.

[3] Abrahamsson P., Warsta J., _Siponen M.,Ronkainen J. *New directions on agile methods: a comparative analysis***,** Proceedings of the 25th International Conference on Software Engineering,2003 .

[4] A. Cockburn, *Agile Software Development*, Addison-Wesley, Boston, 2002.

[5] Nerur, S., Mahapatra, R. and Mangalaraj, G, *Challenges of Migrating to Agile Methodologies*, Communications of the ACM, Vol. 48, No. 5, May 2005, pp. 72-78.

[6] Chau, T.; Maurer, F.; Melnik, G.(2003). *Knowledge sharing: agile methods vs. Tayloristic methods*, Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003. WET ICE 2003. Proceedings.

[7] Mobile Electronic Transactions Ltd. MeT White Paper on Mobile Transactions (2003); www.mobiletransaction.org.

[8] Summerville, I., "Software Engineering" seventh edition 2003.ISBN-8129708671.Pearson Education.

[9] Coplien J., Hoffman D., Weiss D., *Commonality and Variability in Software Engineering*, IEEE Software, 37-45, 1998

[10] Weiss D. M., Lai C. T. R., *Software Product-line engineering: A Family based Sofwtare development Process*, Addison Wesley, 1999.

[11] Karlsson, F., Ågerfalk, P.J. (2004) *Method Configuration: Adapting to Situational Characteristics While Creating Reusable Assets.* In Information and Software Technology 46, 2004, 619-633.

[12] Wistrand, K., Karlsson, F. (2004) *Method Components – Rationale Revealed*. In Advanced Information Systems Engineering 16th International Conference, CAiSE 2004, Riga, Latvia, June 7-11, 2004, Proceedings, A. Persson, J. Stirna, Eds. Springer-Verlag, LNCS 3084, Berlin, 2004, 189-201.

[13] Miller, G., *The characteristics of agile software processes.* The 39th international conference of object-oriented language and system(TOOLS39),Santra Barbara, CA, 2001.

[14] Beck, K. (1999a). *Embracing change with Extreme programming*. IEEE computer 32(10):70-77.

[15] Chen. Z**.** *Finding the right data for software cost modeling.*. Center for Software Eng., Univ. of Southern California, Los Angeles ,CA, USA Menzies,T. ; Port,D. ; Boehm,D. Volume: 22 ,Issue:6

[16] Gupta D. and Dwivedi R. *A Step Towards Method Configuration from Situational Method Engineering* .Software Engineering: An International Journal(SEIJ), INDIA, 2012 (51-5

ARTICLE IN PRESS

Optics Communications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Dual-core leaky optical waveguide as an integrated-optic polarizer

Nandam Ashok [a,*], Vipul Rastogi [a], Ajeet Kumar [b]

[a] Department of Physics, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India
[b] Department of Applied Physics, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Article history:
Received 6 April 2012
Received in revised form
10 July 2012
Accepted 13 July 2012

Keywords:
Integrated optics
Planar waveguide
Dual-core waveguide
Leaky waveguide
Waveguide polarizer
Polarization filter

ABSTRACT

We present a design of a dual-core leaky waveguide that can be used as an integrated-optic polarizer. The proposed polarizer works on the principle of mode filtering. The structure is characterized by two cores, namely core-1 and core-2 and a high index layer in the upper-most region, such a structure supports leaky modes. The leakage losses of the modes have been calculated by using the transfer matrix method (TMM). Single polarization operation is ensured by high differential leakage loss between fundamental TE and TM modes. We show TE-pass operation with 0.5 dB/mm loss in TE_0 mode and 13 dB/mm loss in TM_0 mode, and TM-pass operation with 0.36 dB/mm loss in TM_0 mode and 7.45 dB/mm loss in TE_0 mode. Besides single-polarization operation, single-mode operation of the structure is ensured by high leakage loss of all the other higher-order modes.

1. Introduction

Optical waveguide polarizers have attracted considerable interest in integrated optics for their application in coherent optical communication systems and gyroscopes. The possibility of using different materials and integration with other components [1–3] makes the waveguide polarizer more attractive. An optical waveguide polarizer is a device that passes only one polarization state while the other polarization state is eliminated. The design of waveguide polarizers is difficult, because the lowest order TE and TM mode cut off conditions are very close to each other. Several alternatives to design the waveguide polarizer have been reported in the literature. Thyagarajan et al. have proposed a waveguide polarizer employing the resonant tunneling effect in a non-identical directional coupler [4]. The TM mode of the primary waveguide resonantly couples to a thin high-index waveguide layer and subsequently leaks out to the surrounding overlay, resulting in a TE-pass polarizer. However, in this design the high-index guiding layer is extremely thin and can have fabrication tolerance issues. In another work Jindal and Kumar carried out a comprehensive study of a waveguide polarizers based on the resonant tunneling [5]. Apart from these there have been several studies on waveguide polarizers based on metal-clad structures [6–10]. However, metal-clad optical waveguide based polarizers exhibit a high propagation loss. There have also been reports on waveguide polarizer involving birefringent polymer on glass [11] multilayer structure [12] and lithium niobate waveguide polarizers [13,14]. Long-period waveguide grating structures can also be utilized as effective waveguide polarizers [15,16]. Use of photonic band gap structures is another way of achieving single polarization [17]. Over the past few years we have been designing leaky optical waveguides for potential applications in high power amplifiers and lasers [18–20]. These designs work on the principle of higher-order mode discrimination and can lead to single-mode operation of a thick-core and otherwise multimode waveguides. In this paper we extend our leaky waveguide work [20] to design a waveguide polarizer using InGaAs/AlGaAs semiconductor material layers on a glass substrate. Such systems of materials can cater the range of wavelengths 0.7–1.55 μm. The design is based on the mode discrimination. Our design supports 2 TE and 2 TM modes but all the modes are leaky due to the high-index uppermost leaky layer. We have analyzed the proposed structure by calculating the leakage losses of modes by using the TMM. To achieve an efficient TE/TM pass polarizer, we have chosen the design parameters in such a way that TE_0/TM_0 mode of the structure propagates with low loss and TM_0/TE_0 mode leaks out through the leaky layer. Single-mode operation of the structure is also ensured by having high leakage loss for all the higher-order modes. The proposed design is amenable to fabrication by the Epitaxial lift-off technology, which is suitable for fabrication of high quality semiconductor material layers on arbitrary substrates including glass [21,22]. Numerical simulations predict that with the appropriate parameters we can achieve a polarization extinction ratio (PER) of 58 dB and insertion loss (IL) of

* Corresponding author. Tel.: +91 9219604999.
E-mail address: nandam.ashok@gmail.com (N. Ashok).

0030-4018/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.optcom.2012.07.055

0.91 dB for TE pass with a device length of 5 mm and a PER of 73 dB with IL of 0.13 dB for TM pass with a device length of 10 mm at 700 nm wavelength.

2. Waveguide design and method of analysis

The refractive index profile of a dual-core leaky planar waveguide is shown in Fig. 1. The design consists of two cores namely core-1 and core-2 of same refractive indices n_1 and widths a_1 and a_2 respectively formed on a substrate of refractive index n_s. The two cores are separated by an inner clad of refractive index n_2 and width b. The structure is made leaky by having the uppermost leaky layer of refractive index n_1, which is separated from core-2 by an outer clad of refractive index n_3 and width c. Here we have considered step-index profiles in the two cores. However, the design can always be carried out for graded-index profile cores such as the ones obtained in ion-exchanged waveguides. All the modes supported by the waveguide are leaky in nature and have complex propagation constants. The propagation constants of the modes can be calculated by using the transfer matrix method [23,24], which is an established method for multilayer structure analysis. To implement the TMM in multilayer structure, one can write the electric field variation in the ith layer of width d_i and refractive index n_i as

$$E_i = \begin{cases} A_i \cos[k_i(x-d_{i+1})]+B_i \sin[k_i(x-d_{i+1})], & \kappa_i^2 > 0 \\ A_i \cosh[k_i(x-d_{i+1})]+B_i \sinh[k_i(x-d_{i+1})], & \kappa_i^2 < 0 \end{cases}$$

$$k_i = |\kappa_i|,\ \kappa_i^2 = k_0^2(n_i^2 - n_{eff}^2),$$

$$k_0 = 2\pi/\lambda,\ n_{eff} = \beta/k_0 \quad (1)$$

where k_0, β, and n_{eff}, are the free space wavenumber, propagation constant and effective index of the mode respectively. Applying the appropriate boundary conditions, the field coefficients in the ith and $(i+1)$th layers can be linked through the given matrix form:

$$\begin{pmatrix} A_{i+1} \\ B_{i+1} \end{pmatrix} = S_i \begin{pmatrix} A_i \\ B_i \end{pmatrix} \quad (2)$$

For TE mode S_i can be expressed as

$$S_i = \begin{pmatrix} \cos\Delta_{i+1} & -(k_i/k_{i+1})\sin\Delta_{i+1} \\ \sin\Delta_{i+1} & (k_i/k_{i+1})\cos\Delta_{i+1} \end{pmatrix} \quad \text{for } \kappa_i^2 > 0$$

$$= \begin{pmatrix} \cosh\Delta_{i+1} & (k_i/k_{i+1})\sinh\Delta_{i+1} \\ \sinh\Delta_{i+1} & (k_i/k_{i+1})\cosh\Delta_{i+1} \end{pmatrix} \quad \text{for } \kappa_i^2 < 0 \quad (3)$$

where S_i is the transfer matrix of the ith layer and $\Delta_i = k_i(d_i - d_{i-1})$. One can relate the field coefficients of the last layer A_N and B_N in terms of the first layer A_1 and B_1:

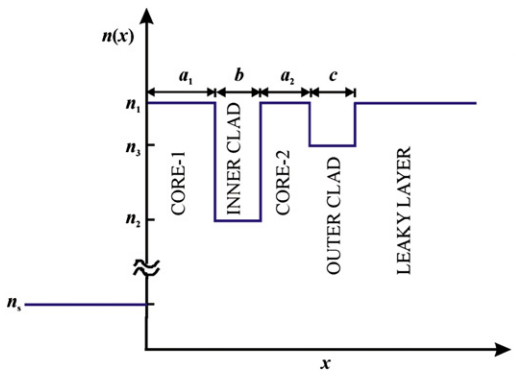


Fig. 1. Refractive index profile of the proposed dual-core leaky waveguide.

$$\begin{pmatrix} A_N \\ B_N \end{pmatrix} = f \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} \quad (4)$$

Here $f = S_{N-1}S_{N-2}\ldots S_1$ and

$$f = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \quad (5)$$

For a leaky waveguide, in the high-index outermost region only the outgoing wave exists and no amplifying term is allowed in the low-index substrate region. Applying these conditions we obtain $A_1 = iB_1$, and $A_N = -B_N$. From Eqs. (4) and (5), we can write

$$A_N = (f_{12} + if_{11})B_1 \text{ and } B_N = (f_{22} + if_{21})B_1 \quad (6)$$

We can get an eigenvalue equation $F(\beta)=0$ from Eq. (6), and the propagation constant (β) can be determined from the eigenvalue equation. The complex propagation constant of the waveguide can be written as $\beta = \beta_r + i\beta_i$. A plot of $1/|F|^2$ as a function of β shows a number of resonance peaks which are Lorentzian in shape. The value of β corresponding to a resonance peak gives the real part of the propagation constant β_r and the FWHM of the Lorentzian gives the imaginary part β_i of a mode. The real part of the propagation constant estimates the effective index of the mode, while the imaginary part estimates the leakage loss of the mode. The leakage loss of the mode has been utilized to achieve the single polarization.

3. Numerical results and discussion

The proposed polarizer works on the principle of mode-discrimination. The parameters of the structure can be appropriately chosen to resonantly leak out either of the TE and TM modes and to retain only one polarization. We have studied the effect of various design parameters to optimize the performance of the polarizer. First of all, we have tuned the refractive index of outer clad n_3 to achieve the desired condition of resonance coupling and then have studied the effect of other parameters to achieve the desired single-polarization and single-mode (SPSM) output. We have chosen the following parameters for numerical calculations:

$$n_s = 1.5,\ n_1 = 3.39,\ n_2 = 3.15,\ a_1 = 1.2\ mm,$$
$$b = 0.80\ \mu m,\ c = 0.80\ \mu m,\ a_2 = 0.87\ \mu m, \text{ and } \lambda = 1.55\ \mu m \quad (7)$$

Variation of leakage losses of the fundamental TE and TM modes with n_3 is shown in Fig. 2. Leakage loss curves show clear resonance peaks at $n_3 = 3.302$ for TM_0 mode and at $n_3 = 3.322$ for TE_0 mode. This is the result of resonant coupling of TM_0/TE_0 mode of core-1 to highly leaky core-2. Operating the structure at $n_3 = 3.302$ results in a TE-pass polarizer while with $n_3 = 3.322$ the structure can be used as a TM-pass polarizer. The differential leakage can be further enhanced by optimizing various other parameters to obtain an efficient integrated optic polarizer. To illustrate the resonance more clearly we have also plotted the spectral variations of effective indices of TM_0 modes of individual cores for TE-pass polarizer as shown in Fig. 3. We can see that the curves intersect around $\lambda = 1.545$ μm, which is the resonance wavelength corresponding to waveguide parameters defined by Eq. (7).

3.1. TE-pass polarizer

A TE-pass polarizer requires low loss for TE_0 mode and high loss for TM_0 and other higher order modes. To achieve this we have optimized coupling of power from TM_0 mode of core-1 to highly leaky core-2 by tuning the value of outer clad index n_3 as shown in Fig. 2. From the figure it is clear that at $n_3 = 3.302$ there

Please cite this article as: N. Ashok, et al., Dual-core leaky optical waveguide as an integrated-optic polarizer, Optics Communications (2012), http://dx.doi.org/10.1016/j.optcom.2012.07.055

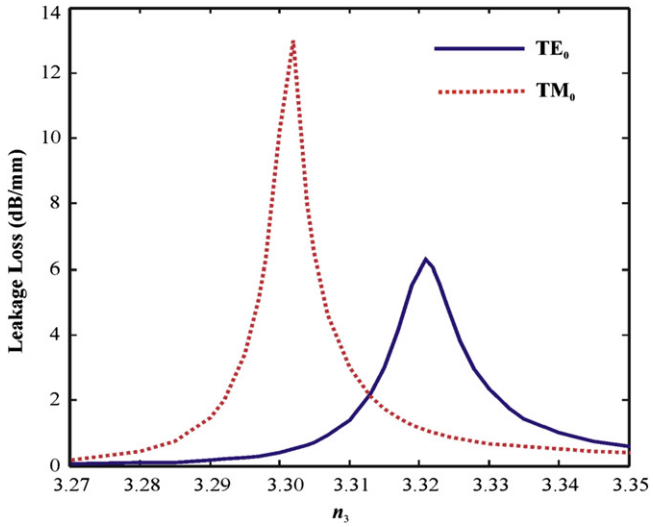


Fig. 2. Variation of leakage losses of the modes with outer-clad refractive index n_3 for the parameters defined by Eq. (7).

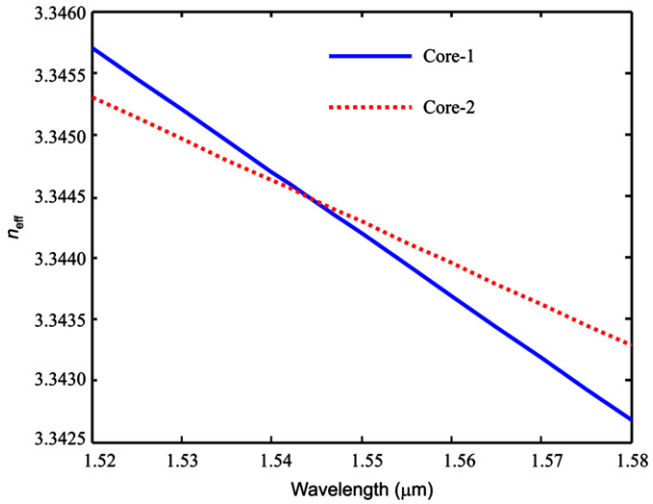


Fig. 3. Spectral variation of effective indices of TM_0 modes of individual cores corresponding to waveguide parameters defined by Eq. (7).

is resonant coupling of power and the leakage loss of the TM_0 mode is maximum. With these parameters the waveguide helps in leaking out TM_0 mode while introducing nominal loss to TE_0 mode. The mode fields of the passing mode (TE_0) and the discriminated mode (TM_0) at resonance are shown in Fig. 4. We can see that at resonance significant the amount of power appears in core-2 for the TM_0 mode. However, for the TE_0 mode which is far from the condition of resonance, the modal field is slightly affected by the presence of outer core. The structure also causes sufficiently high loss to all the other higher order modes and leads to SPSM operation. The SPSM operation of the structure depends on various waveguide parameters such as width of inner and outer clad and refractive index of core-2. Fig. 5 shows the effect of outer clad thickness c on the leakage losses of various modes supported by the structure. We can see that for small values of c the leakage losses of all the higher order modes are very high in comparison to the loss of TE_0 mode. However, the loss of TM_0 mode is also small and the differential leakage loss between the TE_0 and TM_0 modes is low. At around $c=0.8$ μm the leakage loss of TM_0 mode peaks up, which is the signature of resonant

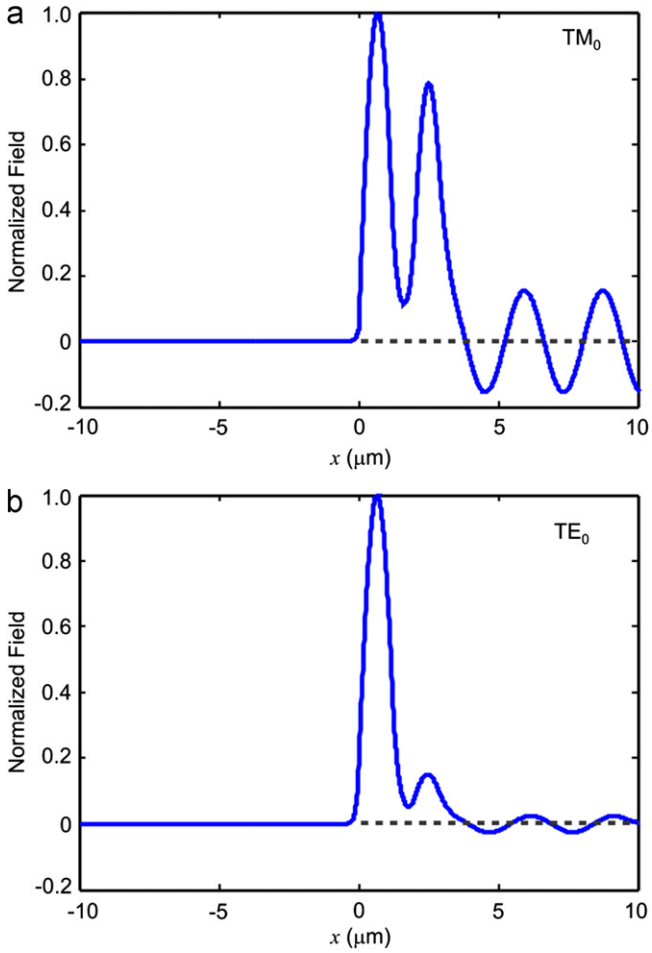


Fig. 4. (a) Mode field of TM_0 mode (discriminated mode) at the condition of resonance ($n_3=3.302$). (b) Mode field of TE_0 mode (passing mode) at $n_3=3.302$.

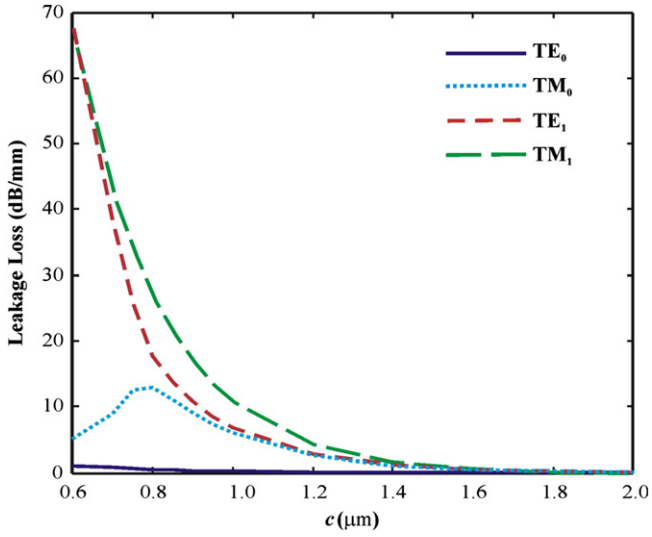


Fig. 5. Variation of leakage losses of different supported modes as a function of c for $n_3=3.302$.

coupling. At this value of c there is a high differential leakage loss between the TE_0 and TM_0 modes and also much higher leakage loss in all the other higher order modes. With this particular value of c the structure effectively behaves as a TE-pass polarizer.

Please cite this article as: N. Ashok, et al., Dual-core leaky optical waveguide as an integrated-optic polarizer, Optics Communications (2012), http://dx.doi.org/10.1016/j.optcom.2012.07.055

The structure with $c=0.8$ μm introduces a loss of 0.5 dB/mm to TE_0 and 13 dB/mm loss to the TM_0 mode for TE-pass operation at 1.55 μm wavelength.

Next, we have studied the effect of inner-clad thickness, b on the leakage losses of the different modes. To illustrate this we have plotted the variation of leakage losses of various modes with the inner-clad thickness, b at $c=0.8$ μm. The loss curves are shown in Fig. 6 and help us to choose the design parameter b. We can see that the leakage losses of all the modes decrease with b. However, the behavior of TM_0 mode is relatively flat up to $b=0.8$ μm and then the loss decreases sharply. We, therefore, choose $b=0.8$ μm to have high differential leakage loss between the TE_0 and TM_0 modes and at the same time keeping the loss of TE_0 mode as low as possible. One can note that, leakage loss of TE_0 mode increases with decrease in the value of b. This shows that loss of TE_0 mode is sensitive to only smaller values of b and discrimination of TM_0 and higher order modes is difficult with large values of b.

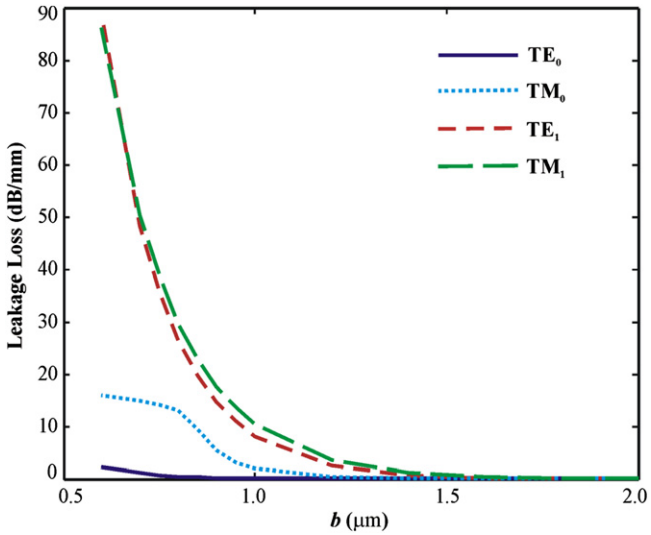


Fig. 6. Variation of leakage losses of modes as a function of inner-clad width b for $c=0.8$ μm and $n_3=3.302$.

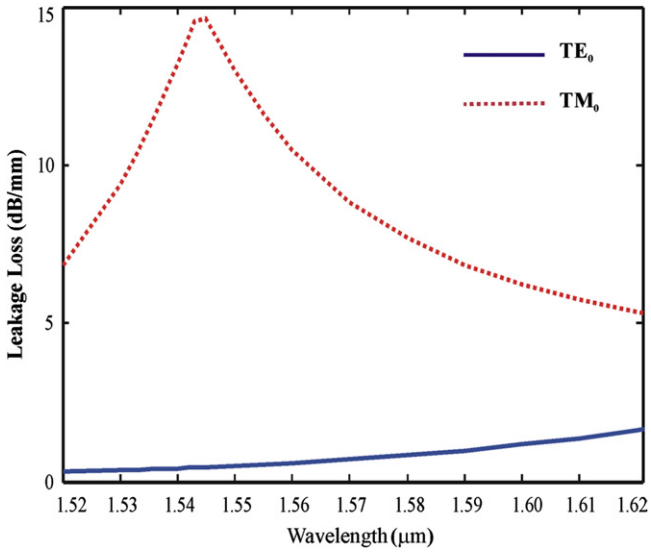


Fig. 7. Spectral variation of leakage losses of the modes of TE-pass filter for $n_3=3.302$, $c=0.8$ μm, and $b=0.8$ μm.

Spectral performance of the TE-pass filter has also been studied and is shown in Fig. 7. A peak in the loss curve of the TM_0 mode is observed at around 1.55 μm, which is the signature of resonance. One can have a 20 dB discrimination against TM_0 mode in the wavelength range 1.52–1.59 μm by a 3 mm long waveguide and obtain a TE-pass polarizer. Such spectral characteristics of the structure can also be used to design a band rejection filter for a particular polarization. The width of the rejection band can be controlled by the device length. For example, a 2 mm long structure can discriminate a wavelength band of 1.532–1.563 μm corresponding to TM-polarization by introducing 20 dB loss.

PER and IL are the important parameters of a polarizer when used in an optical system [25]. We have designed polarizers for different wavelengths and have calculated their PER and IL. The results for a 5 mm long waveguide are listed in Table 1. We have also plotted the variation of PER and IL as a function of waveguide length in Fig. 8. We can see that there is a trade-off between the PER and IL. At 1.55 μm wavelength the IL for waveguide length, $L=5$ mm is 2.5 dB, which is quite high. In order to use it with relatively smaller IL of about 1 dB, one has to be content with a PER of 25 dB. However, at 0.7 μm wavelength the structure offers very high PER of 58 dB with mere 0.91 dB IL when operated with $L=5$ mm.

3.2. TM-pass polarizer

TM-pass operation can be achieved by choosing $n_3=3.322$ where the loss of TE_0 mode peaks up as shown in Fig. 2. At this value of n_3 the leakage losses of the TE_0 and TM_0 modes are 6 and 1 dB/mm respectively. The differential leakage loss between the two modes can be further increased by optimizing the values of inner and outer clad widths.

In Fig. 9 we have plotted the variation of leakage losses of the modes with outer clad width c. We can see that leakage loss of TM_0 mode does not vary sharply with c. The leakage loss of the TE_0 mode peaks up at around $c=1$ μm due to the resonant coupling. Clearly, the value of c around 1 μm results in maximum differential leakage loss of the modes and is the desired value for outer clad width for a TM-pass polarizer. Fig. 10 illustrates the variation of leakage losses of the modes as a function of inner-clad width, b for $c=1$ μm. Differential leakage loss decreases with b. Small inner-clad width is preferable for efficient mode filtering. $b=0.85$ μm introduces a high differential leakage loss between TE_0 and TM_0 modes and a nominal loss to TM_0 mode for TM-pass polarizer.

The spectral variation of leakage losses of the modes for TM-pass filter is shown in Fig. 11. From the figure one can note that a 5 mm long waveguide can work efficiently as a TM-pass filter in the wavelength range of 1.51–1.59 μm by introducing a 20 dB loss to TE_0 mode.

Table 1
Dual-core leaky waveguide-based TE-pass polarizers designed for various wavelengths.

Wavelength (μm)	Design parameters	PER (dB)	IL (dB)
1.55	$a_2=0.87$ μm, $b=0.80$ μm, $c=0.80$ μm, $n_3=3.302$	62.5	2.5
1.30	$a_2=0.91$ μm, $b=0.70$ μm, $c=0.74$ μm, $n_3=3.313$	60.5	1.0
0.700	$a_2=0.98$ μm, $b=0.36$ μm, $c=0.45$ μm, $n_3=3.351$	58.0	0.91

Please cite this article as: N. Ashok, et al., Dual-core leaky optical waveguide as an integrated-optic polarizer, Optics Communications (2012), http://dx.doi.org/10.1016/j.optcom.2012.07.055

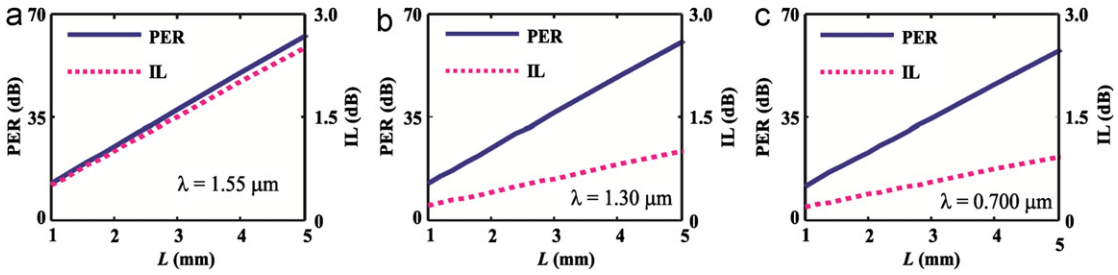


Fig. 8. PER and IL of the TE-pass polarizers designed at different wavelengths: (a) 1.55 μm, (b) 1.3 μm, and (c) 0.7 μm.

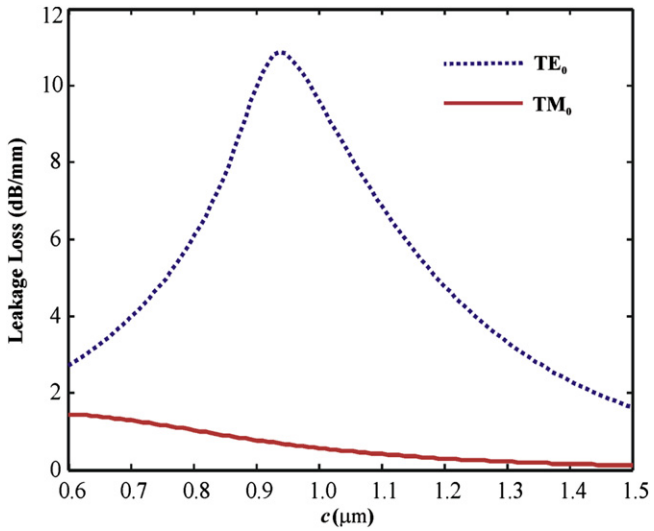


Fig. 9. Variation of leakage losses of the modes as a function of c for $n_3=3.322$.

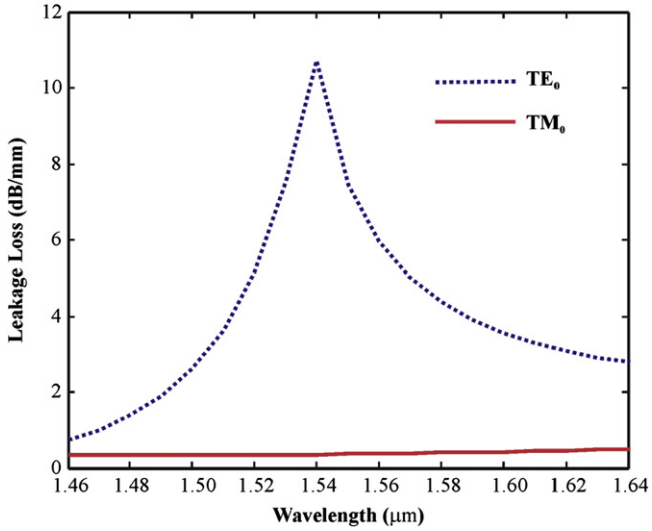


Fig. 11. Spectral variation of leakage losses of the modes of TM-pass filter for $n_3=3.322$, $c=1$ μm, and $b=0.85$ μm.

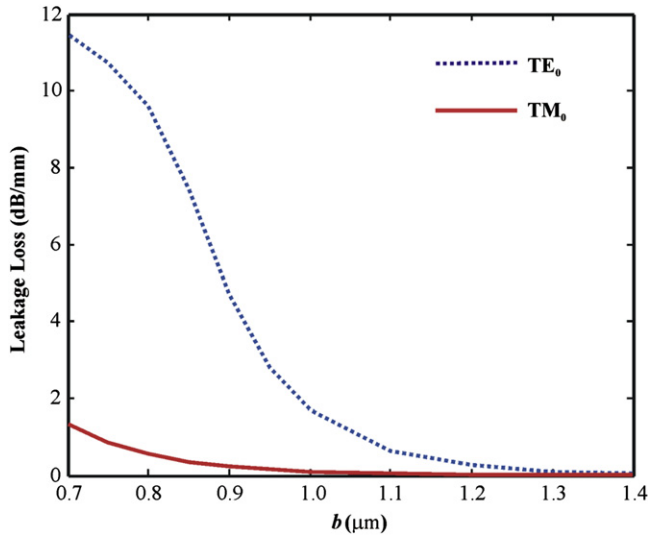


Fig. 10. Variation of leakage losses of modes as a function of inner-clad width b for $c=1$ μm and $n_3=3.322$.

In Table 2 we have listed the design parameters, PER and IL of the designed polarizers at 1.55-, 1.30-, and 0.7 μm wavelength for $L=10$ mm. In Fig. 12 we have plotted the PER and IL of the design as a function of length for three different values of wavelength. We could achieve an IL as low as 0.13 dB and PER of 73 dB at 0.7 μm wavelength.

Table 2
Dual-core leaky waveguide-based TM-pass polarizers designed for various wavelengths.

Wavelength (μm)	Design parameters	PER (dB)	IL (dB)
1.55	$a_2=0.87$ μm, $b=0.85$ μm, $c=1.00$ μm, $n_3=3.322$	71	3.6
1.30	$a_2=0.93$ μm, $b=0.71$ μm, $c=0.85$ μm, $n_3=3.323$	70	1.7
0.700	$a_2=0.98$ μm, $b=0.45$ μm, $c=0.59$ μm, $n_3=3.358$	73	0.13

To implement this planar structure into a practical device, fabricating channels of large widths and using appropriate out-coupling optics would be required.

4. Conclusions

We have presented a dual-core resonant leaky waveguide structure as a TE or TM-pass polarizer. The proposed polarizer works on the principle of polarization mode discrimination in conjunction with resonant tunneling. The structure has been analyzed and the design has been carried out for an SPSM device by using the TMM. We show TE-pass operation with TE_0 loss of 0.5 dB/mm and TM_0 loss of 13 dB/mm and TM-pass operation with TM_0 loss of 0.36 dB/mm and TE_0 loss of 7.45 dB/mm at 1.55 μm wavelength. A waveguide with such a single-polarization

Please cite this article as: N. Ashok, et al., Dual-core leaky optical waveguide as an integrated-optic polarizer, Optics Communications (2012), http://dx.doi.org/10.1016/j.optcom.2012.07.055

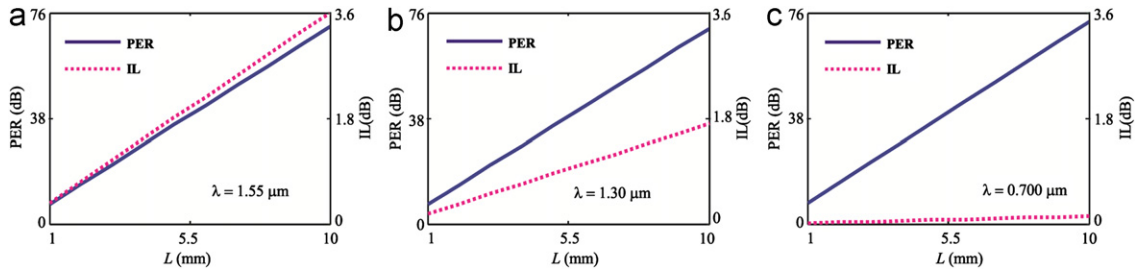


Fig. 12. PER and IL of the TM-pass polarizers designed at different wavelengths: (a) 1.55 μm, (b) 1.3 μm, and (c) 0.7 μm.

single-mode operation should be attractive for coherent optical communication systems and gyroscopes.

Acknowledgements

One of the authors (Nandam Ashok) acknowledges the financial support provided by Indian Institute of Technology, Roorkee, Ministry of Human Resources and Development (MHRD), Government of India. This work has been partially supported by UK–India Education and Research Initiative (UKIERI) Major Award.

References

[1] G.L. Bona, Microsystem Technologies 9 (2003) 291.
[2] A. Boltasseva, T. Nikolajsen, K. Leosson, K. Kajer, M.S. Larsen, S.I. Bozhevolnyi, Journal of Lightwave Technology 23 (2005) 413.
[3] R. Selvaraj, H.T. Lin, J.F. McDonald, Journal of Lightwave Technology 6 (1988) 1034.
[4] K. Thyagarajan, S.D. Seshadri, A.K. Ghatak, Journal of Lightwave Technology 9 (1991) 315.
[5] R. Jindal, A. Kumar, Fiber and Integrated Optics. 17 (1998) 299.
[6] J.N. Polky, G.L. Mitchell, Journal of the Optical Society of America 64 (1974) 274.
[7] H.A. Jamid, S.J. Al-Bader, Electronics Letters 24 (1988) 229.
[8] K. Thyagarajan, S. Diggavi, A.K. Ghatak, W. Johnstone, G. Stewart, B. Culshaw, Optics Letters 15 (1990) 1041.
[9] C.H. Chen, L. Wang, IEEE Journal of Quantum Electronics 34 (1998) 1089.
[10] L. Sun, G.L. Yip, Applied Optics 33 (1994) 1047.
[11] A. Morand, C. Sanchez-Perez, P. Benech, S. Tedjini, D. Bosc, IEEE Photonics Technology Letters 10 (1998) 1599.
[12] M. Kobayashi, H. Terui, K. Egashira, Applied Optics 17 (1978) 486.
[13] H. Lin, J. Ning, F. Geng, Chinese Optics Letters 2 (2004) 89.
[14] R.C. Twu, C.C. Huang, W.S. Wang, Microwave and Optical Technology Letters 48 (2006) 2312.
[15] K.S. Chiang, C.K. Chow, H.P. Chan, Q. Liu, K.P. Lor, Electronics Letters 40 (2004) 422.
[16] K.S. Chiang, C.K. Chow, H.P. Chan, K.P. Lor, IEEE Photonics Technology Letters 18 (2006) 1109.
[17] D. Zhao, B. Shi, Z. Jiang, Y. Fan, X. Wang, Applied Physics Letters 81 (2002) 409.
[18] A. Kumar, V. Rastogi, K.S. Chiang, Applied Physics B 85 (2006) 11.
[19] A. Kumar, V. Rastogi, K.S. Chiang, Applied Physics B 90 (2008) 507.
[20] A. Kumar, V. Rastogi, Applied Physics B 92 (2008) 577.
[21] E. Yablonovitch, E. Kapon, T.J. Gmitter, C.P. Yun, R. Bhat, IEEE Photonics Technology Letters 1 (1989) 41.
[22] J.F. Klem, E.D. Jones, D.R. Myers, J.A. Lott, Journal of Applied Physics 66 (1989) 459.
[23] M.R. Ramadas, E. Garmire, A.K. Ghatak, K. Thyagarajan, M.R. Shenoy, Optics Letters 14 (1989) 376.
[24] M.R. Ramadas, R.K. Varshney, K. Thyagarajan, A.K. Ghatak, Journal of Lightwave Technology 7 (1989) 1901.
[25] A. Kumar, R. Jindal, R.K. Varshney, R. Kashyap, Optical Fiber Technology 3 (1997) 339.

Please cite this article as: N. Ashok, et al., Dual-core leaky optical waveguide as an integrated-optic polarizer, Optics Communications (2012), http://dx.doi.org/10.1016/j.optcom.2012.07.055

Effect of welding parameters of Gas Metal Arc welding on weld bead geometry: A Review

Pushp Kumar Baghel
Research Scholar, Department of Mechanical Engineering, Delhi Technological University, Delhi, India
mech.dtu@gmail.com

Abstract: Weld quality comprises bead geometry and its microstructure, which influence the mechanical properties of the weld. This brief review illustrates the effect of pulse parameters on weld quality. The response factors, namely bead penetration, weld width, reinforcement height, weld penetration shape factor and weld reinforcement form factor as affected by arc voltage, wire feed rate, welding speed, gas flow rate and nozzle-to-plate distance has also been analysed

Keywords: GMAW; bead geometr; weld penetration;weld reinforcement; shape factor

Introduction

The gas metal arc welding (GMAW) process is generally accepted today as the preferred joining technique and is commonly chosen for welding large metal structures such as bridges, automobiles, aircraft, aerospace craft and ships owing to its joint strength, reliability, and low cost compared to other joint processes. The demand to increase productivity and quality the shortage of skilled labors and the strict health and safety requirements have lead to the development of the automated and/ or robotic welding process to deal with many of the present problems of welded fabrication.[1]

To have better knowledge and control of GMAW process it is essential to establish the relationship between process parameters and weld bead geometry to predict and control weld bead quality. The fractional factorial technique may be used for establishing quantitative relationships between welding process parameter and weld bead dimensions. [1-5]

Kim I.S. , Basu A. and Siores E. [6] studied the effects of welding process parameters on weld bead penetration for the gas metal arc welding (GMAW) process. Welding process parameters included wire diameter, gas flow rate, welding speed, arc current and welding voltage. The experiment results revealed that weld bead penetration increased as wire diameter, arc current and welding voltage increased.

Kim I.S., Son K.J., Yang Y.S. and Yaragada P.K.D.V. [7] developed a mathematical model for selection of process parameters and the prediction of bead geometry (weld bead width, weld bead penetration and reinforcement) in robotic GMAW. They used factorial design as a guide for optimization of process parameters. Three factors were incorporated into the factorial model: arc current, welding voltage and welding speed. The study revealed that a change of process parameters affects the bead width and bead height more strongly than penetration relatively.

Ganjigatti J.P., Pratihar D.K. and Roychoudhury A. [8] studied an input-output relationship of MIG welding process by using regression analysis based on the data collected as per full factorial design of experiments. The effects of the welding parameters and their interaction terms on different responses analyzed using statistical methods. Both linear as well as non-linear regression analyses were employed to establish the input-output relations.

Ganjigatti J.P., Pratihar D.K. and Roychoudhury A. [9] studied global versus cluster-wise regression analyses for prediction of bead geometry in MIG Weldind process. They observed that the cluster-wise regression analysis performs slightly better than the global approach in predicting weld bead geometric parameters.

Gupta V.K. and Parmar R.S. [10] developed a mathematical model by using fractional technique to predict the weld bead geometry and shape relationships for submerged arc welding of microalloyed steel in the medium thickness range of 10 to 16 mm.

Murugan N. and Gunaraj V. [11] developed a mathematical model to relate the process variables to the weld bead dimensions. Mathematical models developed for submerged arc welding (SAW) of mild steel pipes using factorial technique to predict weld bead width, weld bead penetration and reinforcement of weld bead geometry.

Palani P.K. and Murugan N. [12] developed a mathematical model for prediction of weld bead geometry in cladding by flux cored arc welding. Experiments were conducted to develop models using a three factor, five level factorial design for flux cored stainless steel wire with structural steel as base plate.

Murugan N. , Parmar R.S. and Sud S.K. [13] used response surface methodology (RSM) to establish quadratic relations between the welding process parameters and bead geometry for depositing stainless steel onto structural steel, using automated submerged arc welding.

Murugan N. and Parmar R.S. [14] used response surface methodology (RSM) to establish quadratic relations between the welding process parameters and bead geometry for depositing stainless steel onto structural steel, using gas metal arc welding process.

.

Shahi A.S. and Pandey Sunil [15] studied the effects of welding conditions on dilution of stainless steel cladding produced by gas metal arc welding. Experimental study carried out to analyse the effects of various gas metal arc welding (GMAW) and universal gas metal arc welding (UGMAW) process parameters on dilution in single layer stainless steel cladding of low carbon structural steel plates. This comparative study reveals that dilution obtained in UGMAW process is significantly lower than that with GMAW process due to external preheating of the filler wire in UGMAW process which results in greater contribution of arc energy by resistive heating and as a consequence, significant drop in main welding current values in case of UGMAW process was observed.

Yang L.J., Chandel R.S. and Bibby M.J. [16] derived curvilinear (i.e., non-linear) regression equation to study the relationship between correlation coefficients and standard deviation of error in prediction in submerged arc welding. They also carried out a study to check the feasibility of using linear regression instead of curvilinear regression to model weld features.

Kim I.S., Son K.J., Yang Y.S. and Yaragada P.K.D.V. [17] carried out a sensitivity analysis for a robotic gas metal arc welding process, to determine the effect of measurement errors on the uncertainty in estimated parameters. They used non-linear multiple regression analysis for modeling the process and identified the respective effects of process parameters on the weld bead geometric parameters.

Kim I.S, Son J.S. , Kim I.G. , Kim J.Y. and Kim O.S. [18] determined both linear as well as non-linear multiple regression equations to relate the welding process parameters with the weld bead geometric parameters in robotic CO2 arc welding. The developed response equations were able to predict the weld bead geometry with sufficient accuracy from the process parameters.

Lee J.I. and Rhee S. [19] conducted an investigation of the gas metal arc welding (i.e., butt welding with groove gap) as well as backward (i.e., from response to process parameters) relations were determined through multiple regression analysis and the mean deviation in prediction were seen to lie within 9.5% and 6.5% respectively.

Iordachescu Danut, Quintino Luisa Miranda Rosa and Pimenta Gervasio [20] studied influence of shielding gases and process parameters on metal transfer and bead shape in MIG brazed joints of the thin zinc coated steel plates. They observed that gases influences on the convexity of the bead, its colour, brightness, smoothness and surface pores formation.

Sung B.S. , Kim I.S. , Xue Y., Kim H.H and Cha Y.H. [21] studied Fuzzy regression model to predict the bead geometry in the robotic welding process. They developed Fuzzy linear regression model to study relationships between four process variables (wire diameter, arc voltage, welding speed and welding current) and four quality characteristics (bead width, bead height, bead penetration and bead area) and to control the quality of the GMAW based on the analysis of the bead geometry. The developed model can fit process variables to a linear function which determines not only coefficients of the fitting function for the weld quality, but also process variables for obtaining the optimal bead geometry.

Kanti K. Manikya and Rao P. Srinivasa [22] developed a back propagation neutral network model for the prediction of weld bead geometry in pulsed gas metal arc welding process. The back propagation network (BPN) system is one of the families of artificial neural network techniques used to determine welding

parameters for various arc welding processes. They observed that the results obtained from neural network model are accurate in predicting the weld bead geometry.

Balasubramanian M. , Jayabalan V. , and Balasubramanian V. [23] developed a mathematical model to predict grain size and hardness of argon tungsten pulse current arc welded titanium alloy. They considered input variables peak current, base current, frequency, time and responses as grain size and hardness. They observed that model can be effectively used to predict the hardness and grain size within the range of parameters.

Kolhe P. Kishor and Dutta C.K. [24] studied on the microstructure, phase analysis and mechanical properties, HAZ width of submerged arc weld metal multipass joint and HAZ of 16 mm thick mild steel plate. The bulk hardness, impact energy and microhardness of a multipass welded joint were tested by Rockwell hardness testing machine, Charpy V notch test and Vickers microhardness test. The various sub-zones in the microstructure was observed in the HAZ of Submerged arc weld were spheroidized, partially transformed, grain refined and grain coarsened. The variation in hardness of weld metal, fractured surface and base metal were compared with the microstructure.

Juan Wang and Yajiang LI [25] studied microstructure characterization in weld metals of high strength steels. Microstructural characterization of the weld metals of high strength steels welded under 80% Ar+20%CO2 GMAW and different weld heat inputs was carried out by means of scanning electron microscopy (SEM) and transmission electron microscopy (TEM). The relative contents of acicular ferrite and pro eutectic in the weld metals were evaluated by means of micro image analyzer. They observed that there was acicular ferrite in the grain and some pro eutectic ferrite on the boundary of original austenite grains when the weld heat input was small (9.6 kj/cm), but the microstructure was ferrite side plate when the heat input was larger (22.3 kj/cm).

Geo Ming, Zeng Xiaoyan, Yan Jun and Hu Qianwu [26] studied microstructure characteristics of lager-MIG hybrid welded mild steel. The microstructure of arc zone consists of coarse columnar dendrite and fine acicular dendrite between the columnar dendrites, but that of laser zone was composed of fine equiaxed dendrite in weld center and columnar dendrite around the equiaxed dendrite.

Eroglu M. , Aksoy M. and Orhan M. [27] studied the effect of coarse initial grain size on microstructure and mechanical properties of weld metal and HAZ of low carbon steel. The specimens taken from the hot rolled steel plate were heat treated at 11000C for 450 minute and cooled in a furnace in order to obtain a coarse initial grain size. Then original and grain coarsened were welded using a Submerged arc welding machine with heat input of 0.5,1 and 2 kj/mm. they found from toughness tests that the welded metals of coarse initial grain sized specimens and original specimens exhibited nearly same toughness values with the same heat input, whereas different HAZ toughness values were obtained with the same heat input. Maximum toughness of HAZ of the coarse initial grain sized specimen was achieved with a high input. They obtained also that as the heat input increased, maximum hardness values in the both weld metal and HAZ decreased

Conclusion: Gas Tungsten Arc welding is probably an important joining technique for different metals and alloys and can promote their wider uses in aerospace, aircraft, automotive, electronics and other industries. This review explores the effect of some influencial parameters of GTAW process which effects tensile strength and bead geometry of the joint being fabricated by this process. Review reveals the significance of selection of appropriate parameters in GMAW is highly critical

REFERENCES

[1] Kim I.S. , Basu A. and Siores E. , 1996 “Mathematical models for control of weld bead penetration in GMAW process”. International journal of advanced manufacturing technology, vol 12, pp 393-401

[2] Kim I.S., Son K.J., Yang Y.S. and Yaragada P.K.D.V., 2003 “Sensitivity analysis for process parameters in GMAW processes using a factorial design method”. International journal of machine tools and manufacture, vol 43, pp 763-769

[3] Ganjigatti J.P., Pratihar D.K. and Roychoudhury A. , 2008 “Modeling of MIG welding process using statistical approaches”. International journal of advanced manufacturing technology, vol 35, pp 1166-1190

[4] Ganjigatti J.P., Pratihar D.K. and Roychoudhury A., 2007 “Global versus cluster-wise regression analyses for prediction of bead geometry in MIG Welding process”. Journal of materials processing technology, vol 189, pp 352-366

[5] Gupta V.K. and Parmar R.S. , 1989 “Fractional factorial technique to predict dimensions of the weld bead in automatic submerged arc welding”. IE(I) Journal-MC, vol 70, pp 67-75

[6] Kim I.S. , Basu A. and Siores E. , 1996 “Mathematical models for control of weld bead penetration in GMAW process”. International journal of advanced manufacturing technology, vol 12, pp 393-401

[7] Kim I.S. , Son K.J. , Yang Y.S. and Yaragada P.K.D.V. , 2003 “Sensitivity analysis for process parameters in GMAW processes using a factorial design method”. International journal of machine tools and manufacture, vol 43, pp 763-769

[8] Ganjigatti J.P., Pratihar D.K. and Roychoudhury A., 2008 “Modeling of MIG welding process using statistical approaches”. International journal of advanced manufacturing technology, vol 35, pp 1166-1190

[9] Ganjigatti J.P., Pratihar D.K. and Roychoudhury A. , 2007 “Global versus cluster-wise regression analyses for prediction of bead geometry in MIG Welding process”. Journal of materials processing technology, vol 189, pp 352-366
[10] Gupta V.K. and Parmar R.S. , 1989 “Fractional factorial technique to predict dimensions of the weld bead in automatic submerged arc welding”. IE(I) Journal-MC, vol 70, pp 67-75
[11] Murugan N. and Gunaraj V. , 2005 “Prediction and control of weld bead geometry and shape relationships in Submerged arc welding of pipes”. Journal of materials processing technology, vol 168, pp 478-487
[12] Palani P.K. and Murugan N. , 2006 “Development of mathematical models for prediction of weld bead geometry in cladding by flux cored arc welding”. International journal advanced technology, vol 30, pp 669-679
[13] Murugan N. , Parmar R.S. and Sud S.K. , 1999, “Effect of submerged arc process variables on dilution and bead geometry in single wire surfacing”. Journal of material processing technology, vol 37, pp 767-780
[14] Murugan N. and Parmar R.S. , 1994, “Effect of MIG process parameters on the geometry of the bead in the automatic surfacing of stainless steel”. Journal of material processing technology, vol 41, pp 381-398
[15] Shahi A.S. and Pandey Sunil, 2008 “Modelling of the effects of welding conditions on dilution of stainless steel claddings produced by gas metal arc welding procedures”. International journal advanced technology, vol 30, pp 669-676
[16] Yang L.J. , Chandel R.S. and Bibby M.J. , 1993 “An analysis of curvilinear regression equations for modeling the submerged arc welding process”. Journal of material processing technology, vol 37, pp 601-611
[17] Kim I.S. , Son K.J. , Yang Y.S. and Yaragada P.K.D.V. , 2003 “Sensitivity analysis for process parameters influencing weld quality in robotic GMAW process”. Journal of materials processing technology, vol 140, pp 676-681
[18] Kim I.S, Son J.S. , Kim I.G. , Kim J.Y. and Kim O.S. , 2003, “ A study on relationship between process variables and bead penetration for robotic CO_2 arc welding”. Journal of material processing technology, vol 136, pp 139-145
[19] Lee J.I. and Rhee S. , 2000, “Prediction of process parameters for gas metal arc welding by multiple regression analysis, in; Proceeding of the Institution of mechanical engineers, Part B, vol 214 pp 443-449
[20] Iordachescu Danut, Quintino Luisa Miranda Rosa and Pimenta Gervasio , 2006 “Influence of shielding gases and process parameters on metal transfer and bead shape in MIG brazed joints of the thin zinc coated steel plates”. Materials and Design, vol 27, pp 381-390
[21] Sung B.S. , Kim I.S. , Xue Y., Kim H.H and Cha Y.H. , 2007 “Fuzzy regression model to predict the bead geometry in the robotic welding process”. Acta Metallugica Sinica, vol 20, pp 391-397
[22] Kanti K. Manikya and Rao P. Srinivasa, 2008 “Prediction of bead geometry in pulsed GMAW using back propagation neural network”. Journal of materials processing technology, vol 200, pp 300-305
[23] Balasubramanian M. , Jayabalan V. , and Balasubramanian V. , 2008 “Developing mathematical models to predict grain size and hardness of argon tungsten pulse current arc welded titanium alloy”. Journal of material processing technology, vol 196, pp 222-229
[24] Kolhe P. Kishor and Dutta C.K. , 2008 “Prediction of microstructure and mechanical properties of multipass SAW”. Journal of materials processing technology, vol 197, pp 241-249
[25] Juan Wang and Yajiang LI , 2003 “Microstructure characterization in the weld metals of high strength steels”. Journal material science, vol 26, pp 295-299
[26] Geo Ming, Zeng Xiaoyan, Yan Jun and Hu Qianwu , 2008 “Microstructure characteristics of laser-MIG hybrid welded mild steel”. Journal applied surface science, vol 72, pp 211-218
[27] Eroglu M. , Aksoy M. and Orhan M. , 1999 “Effect of coarse initial grain size on microstructure and mechanical properties of weld metal and HAZ of a low carbon steel”. Material science and engineering, vol 269, pp 59-66

Plasmonics
DOI 10.1007/s11468-012-9421-x

Highly Sensitive Plasmonic Temperature Sensor Based on Photonic Crystal Surface Plasmon Waveguide

Triranjita Srivastava · Ritwick Das · Rajan Jha

Received: 19 April 2012 / Accepted: 6 August 2012

Abstract We propose a highly sensitive temperature sensor based on photonic crystal surface plasmon waveguides comprising different plasmonic active metals such as gold, silver, and aluminum, utilizing surface plasmon resonance phenomenon. We found that the resonance wavelength can be easily and substantially tuned over a broad spectral range by changing the temperature and also by judiciously choosing the different plasmonic metals. Employing coupled mode theory, we found that the proposed sensor can be used in harsh environment with sensitivity as high as ~70 pm/K around telecommunication window.

Keywords Surface plasmon resonance · Temperature sensor · Photonic crystal waveguide

Introduction

Sensing in harsh environments such as in the case of high-power lasers, fire alarm systems, monitoring of furnace operation, or volcanic events requires reliable sensors at high temperature. The temperature sensors utilizing the optical techniques offer a promising direction in the development of sensor technologies due to several advantages as compared to other temperature measurement techniques, e.g., high sensitivity, large temperature range, stability, and immunity of optical signal to environmental disturbances [1]. The well-developed dielectric waveguides [2] and fiber-optic-based temperature sensors [1, 3–6] constitute major category of the optical temperature sensors. A multimode curved waveguide-based temperature sensor proposed by Remouche et al. has been shown to measure up to 200 °C [2], thereby imposing a limit on the temperature measurement by planar waveguide-based temperature sensors. On the other hand, the fiber optic temperature sensors based on the principles of fiber Bragg gratings [1, 3]; surface plasmon resonance (SPR) [4] or modal interferometer [5, 6] are very favorable for constructing remote-distributed sensing networks. However, the fabrication of gratings is complex since it involves tailoring the glass composition either by inscribing the gratings with femtosecond lasers or specific thermal/annealing processes. Moreover, the modal interferometer-based temperature sensor is simple to fabricate but the interference pattern drift in real time making it vulnerable to environment. Also, all these fiber-optic temperature sensors are relatively bulky and problematic to scale down and thus can hardly be used as chip-scale temperature sensors. At the same time, recent developments in nanophotonics opened new perspectives for further miniaturization of photonic components based on surface plasmon polariton waveguides [7, 8]. However, the excitation of the surface plasmon polariton (SPP) is a challenging issue. The conventional route to excite SPP is Kretschmann configuration which deploys various types of high-index prisms for achieving the resonance/phase-matching between the incident wave and SPP, thereby making them unsuitable for miniaturized applications. Another approach to excite SPPs at a thin

T. Srivastava
Department of Applied Physics, Delhi Technological University, Delhi 110042, India

R. Das
School of Physical Sciences,
National Institute of Science Education and Research,
Bhubaneswar, India

R. Jha (✉)
School of Basic Sciences,
Indian Institute of Technology Bhubaneswar,
Toshali Plaza,
Bhubaneswar, India
e-mail: rajaniitd@gmail.com

R. Jha
e-mail: rjhaphy@iitbbs.ac.in

Published online: 16 August 2012

metal stripe comprises dielectric waveguide coupling mechanism [7]. Though miniaturization is achieved by this approach, the phase-matching/resonance between conventional dielectric waveguide mode and pure SPP mode still remains a challenge, owing to high refractive index of waveguide core (~1.45) at infrared frequencies [9]. This can be overcome by exploiting the features of photonic bandgap-guided modes in photonic crystal waveguides (PCW), as the mode effective indices of guided modes are smaller than the core indices. Riding on some of these advantages, in the present paper, we propose a waveguide sensor hereafter referred to as photonic crystal surface plasmon waveguide (PCSPW; see Fig. 1) based on SPR for high and accurate temperature measurement up to 800 K. We have optimized different waveguide parameters such as, number of bilayers, refractive index profile, and channel width to design a highly sensitive temperature sensor for harsh environment. Also, a comparative study among the sensor performance has been presented for the proposed sensor comprising of different plasmonic active metals, such as gold (Au), silver (Ag), and aluminum (Al).

Design Consideration of the Proposed Sensor

The proposed experimental setup to realize the high-temperature sensor is shown in Fig. 1. Employing suitable optical equipments, e.g., polarizer, microscope, etc., light from a highly stable tunable laser is launched into the core of a PCSPW which comprised silica-PCW followed by a plasmonic waveguide consisting of a 50-nm-thick metal layer and analyte (n_a) in a vertical-stack configuration. The PCW is based on widely used lithographic and nanofabrication compatible materials like SiO_2 and TiO_2, such that silica core of refractive index n_c and thickness d_c is symmetrically sandwiched between periodically stratified cladding ($n_1 \equiv TiO_2$ and $n_2 \equiv SiO_2$, of thicknesses d_1 and d_2, respectively). The transmitted light is measured using appropriate detector and it would exhibit a minimum in the transmission spectrum at a particular wavelength, known as resonance wavelength (λ_{res}), implying resonant excitation of SPP wave at metal/analyte interface. Physically, at λ_{res}, the effective index of SPP mode ($n_{SPP} = \sqrt{\varepsilon_m \varepsilon_a / (\varepsilon_m + \varepsilon_a)}$; $\varepsilon_m = -\varepsilon_{mr} + i\varepsilon_{mi}$ and ε_a are the dielectric constants of metal and analyte, respectively) is equal to the mode-effective index of the fundamental PCW mode and hence, the power which is launched into PCW core gets transferred to the metal-analyte interface. It is well known that the resonance condition is extremely sensitive to any change in the physically measurable parameter such as refractive index, concentration, or temperature of the ambient region. The shift in the resonance wavelength (λ_{res}) and full width at half minima (FWHM) of the transmittance curve determines the sensitivity and accuracy of the sensor. For temperature sensor, the shift in the resonance wavelength ($\Delta\lambda_{res}$) due to temperature change (ΔT) determines the sensitivity (= $\Delta\lambda_{res}/\Delta T$) of temperature sensor. Further, how accurately the value of λ_{res} can be determined depends on the FWHM of the transmittance curve. Therefore, the overall performance of a sensor is defined in terms of figure of merit (FOM), which is the ratio of sensitivity and FWHM. In short, to realize a high performance sensor, the

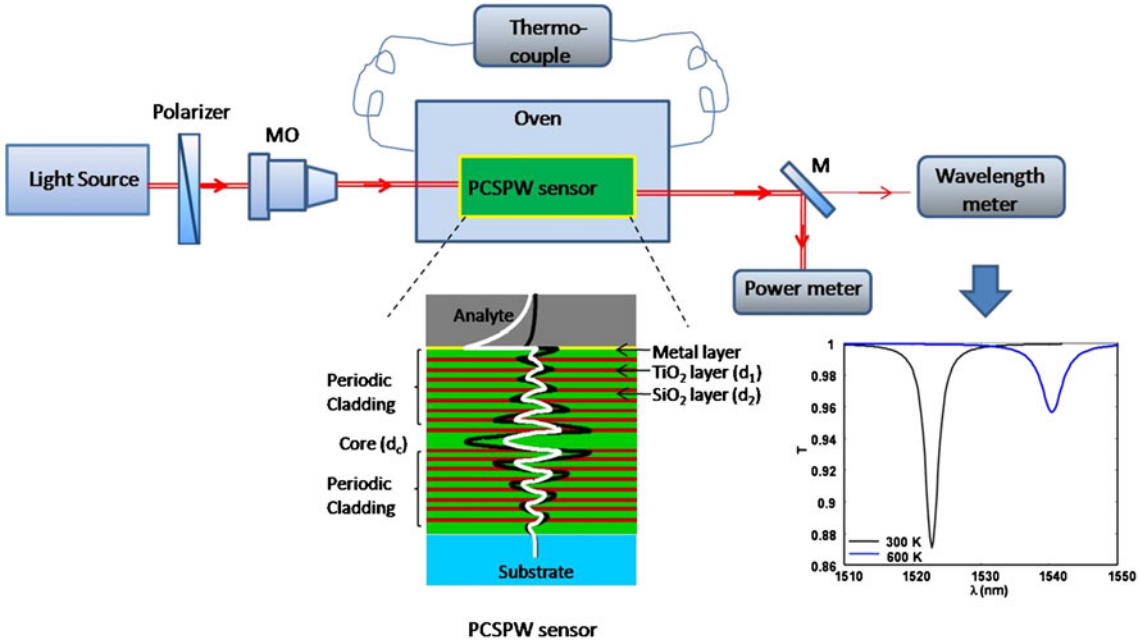


Fig. 1 The experimental setup for measuring transmittance through the proposed sensor (*MO* microscope objective, *M* beam splitter)

sensitivity, i.e., value of $\Delta\lambda_{res}$, should be as high as possible and the value of FWHM should be as low as possible.

It is important to note that in photonic crystal-based waveguides and devices, the waveguidance is achieved by the bandgap guidance mechanism where the waveguide parameters, such as core (d_c) and cladding thicknesses (d_1 and d_2), should be such that PCW mode falls within the photonic bandgap of the periodic cladding region [10]. Thus, to obtain the optimum thicknesses of different layers, we first choose an arbitrary resonance wavelength (λ_{res}) at which coupling between bandgap-guided PCW mode and SPP mode is to be realized [11, 12]. Assuming the silica-PCW- and Au/water-based plasmonic waveguide to be non-interacting, we calculate the mode effective index of SPP mode ($n_{eff(SPP)}$) at λ_{res}, which is equal to that of PCW mode ($n_{eff(PCW)}$). Since at λ_{res}, $n_{(eff)PCW}=n_{(eff)SPP}$, we obtain the core thickness (d_c) and the cladding thicknesses (d_1 and d_2) using Eq. (31) and Eq. (24), respectively of ref. [10]. Following this recipe, we first calculate the waveguide parameters at the room temperature. We choose an arbitrary resonance wavelength $\lambda_{res}=1.5$ μm and evaluated the core and cladding thicknesses to be $d_c=1.352$ μm, $d_1=0.1875$ μm, and $d_2=0.6760$ μm. It is to be mentioned that as the temperature and wavelength change, the refractive indices of all the constituents of the waveguide significantly change and hence, in the paper, we have considered the temperature and wavelength dependence of the refractive indices for all the waveguide constituents as given below.

Temperature- and Wavelength-Dependent Refractive Index of TiO_2

It is known that TiO_2 is an anisotropic material and the temperature-dependent ordinary refractive index is given by the following empirical relation [13]:

$$n_1(\lambda,T)=n(\lambda)+\frac{dn}{dT}(T-T_0) \tag{1}$$

where $n(\lambda)=\sqrt{5.913+\frac{0.2441}{\lambda^2-0.0803}}$, $\frac{dn}{dT}=1.2\times10^{-7}T-6.5\times10^{-5}$, T and T_0 (room temperature) are in degree Celsius.

Temperature- and Wavelength-Dependent Refractive Index of SiO_2

The core and one of the layer in the periodic cladding is composed of SiO_2 ($n_c=n_2$). The temperature- and wavelength-dependent refractive index of SiO_2 is given by the following Sellemeier relation [14]:

$$n_2(\lambda,T)=\sqrt{A+\frac{B}{(1-C/\lambda^2)}+\frac{D}{(1-E/\lambda^2)}} \tag{2}$$

where,

$$A=0.69\times10^{-5}T+1.31552,\ B=2.35835\times10^{-5}T+0.788404,\ C=5.84758\times10^{-7}T+0.0110199,\ D=5.48368\times10^{-7}T+0.91316,\ E=100$$

λ is in micrometers and T is in degree Celsius.

Temperature- and Wavelength-Dependent Dielectric Constant of Gold

The temperature- and wavelength-dependent dielectric constant of metal is given by the following Drude model [15]:

$$\varepsilon_m(\omega,T)=-\varepsilon_{mr}+i\varepsilon_{mi}=1-\frac{\lambda^2}{\lambda_p^2(T)\left(1+i\frac{\lambda}{\lambda_c(T)}\right)} \tag{3}$$

where λ_p and λ_c are temperature-dependent plasma wavelength and collision wavelength, respectively. As the density of electrons and the effective mass of electrons are temperature-dependent, the plasma wavelength is expressed as [16]:

$$\frac{1}{\lambda_p(T)}=\frac{1}{\lambda_{p0}(T)}\left[1+3\gamma_m(T-T_0)\right]^{-1/2} \tag{4}$$

where λ_{p0} is plasma frequency of metal at room temperature, γ_m is the thermal linear expansion coefficient of the metal, and T_0 is the room temperature in Kelvin.

The temperature dependence of collision wavelength is due to phonon–electron scattering and electron–electron scattering and is given by [17, 18],

$$\frac{1}{\lambda_c(T)}=\frac{2\pi c\varepsilon_0}{\lambda_p^2\sigma(0)}\frac{\left\{\frac{1}{10}+\left(\frac{T}{T_D}\right)^5\int_0^{\left(\frac{T_D}{T}\right)}\frac{z^4}{e^z-1}dz\right\}}{\int_0^1\frac{z^5}{(e^z-1)(1-e^{-z})}dz}+\frac{\pi^3\Gamma\Delta}{12chE_f}\left\{(k_BT)^2+\left(\frac{h\omega}{4\pi^2}\right)^2\right\} \tag{5}$$

where ε_0 is electric permittivity of free space, $1/\sigma(0)$ is dc resistivity, T_D is the Debye temperature, Γ is a constant giving the average over the Fermi surface of the scattering probability, Δ represents the fractional unklapp scattering, E_f is the Fermi energy, and the other symbols have their usual meaning. The values for all the above-mentioned parameters used in Eqs. (3), (4), and (5) to determine the temperature- and wavelength-dependent dielectric constant for different

plasmonic active metals such as Au, Ag, and Al are given in Table 1.

Temperature- and Wavelength-Dependent Refractive Index of Water and Steam

The wavelength-, temperature-, and density-dependent refractive index of water and steam according to Lorentz–Lorentz relation derived by Scheibener et al. [19] is given in

$$n_a(\lambda, T) = \sqrt{\frac{2Q+1}{1-Q}} \tag{6}$$

where

$$Q = \rho^* \left\{ a_0 + a_1\rho^* + a_2 T^* + a_3\lambda^{*2}T^* + \frac{a_4}{\lambda^{*2}} + \frac{a_5}{\left(\lambda^{*2} - \lambda_{\mathrm{UV}}^2\right)} + \frac{a_6}{\left(\lambda^{*2} - \lambda_{\mathrm{IR}}^2\right)} + a_7\rho^{*2} \right\}$$

$$\rho^* = \rho/\rho_0,\ T^* = T/T_0,\ \lambda^* = \lambda/\lambda_0$$
$$\rho_0 = 1000\mathrm{kg/m}^3\ T_0 = 273.15K,\ \lambda_0 = 0.589\mu m$$
$$a_0 = 0.243905091,\ a_1 = 9.53518094 \times 10^{-3},$$
$$a_2 = -3.64358110 \times 10^{-3}, a_3 = 2.65666426 \times 10^{-4},$$
$$a_4 = 1.59189325 \times 10^{-3},\ a_5 = 2.45733798 \times 10^{-3},$$
$$a_6 = 0.897478251,\ a_7 = -1.63066187 \times 10^{-2},$$
$$\lambda_{UV} = 0.2292020,\ \lambda_{IR} = 5.432937$$

ρ is density of water (in kilograms for cubic meter), T is temperature (in Kelvin), and λ is wavelength (in micrometers). For our calculation, we have assumed that the thickness of different metals and dielectric layers exhibits negligible variation as temperature changes.

Results and Discussion

It is known that with increase in temperature, the refractive index of all the constituents of the waveguides significantly changes, and hence the modal characteristics and the sensing performance of the proposed structure will accordingly change. The response of the PCSPW with respect to temperature is discussed below.

In order to study the modal characteristics and performance of the proposed temperature sensor, we considered eight unit cells in periodic cladding along with substrate refractive index n_s=1 and metal thickness 50 nm. Employing standard transfer matrix method [12], we obtained the mode effective indices of the fundamental mode of individual PCW. Figure 2 illustrates the variation of the mode effective indices (n_{eff}) of the fundamental mode of individual PCW and the SPP mode of plasmonic waveguide composed of Au (Au/analyte interface) with respect to wavelength for three different temperatures T=300, 500, and 700 K. It is observed that the dispersion slope for fundamental PCW mode is steeper as compared to that of the SPP mode. Also, the n_{eff} of the fundamental PCW mode and SPP mode decreases with wavelength at different temperatures. It is interesting to observe that at higher temperatures, say at 700 K (dashed magenta curve), the n_{eff} values of the fundamental PCW mode are larger than the n_{eff} values at lower temperature, say 300 K (dashed brown curve). However, a complimentary behavior is seen for the SPP modes as shown in Fig. 2. This can be explained in terms

Table 1 Parameters used for determination of wavelength- and temperature-dependent dielectric constant of various metals

Parameters	Gold (Au)	Silver (Ag)	Aluminum (Al)
λ_{p0} (m)	1.6826×10^{-7}	1.4541×10^{-7}	1.0657×10^{-7}
γ_m (K^{-1})	1.42×10^{-5}	1.89×10^{-5}	2.40×10^{-5}
Γ	0.55	0.55	0.52
$1/\sigma(0)$	1.32×10^{-6}	1.16×10^{-6}	3.90×10^{-6}
T_D (K)	170	215	396
Δ	0.77	0.73	0.75
E_f (eV)	5.53	5.48	11.7

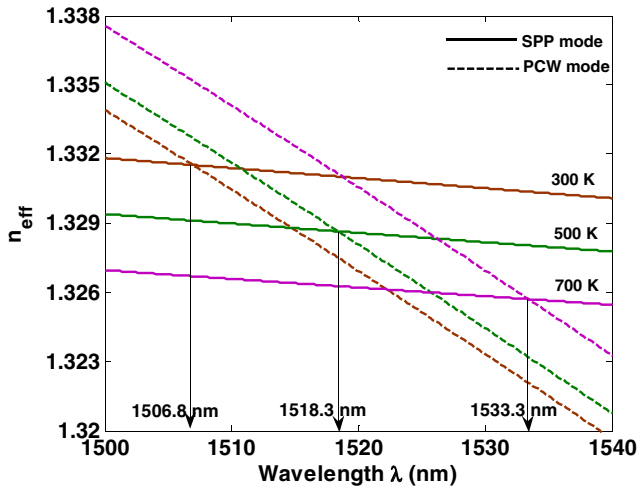


Fig. 2 Variation of mode effective index (n_{eff}) as a function of wavelength (λ), for fundamental mode of PCW and SPP mode of plasmonic waveguide composed of Au/water at three temperatures (T=300, 500, and 700 K)

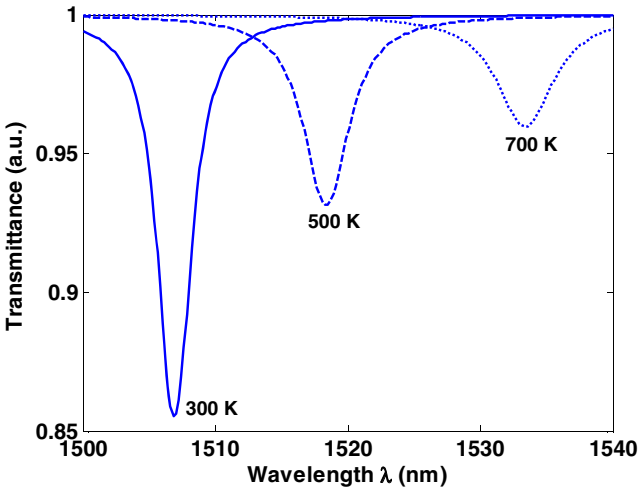


Fig. 3 Transmittance through core of PCW as a function of wavelength (λ) for the proposed PCSPW composed of Au, at T=300, 500, and 700 K

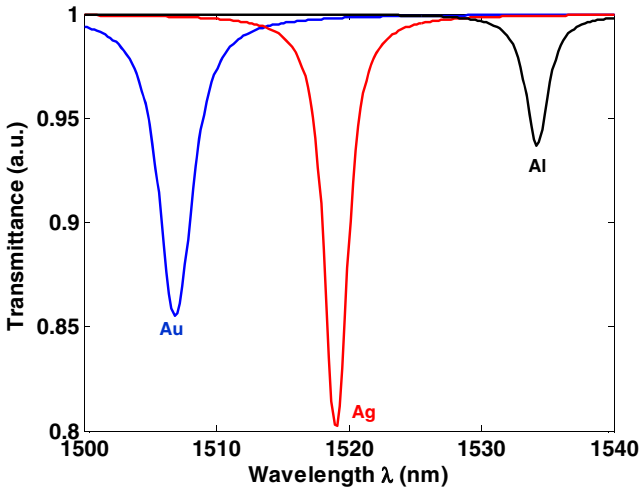


Fig. 5 Transmittance through core of PCW as a function of wavelength (λ) for the proposed PCSPW composed of Au, Ag, and Al, at T=300 K

of thermo-optic coefficient. The refractive index of silica-based core (n_c) of the PCW exhibits a positive thermo-optic coefficient, whereas the other sensor constituent materials such as TiO_2, water, and Au have negative thermo-optic coefficient. Thus with increasing temperature, the refractive index of core of the PCW increases and this results in increased mode effective index (n_{PCW}) of the fundamental PCW mode. Whereas, the effective index of SPP mode ($n_{eff(SPP)} = \sqrt{\varepsilon_m \varepsilon_a / (\varepsilon_m + \varepsilon_a)}$) depends only on the dielectric constant of water and Au and therefore, it decreases with increase in temperature. It is found (as shown in Fig. 2) that the dispersion curves for the PCW and SPP modes intersect at λ_{res}=1506.8, 1518.3, and 1533.3 nm at T=300 K (brown curves), 500 K (green curves), and 700 K (magenta curves),

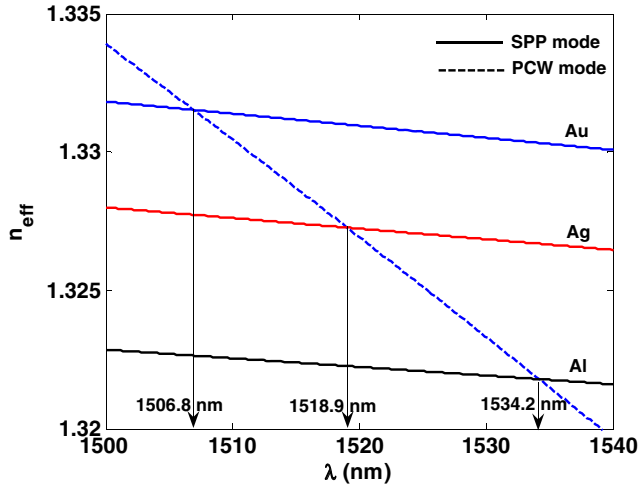


Fig. 4 Variation of mode effective index (n_{eff}) as a function of wavelength (λ), for fundamental mode of PCW and SPP mode of plasmonic waveguide composed of different metals: Au, Ag, and Al, at temperature (T=300 K)

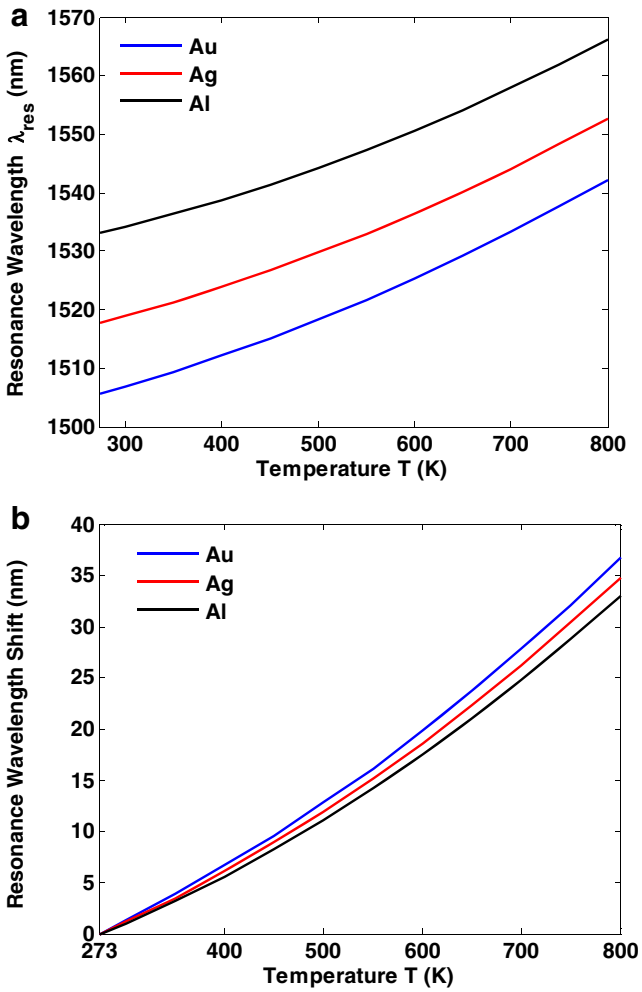


Fig. 6 Variation of **a** resonance wavelength λ_{res} and **b** shift in resonance wavelength ($\Delta\lambda_{res}$) of the sensor composed of Au, Ag, and Al as a function of temperature

respectively. The intersection wavelength corresponds to the resonance wavelength, at which the modal power launched into the core of PCW gets transferred to the Au/analyte interface [8], thereby exhibiting a dip in the transmittance curve as discussed below.

The sensor performance is obtained from the SPR transmittance curve. We employed a widely used coupled mode theory [20] to evaluate the transmittance through the proposed temperature sensor. Figure 3 illustrates the normalized transmittance through the core of PCW, as a function of wavelength for the proposed PCSPW composed of Au as SPR active metal, at T=300, 500, and 700 K for interaction length equal to coupling length at λ_{res}. It is evident that the resonance wavelength gets red shifted with increasing temperature. Also, the transmission curve is narrower (FWHM=3.0 and 5.2 nm at 300 and 700 K, respectively) at a lower temperature, which is essentially due to weaker dispersion of SPP mode at lower temperature, thereby exhibiting high contrast in the dispersion slopes of PCW mode and SPP mode (Fig. 2).

As discussed earlier, the performance of the sensor strongly depends upon the constituent metal. Figure 4 illustrates the variation of the mode effective indices (n_{eff}) of the fundamental mode of individual PCW and the SPP modes of plasmonic waveguide composed of Au, Ag, and Al with respect to wavelength at temperature T=300 K. It is found that due to the strong dispersion properties of different metals, the corresponding n_{eff} curves vary; thereby, the PCW and SPP modes intersect at λ_{res}=1506.8, 1518.3, and 1533.3 nm, for Au (blue curve), Ag (red curve), and Al (black curve), respectively, i.e., the resonance wavelength for Al is the highest. This suggests that one can easily tune the resonance wavelength (almost over 30 nm at a fixed temperature) to a significant extent by properly selecting the plasmonic metal.

Further, Fig. 5 illustrates the normalized transmittance through the core of PCW as a function of wavelength for the proposed PCSPW composed of Au, Ag, and Al at temperature T=300 K. It is found that the transmittance for Ag is the lowest, whereas for Al is the highest. Also, the FWHM for Al (2.01 nm) and Ag (2.06 nm) is comparable to each other but smaller than that of Au (3.01 nm), implying large detection accuracy for Al.

In order to quantify the sensor performance, we obtained the variation in the resonance wavelength (λ_{res}) and shift in the resonance wavelength ($\Delta\lambda_{res}$) with respect to temperature of the ambient region as illustrated in Fig. 6a, b, respectively. It is to be mentioned here that the shift in the resonance wavelength is obtained with reference to the resonance wavelength evaluated at 273 K. The figure shows that the resonance wavelength and shift in resonance wavelength both increase nonlinearly with increasing temperature. The resonance wavelength for Au-based sensor at 273 and 800 K is 1,505.5 and 1,542.2 nm, respectively. Thus, the overall shift in resonance wavelength $\Delta\lambda_{res}$=37 nm for ΔT=527 K results in sensitivity of 70 pm/K as shown in Fig. 6a. As the resonance condition changes with increase in temperature, there is large shift in resonance wavelength with temperature and hence the sensitivity increases. It is also observed that the shift in resonance wavelength is slightly larger for Au, as compared to Ag and Al. The obtained sensitivity for Ag and Al is 66 and 63 pm/K, respectively, as shown in Fig. 6b. Also, the calculated FOM for Au, Ag, and Al is 0.023, 0.032, and 0.031/K, respectively, at 300 K, thereby exhibiting high FOM for Ag- and Al-based sensors.

Conclusion

The proposed PCSPW has been optimized for 800 K owing to the melting point of different SPR active element, but the proposed concept in principle may be utilized for much higher temperature measurement. Employing the coupled-mode theory, we have found that the gold based sensor in our proposition would exhibit a sensitivity of 70 pm/K in harsh environmental conditions. Though one may use silver and aluminum for better detection accuracy along with relatively high sensitivity, but such metals would oxidize in the ambient environment at high temperature and humid conditions. Thus, our findings show that the proposed sensor can be widely deployed in petrochemical industry hazard control, aircraft industry, space security and control, air conditioning control, fire detection, and any harsh environment by riding on the advantage of latest fabrication technology for designing such chip-scale temperature sensor.

References

1. Lee B (2003) Review of the present status of optical fiber sensors. Opt Fiber Technol 9(2):57–79
2. Remouche M, Mokdad R, Chakari A, Meyrueis P (2007) Intrinsic integrated optical temperature sensor based on waveguide bend loss. Optics & Laser Technology 39:1454–1460
3. Chomát M, Čtyroký J, Berková D, Matějec V, Kaňka J, Skokánková J, Todorov F, Jančárek A, Bittner P (2006) Temperature sensitivity of long-period gratings inscribed with a CO_2 laser in optical fiber with graded-index cladding. Sensor and Actuator B Chem 119(2):642–650
4. Sharma AK, Jha R, Gupta BD (2007) Fiber-optic sensors based on surface plasmon resonance: a comprehensive review. IEEE Sens J 7(8):1118–1129
5. Choi HY, Park KS, Park SJ, Paek UC, Lee BH, Choi ES (2008) Miniature fiber-optic high temperature sensor based on a hybrid structured Fabry–Perot interferometer. Opt Lett 33:2455–2457
6. Jha R, Villatoro J, Badenes G, Pruneri V (2009) Refractometery based on photonic crystal fiber modal interferometer. Opt Lett 34:617–619
7. Homola J, Čtyroký J, Skalsky M, Hradilova J, Kolarova P (1997) A surface plasmon resonance based integrated optical sensor. Sensors Actuator B Chem 38–39:286–290

8. Srivastava T, Das R, Jha R (2010) Design consideration and propagation characteristic of channel Bragg plasmon coupled waveguide. Appl Phys Lett 97:213104–3
9. Dostalek J, Ctyroky J, Himola J, Brynda E, Skalsky M, Nekvindova P, Spirkova J, Skvor J, Schrofel J (2001) Surface plasmon resonance biosensor based on integrated optical waveguide. Sens Actuators B 76:8
10. West BR, Helmy AS (2006) Properties of the quarter-wave Bragg reflection waveguide: theory. J Opt Soc Am B 23:1207
11. Srivastava T, Jha R, Das R (2011) High performance bimetallic SPR sensor based on periodic multilayer waveguide. IEEE Photon Tech Lett 23:1448
12. Srivastava T, Das R, Jha R (2011) Highly accurate and sensitive surface plasmon resonance sensor based on channel photonic crystal waveguide. Sensors and Actuator B 157:246
13. Toyoda T, Yabe M (1983) The temperature dependence of the refractive indices of $SrTiO_3$ and TiO_2. J Phys D: Appl Phys 16: L251–L255
14. Ghosh G, Endo M, Iwasaki T (1994) Temperature dependent Sellmeier coefficients and chromatic dispersions for some optical fiber glasses. Journ Lightwave Technol 12(8):1338–1342
15. Sharma AK, Pattanaik HS, Mohr GJ (2009) On temperature sensing capability of a fibre optic SPR mechanism based on bimetallic alloy nanoparticles. J Phys D: Appl Phys 42:045104
16. Chiang HP, Leung PT, Tse WS (1998) The surface plasmon enhancement effect on absorbed molecules at elevated temperatures. J Chem Phys 108:2659–2660
17. Holstein T (1954) Optical and infrared volume absorptivity of metals. Phys Rev 96:535–536
18. Lawrence WE (1976) Electron–electron scattering in the low temperature resistivity of the noble metals. Phys Rev B 13:5316–5319
19. Schiebener P, Straub J, Levelt Sengers JMH, Gallaghen JS (1990) Refractive index of water and steam as function of wavelength, temperature and density. J Phys Chem Ref Data 19:677–717
20. Yariv A (1973) Coupled-mode theory for guided-wave optics. J Quant Electron QE-9(9):919–933

KHAGOL

No. 91 - JULY 2012

IUCAA

A quarterly bulletin of the
Inter-University Centre for
Astronomy and Astrophysics
(An Autonomous Institution of the University Grants Commission)

ISSN 0972-7647

Available online at http://www.iucaa.ernet.in/Resources.html

Editor :
Swara Ravindranath
(swara@iucaa.ernet.in)

Editorial Assistant :
Manjiri Mahabal
(mam@iucaa.ernet.in)

Contents...

IUCAA Associates Fest

The Visiting Associates Programme of IUCAA is an integral part of its efforts to spread astronomy and astrophysics in universities and colleges in the country. A number of associates visit during the summer vacation period. The team associated with the Visitor Academic Programmes made special efforts this year to increase the number of associates coming in the summer and to foster interaction between them and the faculty at IUCAA. As a part of these overall efforts, the very first Associates Fest was organized during June 25-29, this year. This gave an opportunity for the Visiting Associates and a few former associates to highlight their own work, through presentations and discussions over the period of the Fest. There were also presentations by faculty from IUCAA, NCRA and IISER, Pune on present and upcoming facilities including the instrumentation programme at IUCAA and two mega projects, Thirty Metre Telescope (TMT) India and LIGO-India in which IUCAA will be a nodal institution. The talks were delivered by Sk. Saiyad Ali, Dipankar Bhattacharya, Mamta Dahiya, Sushant Ghosh, Yashwant Gupta, S.N.A. Jaaffrey, Joe Jacob, Md. Mehedi Kalam, Ajit Kembhavi, Pushpa Khare, V.C. Kuriakose, Vijay Mohan, Sailo Mukherjee, S.K. Pandey, Sajeeth Philip, Anirudh Pradhan, A.N. Ramaprakash, C.D. Ravikumar, Saibal Ray, T.R. Seshadri, H.P. Singh,

Tarun Souradeep, R. Srianand, Kandaswamy Subramanian, Prasad Subramanian, and Pranjal Trivedi. The talks were followed by a panel discussion, in which Dr. Anil Kakodkar, Chairman, Governing Board of IUCAA participated. The discussion focused on the ways in which the associates, students and other interested persons from the university could systematically participate in the exciting new projects.

A detailed report on the discussion will be provided in due course. About 20 associates, some former associates and the IUCAA faculty members participated in the Fest. It was agreed at the end of the event that such a Fest should be organized every year and associates should be encouraged to participate in large numbers.

Welcome...

Sujoy Kumar Modak, who has joined as a Post-Doctoral Fellow. His areas of research are Gravitation, Cosmology, Black Hole Physics, and Gauge / Gravity Duality.

Angel Ruiz, who has joined as a Post-Doctoral Fellow. His areas of research are Extragalactic Astronomy, AGN, the Co-evolution of Supermassive Black Holes, and Galaxy Formation.

Introductory Summer School on Astronomy and Astrophysics (for College/University Students) and Vacation Students' Programme

The Introductory Summer School on Astronomy and Astrophysics, was conducted at IUCAA, during May 7 - June 8, 2012. Twenty-four students from across the country, who were in their final year of B.Sc., or first year of M.Sc., as well as second/third year of engineering participated in the school. The programme consisted of a series of lectures on a wide range of topics in theoretical and observational astronomy, data analysis, and problem solving sessions. Facilities like library, internet access and computing, etc. were provided to the students during their stay. The main areas on which the lectures were given include, Astrophysical Processes, Stars and Stellar Systems, The Sun, Gravitation and Cosmology, Telescopes and Instruments, and Data Analysis. Almost all the lectures were given by faculty members and visiting associates of IUCAA. Prasad Subramanian from IISER-Pune gave a couple of lectures on introduction to the Sun. The lectures were followed by hands on demonstration sessions.

The participants were taken for a visit to the IUCAA Girawali Observatory to see the 2-m optical and infrared telescope. They also visited the GMRT, operated by NCRA. The students were assigned to do a project in groups consisting of three students. The aim of these short-term projects was to give the students a feel of working with real scientific data and analysis methods. At the end of the school, the students made posters as well as oral presentations which were very well appreciated by the IUCAA faculty members.

An interactive session was arranged between IUCAA research scholars and post-doctoral fellows, and the summer school students, to give the summer school students an opportunity to ask questions related to life

as a researcher and regarding the research career in Astronomy.

The faculty coordinators for the Summer School were Durgesh Tripathi and Ajit Kembhavi.

The Vacation Students' Programme (VSP)

The Vacation Students' Programme (VSP), for students in their penultimate year of M.Sc. (Physics) or engineering degree course was held during May 7 - June 22, 2012. Exceptionally motivated final year B.Sc. and second year engineering students were also invited. This year, twelve students participated in this programme. The participants attended about 50 lectures, dealing with a wide variety of topics in Astronomy and Astrophysics, given by the academic members of IUCAA. They also did a project with one of the faculty members of IUCAA during this period. R. Srianand was the faculty coordinator of this programme.

E-Books...

The IUCAA library has recently acquired 183 e-books in physics and astronomy published by Cambridge University Press (CUP) and 78 e-books from the Oxford Scholarship Online, which can be accessed at **http://www.iucaa.ernet.in/~library/eBooks.html.** These e-books can be accessed by the users within IUCAA. At present, certain IUCAA Resource Centres can access these books and it is planned to extend the facility to all the IRCs in the near future.

School Students' Summer Programme and Astronomy Camp

IUCAA Public Outreach group conducted the regular School Students' Summer Programme as well as an Astronomy Camp from April 11 to May 20, 2012. One hundred and sixty students of class VIII/IX/X participated in this programme, who were nominated by their schools.

The highlight of the Camp was the newly developed content to give an overview of Astronomer's tools to school students, while getting them to try out what they study in school. Topics covered ranged from simple geometry and statistics to optics and spectroscopy. The students used the Samrat Yantra to study Trigonometry, as well as the Celestial Motions of the Sun. They were able to understand the scales of the solar system by finding the ratio of the distances of the Planets to the Sun. They were also introduced to the basics of aero-modelling and flight. Star maps were distributed, and sky watching practice sessions were held for all the students on Friday evenings. Throughout the duration, effort was taken to clarify misconceptions and to give a better understanding of motives of Science.

The programme was coordinated by Samir Dhurde. The experimental devices used were designed by Arvind Paranjpye (Director, Nehru Planetarium, Mumbai), who also gave four talks. They were constructed by Maharudra Mate. Considerable teaching help was provided by summer interns, Chaitanya Thakur and Sujay Mate, both first year B.Sc. students from Fergusson College, Pune, and IISER, Pune respectively. The ability of the students of that age to put data into a graphical form is being assessed, in collaboration with Amit Dhakulkar, a research student at HBCSE-TIFR, Mumbai.

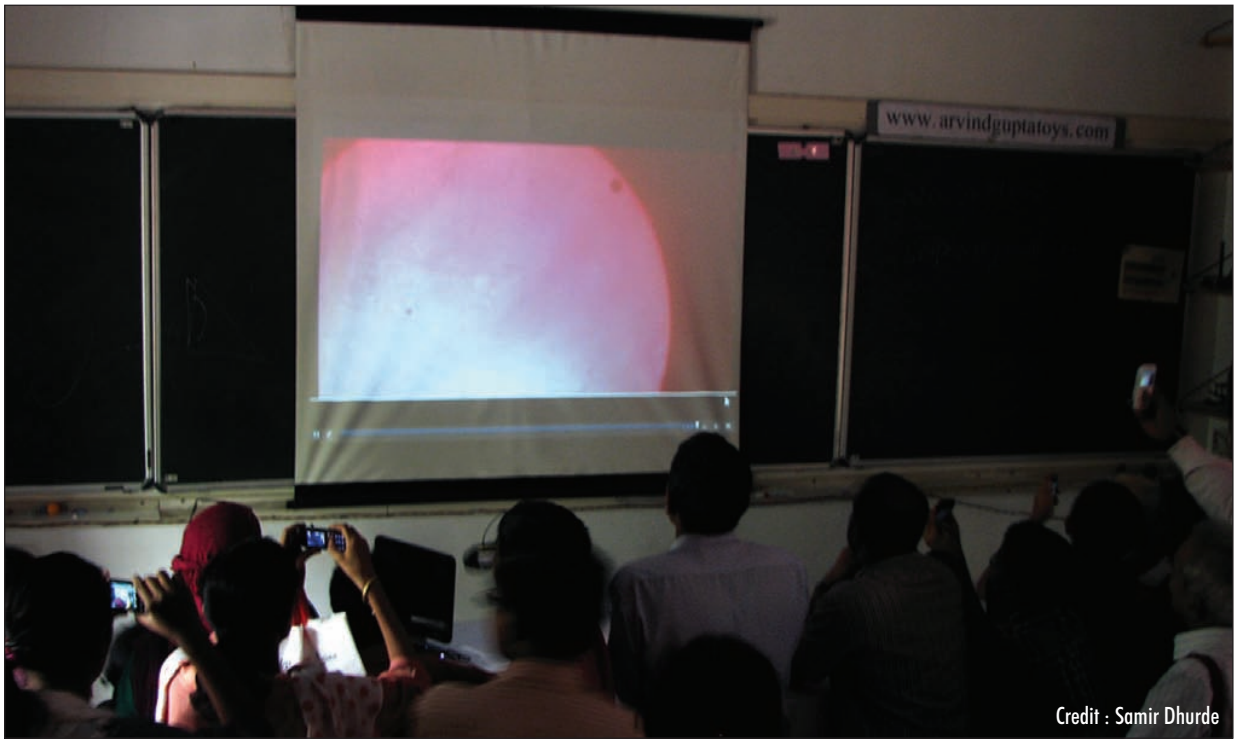
The regular, School Students' Summer Programme continued in parallel, in which about 21 students did projects with six academics from IUCAA, namely, J. V. Narlikar, Krishna Parattu, Pushpa Khare, Ranjan Gupta, Shalima P. and Suprit Singh. The students were given access to the main IUCAA library, and the facilities at the Muktangan Vidnayan Shodhika, including the library, computers and workshops. On completion of their work, on the last working day, the student teams made presentations on the work done by them during that week.

Educational Campus Visit

Over 3 weeks in May 2012, IUCAA conducted educational campus visits for ~1500 high school students from many districts of Maharashtra in collaboration with the INSPIRE programme conducted by IIIT, Pune. During the visit, the students spent time in the Science Park aided by IUCAA trained volunteers, and attended a talk discussing astronomy and its scope for Indian students.

Transit of Venus

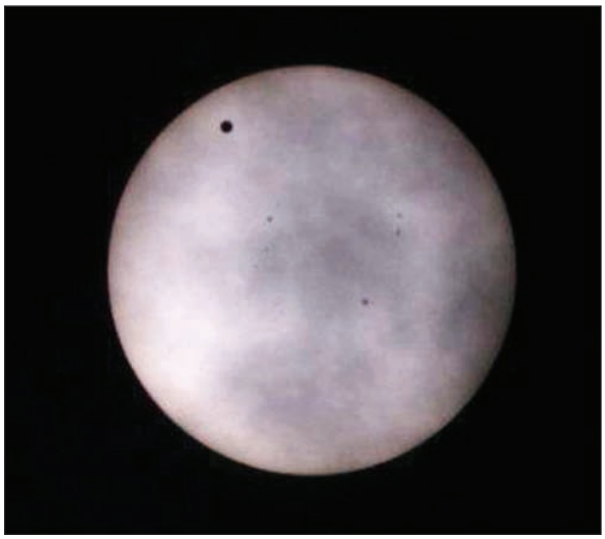
June 6, 2012



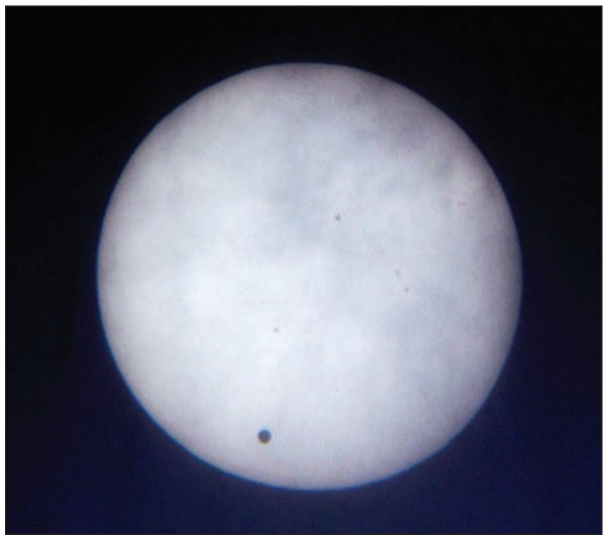
Credit : Samir Dhurde

On June 6, 2012 more than 700 people gathered at the IUCAA Science Centre to see the Transit of Venus. Despite a cloud cover, many woke up early to see the event, which had already started before the Sun rose at 6.00 a.m. The enthusiasm for this rare event that happens twice in a century separated by eight years, was not dampened by the drizzle. The clouds played spoilsport while live telecasts from other Indian cities could be seen on TV. A webcast from the Keck Observatory was shown to the public in the Science Centre.

The Public Outreach team, with the help of Akashmitra volunteers, had prepared six telescopes and many solar image projection units. Hundreds of solar eclipse goggles had also been distributed. This year's summer school students had been trained to make solar projectors to share the event with those around them. However, the clouds continued till 9:15 a.m., and when the Sun peeked out, all the patient viewers got to see live the round beauty spot on the face of the Sun. This was for the second time, after 2004, that IUCAA was successful in sharing this event with so many people. The picture shows the event, as seen from Pune by an amateur astronomer, Kshitija Kelkar, an ex-project student at IUCAA. It also shows some sunspots that were actually on the Sun's surface on that day.



Credit : Kshitija Kelkar



Credit : V.C. Kuriakose, Visiting Associate of IUCAA, and his students at CUSAT, Kochi (Using 6" IUCAA telescope)

Seminars

12.04.2012	Rishi Khatri on *Bose-Einstein condensation of CMB and damping of sound waves in the early universe*
04.05. 2012	Dipanjan Mukherjee on *MHD of accretion columns on neutron star poles and cyclotron spectra*
09.05.2012	Hadi Rahmani Bayegi on *Constraining the variation of fundamental constants at z ~ 1.3 using 21-cm absorbers*
15.05.2012	Sanved Kolekar on *Some issues in horizon thermodynamics*
17.05.2012	Suprit Singh on *Complex effective path: A semi-classical probe of quantum effects*
25.05.2012	Vikram Rana on *The nuclear spectroscopic telescope array (NuSTAR)*
29.05.2012	Bhaswati Bhattacharyya on *Discovery of interesting millisecond pulsars at the positions of Fermi LAT unassociated sources and results from followup study*
30.05.2012	Aditya Rotti on *Weak lensing probes of cosmology*
31.05.2012	P. Shalima on *Dust properties of extragalactic sources*
01.06.2012	Bibhas Majhi on *Noether current, Virasoro algebra and entropy*
04.06.2012	Sibasish Laha on *A study of warm absorbers in Seyfert galaxies in X-rays*
06.06.2012	Kaustubh Vaghmare on *The ones in the middle*
11.06.2012	Charles Jose on *Spatial clustering of high redshift LBGS*
12.06.2012	Ramesh Chandra on *Multiple surges/jets and associated flares from NOAA Ar10484 on 23 and 25 October, 2003*
12.06.2012	Badam Singh Kushvah on *Stability and chaos in dynamical systems*
12.06.2012	Debbijoy Bhattacharya on *Blazar: Long term variability and their luminosity function*
13.06.2012	Pramod Kumar Samal on *Question of isotropy of universe*
13.06.2012	Sanil Unnikrishnan on *Non-canonical scalars and their application in cosmology*
14.06.2012	Gaurav Goswami on *Uncovering cosmic inflation*
15.06.2012	Jayanti Prasad on *Optimization problems in cosmology*
18.06.2012	Prakash Sarkar on *Exploring the cosmic web in the galaxy redshift survey*
20.06.2012	Sowgat Muzahid on *QSO absorption line systems : Intervening and associated*
22.06.2012	Surajit Paul on *Effect of galaxy cluster mergers on cluster medium and void*

Announcements

Workshop on Cosmology (November 8 - 10, 2012)

Bangalore University will be hosting an IUCAA sponsored Workshop on Cosmology during November 8 - 10, 2012. The workshop is intended for research scholars and college/ university teachers. Topics proposed to be covered include large scale structures of the universe, cosmological models, early universe, primordial nucleosynthesis, microwave background radiation, dark matter and dark energy. To participate in this workshop please send your resume by email to B.A. Kagali before September 15, 2012. Participants will be provided with free hospitality and accommodation during this workshop. Limited travel support will be available. Those who require travel support should mention that in their applications.

Contact:

Professor B. A. Kagali,
Department of Physics, Bangalore University, Bangalore - 560 056.
Ph. : 080-22961475, Email : bakagali@gmail.com

BITS-IUCAA workshop on Gravitational Waves Data Analysis (December 17 - 21, 2012)

Birla Institute of Technology and Science (BITS), Pilani, K.K. Birla Goa Campus and the Inter-University Centre for Astronomy and Astrophysics (IUCAA) are jointly organising a Workshop on Gravitational Wave Data Analysis at the BITS, Pilani, K.K. Birla Goa Campus during December 17-21, 2012. Students and young researchers with strong motivation to pursue future studies in Gravitational Waves are specially encouraged to participate in this workshop.

The participants will be introduced to the basic concepts of Gravitational Waves, their sources and detection, and techniques of data analysis Interested candidates may send their CV to S. K. Sahay before September 30, 2012, along with a covering letter briefly mentioning their motivation to attend the workshop. Selected participants will be provided with free hospitality during the workshop. Limited travel support, as admissible, may be available for participants on advance request.

Contact :

Professor S. K. Sahay
Department of Computer Science and Informations Systems
BITS, Pilani - K.K. Birla Goa Campus, NH-17B, By Pass Road,
Zuarinagar, Goa - 403726.
Ph. : 0832-2580-243, Fax : +91-832-2557-033,
Email : ssahay@bits-goa.ac.in

Introductory School in Astronomy and Astrophysics (January 29 - 31, 2013)

Srikishansarda College, Hailakandi, Assam, will be hosting an IUCAA sponsored Introductory School in Astronomy and Astrophysics. This workshop is intended for young researchers (final year postgraduate students, research scholars, and college / university teachers) who are working in this area or plan to do so in future. To participate in this workshop please send your resume by email to A. K. Das before October 31, 2012. Participants will be provided with free hospitality and accommodation during this school. Limited travel support will be available. Those who require travel support should mention that in their applications.

Contact :

Dr. Ashok Das
Srikishansarda College, Hailakandi - 788151, Assam.
Ph. / Fax : 03844-222409, Email : ashok.dashok.das93@gmail.com

Visitors Expected

July 2012

A. Abdujabbarov, Nuclear Institute of Physics, Uzbekistan; Bobomurat Ahmedov, Nuclear Institute of Physics, Uzbekistan; B.G. Anandarao, Physical Research Laboratory, Ahmedabad; KG. Arun, Chennai Mathematical Institute ; Abhay Ashtekar, University of Pennsylvania, USA; Jasjeet Bagla, IISER, Mohali; Raj Bali, University of Rajasthan, Jaipur; S.K. Banerjee, University of Petroleum and Energy Studies, Dehradun; Rupal Basak, TIFR, Mumbai; David Buckley, SAAO, South Africa; Soumya Chakravarti, Presidency University, Kolkata; Phil Charles, SAAO, South Africa; Hauen Chung, Korea Institute of Advanced Study, Korea; Prathamesh Dalvi, BITS, Goa; Sudipta Das, ISI, Kolkata; Jishnu Dey, Presidency University, Kolkata; Mira Dey, Presidency University, Kolkata; P.S. Gaikwad, University of Pune; Sunandan Gangopadhyay, West Bengal State University, Kolkata; Fred Gent, University of Newcastle, United Kingdom; Sushant Ghosh, Jamia Millia Islamia, Delhi; Sarah Gibson, High Altitude Observatory, USA; Mubashir Hamid, University of Kashmir, Srinagar; K.P. Harikrishnan, The Cochin College, Kochi; Reju Sam John, Pondicherry University; Amritaksha Kar, Assam University, Silchar; Enrico Kotze, SAAO, South Africa; Marissa Kotze, SAAO, South Africa; Shailesh Kulkarni, Sogang University, South Korea; Saurabh Kumar, Hindalco Industries Ltd., Uttar Pradesh; Suresh Kumar, Delhi Technological University; Debojoti Kuzur, IISER, Kolkata; Naval Kishor Lohani, Govt. P.G. College, Uttarakhand; Bari Maqbool, University of Kashmir, Srinagar; Anand Narayanan, IISST, Thiruvananthapuram; Ravikumar R. Panchal, B.H. Gardi College of Engineering and Technology, Rajkot; P.N. Pandita, North Eastern Hill University, Shillong; A. Rajoelimanana, SAAO, South Africa; Fidy Ramamonjisoa, University of Kwazulu-Natal, South Africa; L. Resmi, IISST, Thiruvananthapuram; Anirban Saha, West Bengal State University, Kolkata; Rathin Sarma, Hojai College, Assam; Asoke Sen, Assam University, Silchar; Anand Sengupta, IIT, Gandhinagar; Kiran Shanker, University of Allahabad, Allahabad; Pankaj Sheoren, Jamia Millia Islamia, Delhi; Dipak Singh, ERNET, Delhi; Nishant Singh, Raman Research Institute, Bangalore; Srividya Subramanian, USA; Lijo Thomas, Christ University, Bangalore; Pranjal Trivedi, University of Delhi; Jithesh V., University of Calicut, Kerala; Naveel Wani, University of Kashmir, Srinagar; Mahesh Kumar Yadav, Sant Ganinath Govt. P.G. College, Uttar Pradesh; and A.A. Zdziarski, N. Copernicus Astronomical Centre, Poland.

August 2012

Gazi Ameen Ahmed, Tezpur University, Assam; Asoke Kumar Sen, Silchar, Assam University; Trina Chakraborty, ISI, Kolkata; Chanchal Chawla, Sant Longowal Institute of Engineering and Technology, Sangrur; Sayantan Choudhury, ISI, Kolkata; Sudipta Das, ISI, Kolkata; Reda El-Bendary, NRIAG, Egypt; Sushant Gupta, BBA Central University, Lucknow; Rekha Jaiswal, Hindu Post-Graduate College, Uttar Pradesh; Kanti Jotania, The M.S. University of Baroda, Vadodara; Barun Kumar Pal, ISI, Kolkata; Anirudh Pradhan, Hindu Post-Graduate College, Uttar Pradesh; Souvik Pramanik, ISI, Kolkata; Isabel V. Rosario, University of Valencia, Spain; and Christopher Tout, University of Cambridge, United Kingdom.

September 2012

Andreas Finke, University of Heidelberg, Germany; and Bhola Ishwar, BRA Bihar University, Muzaffarpur.

Long term visitors

Pushpa Khare, and Sanjeev Dhurandhar.

Visitors

April - June 2012

Marcela Morillo Acosta, Rupali Ahire, Syed Moosa Ali, Sk. Saiyad Ali, Pavan Kumar Aluri, Rizwan Ul-Haq Ansari, Bidisha Bandyopadhyay, Sudhanshu Barway, Sandip Bhattacharya, N. Bijoy, Sukanta Bose, Shuvendu Chakraborty, Rabin Chhetri, Steve Crawford, Pratik Dabhade, Mamta Dahiya, Drisya K., Madhuri Gaikwad, Ritesh Ghosh, Shaon Ghosh, Sushant Ghosh, M.K. Haris, Ananda Hota, Bala Iyer, Deepak Jain, S.N.A. Jaaffrey, Joe Jacob, Kanti Jotania, Minu Joy, V. Jithesh, Md. Mehedi Kalam, Tejas Kale, Atish Kamble, Rishi Khatri, Gopala B. Krishna, Atmjeet Kumar, V.C. Kuriakose, Rakesh Lakshman, Sruthil S.B. Lal, Michael Little, Nikunj Maheswari, Nairwita Mazumder, Sneha More, Arunava Mukherjee, Sailo Mukherjee, Rajesh Nayak, Mayukh Pahari, Sambit Kumar Panda, S.K. Pandey, Preeti Panjwani, Shankar Dayal Pathak, Arvind Paranjpye, Devraj Pawar, Pramod Pawar, Kedar Phadke, Sajeeth Philip, Alok Prabhakar, Anirudh Pradhan, C.P. Ranjith, C.D. Ravikumar, Saibal Ray, Rupak Roy, Amartya Jyoti Saha, Anuradha Samajdar, Subroto Sarkar, Bhim Prasad Sarmah, T.R. Seshadri, Marsha Shams, K. Shanthi, Arvind K. Singh, Balveer Singh, Dipal Singh, H.P. Singh, Nishant Singh, Parth Singh, Jay Singla, Radha Srinivasan, T.P. Srinivasan, Firoza Sutaria, Eric Tatulli, Navita Thakkar, Peter Tino, Pranjal Trivedi, Nilkanth Dattatray Vagshette, and M. Vivek.

Also there were about 35 students attending the Introductory Summer School on Astronomy and Astrophysics, and the Vacation Students' Programme.

IUCAA Preprints

Listed below are the IUCAA preprints released during April - June 2012. These can be obtained from the IUCAA library (library@iucaa.ernet.in). The preprints can also be freely downloaded from http://www.iucaa.ernet.in/~library/main.html.

Durgesh Tripathi, Helen E. Mason and James A. Klimchuk, *Active region moss: Doppler shifts from Hinode/EIS observations,* IUCAA-04/2012; Rakesh K. Rai and Shantanu Rastogi, *Modeling anomalous/non-CCM extinction using nanodiamonds,* IUCAA-05/2012; Naseer Iqbal, Naveel Ahmed, Mubashir Hamid and Tabasum Masood, *Correlation functions for extended mass galaxy clusters,* IUCAA-06/2012; Pranjal Trivedi, T. R. Seshadri and Kandaswamy Subramanian, *Cosmic microwave background trispectrum and primordial magnetic field limits,* IUCAA-07/2012; Anju Maurya, Shantanu Rastogi and Gael Rouille, Friedrich Huisken and Thomas Henning, *Experimental and theoretical study on the infrared spectroscopy of astrophysically relevant PAH derivatives 2- and 9-vinylanthracene,* IUCAA-08/2012.

Know Thy Clouds - 10

Arvind Paranjpye
Director, Nehru Planetarium, Mumbai.

Nimbostratus : Brings continuous rains

As we have said in the last article, Stratus[1] gets their name from Latin, which means 'layer' and nimbo is associated with Latin word "nimbus" for rain. Nimbostrati are layered clouds, which bring rains, as would be expected during the monsoon. Unlike Cumulonimbus[2], these clouds bring long continuous spell of rains. These clouds are seen across the world, but are more common at the latitudes close to the equator.

These clouds are formed when altostratus (to be discussed in the next issue) descend to lower altitudes. These clouds do not have much of vertical development but the layer can be as thick as 2000 metres. Nimbostratus is somewhat similar to Altocumulus[3] and Stratocumulus[4], but these clouds show patterns and they have well-defined base.

Nimbostrati are rather difficult clouds to photograph, as can be seen in the photographs that were taken on Mumbai-Pune highway at Lonavala ghat. These clouds have generally grey base and lack contrast. The cloud blocks the sun entirely behind it and create generally a wet, gloomy and dull rainy weather that can last for days together.

At times, the cloud breaks and one can see small blobs of clouds right below the main nimbus clouds. These are called scuds. The scuds are produced due to re-condensation of water saturated air.

Due to its featureless appearance and lack of contrast, it is difficult to estimate the speed and the direction in which the cloud is traveling. This is where scuds have been found useful.

Like scuds, another interesting feature often associated with nimbostratus is virga. It is streak or shaft of cloud that appears to hang from the cloud. Virga never reaches the ground. It evaporates well above the ground level.

Name	:	Nimbostratus
Symbol	:	Ns
Height	:	upto 2000 m
Symbol	:	

[1] **Stratus** : Know Thy Clouds - 9, April 2012, Khagol No. 90
[2] **Cumulonimbus** : Know Thy Clouds - 3,October 2011, Khagol No. 84
[3] **Altocumulus** : Know Thy Clouds - 5, April 2011, Khagol No. 86
[4] **Stratocumulus** : Know Thy Clouds - 6, July 2011, Khagol No. 87

Virga

Nimbostrati

Khagol (the Celestial Sphere) is the quarterly bulletin of IUCAA

We welcome your responses at the following address:

IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India.
Phone : (020) 25691414; 25604100 Fax : (020) 25604699
email : publ@iucaa.ernet.in Web page : http://www.iucaa.ernet.in/

Mutual Authentication Protocol Using Hyperelliptic Curve Cryptosystem in Constrained Devices

Kakali Chatterjee[1], Asok De[2], and Daya Gupta[1]
(Corresponding author:Kakali Chatterjee)

Computer Engineering Department, Delhi Technological University, Delhi, India[1]
Ambedkar Institute of Advanced Communication Technologies & Research, Delhi, India[2]
(Email: kakali2008@gmail.com)

(*Received Aug. 10, 2011; revised and accepted Feb. 22, 2012*)

Abstract

Hyperelliptic Curve Cryptosystem (HECC) is well suited for secure communication in wireless sensor network as limited resources (storage, time or power) on sensor nodes restrict the use of conventional security techniques. We can construct genus 2 HECC on 80-bit finite fields in order to achieve the same security level as 160-bit ECC or 1024-bit RSA. This paper proposes a mutual authentication protocol based on Hyperelliptic Curve Digital Signature Algorithm for a secure access in constrained devices which allows both the entities to verify each other's authenticity. Our experimental result shows that performance of the proposed system is comparable with that of ECC.

Keywords: Hyperelliptic Curve Cryptosystem (HECC), Hyperelliptic Curve Digital Signature Algorithm (HECDSA), mutual authentication protocol

1 Introduction

With the rapid development of the information technology, wireless networks are now extensively used to transmit critical information relating to monitoring of real time data. The security mechanisms are essential to ensure integrity, confidentiality and authenticity of the data. Implementation of suitable cryptosystem in this environment is challenging as these networks consist of many tiny and smart devices which are constrained in terms of memory, computing power and energy supply. While considering the different security threats involved in wireless sensor network, different symmetric /asymmetric algorithms are proposed in [6, 10, 15, 21]. In general these approaches either need pre-distributed keys which mean a higher configuration effort before deployment or they produce much traffic which results in higher energy consumption [20].

In case of symmetric algorithms memory usage is reduced but an adversary can easily eavesdrop on communication or a stealing of node can be possible. To avoid these types of attacks, a mutual authentication protocol based on ECC has been proposed [22]. The principal attraction of ECC compared to RSA is that it offers equal security for a far smaller key size, thereby reducing processing overhead. ECC based mutual authenticated key agreement protocol was already established for Wireless LAN security [2, 14]. So far, several protocols have been proposed to provide robust mutual authentication and key establishment for Wireless LAN. The significant improved performance of some of the protocols in computational and communicational load over many other key agreement protocols were compared and discussed in [3]. These protocols utilize ECDSA signature technique that enhances the security of user authentication and key exchange. However, the security level can also be increased using hyperelliptic curves because it has some advantage over ECC. For HECC over a finite field one needs 40-bits to 80-bit long operands to compute the group operations for these curves. In the case of ECC we have to work with operand lengths of approximately 160-bit whereas in the case of RSA, the operands will be approximately 1024-bit in order to achieve the same security. HECC is, therefore, more suitable for implementation in the constrained platforms in wireless networks.

Hyperelliptic Curve Cryptosystem (HECC) was proposed by Koblitz [13] in 1989 based on the discrete logarithm problem on the Jacobian of hyperelliptic curves over finite fields. The main difference between ECC and HECC is in group operation because these consist of different sequences of operations. Unlike elliptic curves, the points on the hyperelliptic curve do not form a group. The additive group on which the cryptographic primitives are implemented is the divisor class group. Each element of this group is a reduced divisor. Divisor group operations of HECC are more complex compared to point operation of ECC for implementing the cryptographic primitives. Hence it is challenging to implement HECC in constrained environment.

Different aspects of Hyperelliptic curve cryptosystem are discussed in Avanzi [1], Menezes et al. [16], Pelzl et al. [17, 18]. We have discussed Evolution of Hyperelliptic Curve Cryptosystems in [9]. Current Research on HECC emphasize on finding efficient methods to select secure hyperelliptic curves, fast operations on the Jacobians and implementation of

HECC for use in practical applications to enhance network security.

Our contributions in this paper are as follows:

(1) We propose a mutual authentication protocol based on HECDSA for a secure communication in wireless network for constrained devices. The proposed protocol is a secure authenticated protocol as it can resist the possible attacks from both internal users and external hackers as discussed in the Security Analysis.

(2) We have implemented our proposed authentication protocol over the finite field $\mathbb{F}_p$ in affine co-ordinates in Netbeans IDE 6.8. Our experimental result shows the timings of basic operations of the proposed Hyperelliptic curve based scheme which enables us to exchange keys, sign & authenticate documents and we have compared these timings to study the relative performance of HECC with that of ECC.

The rest of the paper is organized as follows: Section 2 presents Mathematical Background; Section 3 provides Proposed Mutual Authentication Protocol; Section 4 discusses Security Analysis; Section 5 presents Implementation Results; Finally, we conclude the paper in Section 6.

2 Mathematical Background

2.1 Arithmetic of Hyperelliptic Curves

Let $\mathbb{F}$ be a finite field, and let $\overline{\mathbb{F}}$ be the algebraic closure of $\mathbb{F}$ [18]. A hyperelliptic curve C of genus $g \geq 1$ over $\mathbb{F}$ is the set of solutions $(u, v) \in \mathbb{F} \times \mathbb{F}$ to the equation $C: v^2 + h(u)v = f(u)$. The polynomial $h(u) \in \mathbb{F}[u]$ is of degree at most g and $f(u) \in \mathbb{F}[u]$ is a monic polynomial of degree $2g+1$. For odd characteristic it suffices to suppose that $h(u) = 0$ and $f(u)$ is square free. If no point on the curve over the algebraic closure $\overline{\mathbb{F}}$ of $\mathbb{F}$ satisfies both partial derivatives $2v + h(u) = 0$ and $h'(u)v - f'(u) = 0$, then the curve is said to be non-singular.

The points of curve C generate a set called Jacobian. The Jacobian of the curve C is the quotient group $\mathbb{J} = \mathbb{D}^0/\mathbb{P}$, where $\mathbb{D}^0$ is the set of divisors of degree zero, and $\mathbb{P}$ is the set of divisors of rational functions. The element of Jacobian over $\mathbb{F}$, denoted by $\mathbb{J}_C(\mathbb{F})$, can be represented uniquely by a divisor $D = \sum m_i P_i$, $m_i \in \mathbb{Z}$, is a finite formal sum of $\overline{\mathbb{F}}$ points. Its degree is the sum of the coefficients $\sum m_i$. The set of all divisors form an Abelian group denoted by $\mathbb{D}(C)$. The set of degree zero divisors $\mathbb{D}^0$ forms a subgroup of $\mathbb{D}(C)$ [16].

Cantor shows that each element of the Jacobian can be represented in the form $D = \sum_{i=1}^{r} P_i - r.\infty$ such that for all $i \neq j$, P_i and P_j are not symmetric points [7]. Such a divisor is called a semi-reduced divisor. Each element of the Jacobian can be represented uniquely by such a divisor, subject to $r \leq g$. Such divisors are referred to as reduced divisors. We use the reduced divisor in addition of $\mathbb{J}_C$. Cantor's algorithm is used for doing arithmetic in general hyperelliptic curve which applies to any genus and characteristic. This transfer the group laws in a sequence of Composition and Reduction using only polynomial arithmetic. Group operations on a Jacobian are performed in two steps: addition of generic divisors and doubling of generic divisors. Addition of divisor classes means multiplication of ideal classes, which consists in a composition of the ideals and a first reduction to a basis of two polynomials. The output of this algorithm is called semi-reduced divisor. Then the second algorithm (reduction) is used to find the unique representative in the class.

The equivalence classes of the Jacobian are represented by a unique reduced divisor (which is represented using Mumford representation) upon which we perform the group law. Each unique reduced divisors can be represented via a unique pair of polynomials $u(x)$ and $v(x)$, $u,v \in \mathbb{F}_q[x]$, where

(1) u is monic

(2) $\deg v < \deg u \leq g$

(3) $u \mid v^2 + vh - f$

This is known as Mumford representation. Mumford proposed a convenient way to represent each reduced divisors as $D = (u(x), v(x))$, where $u(x)$ is a monic polynomial with $\deg(u(x)) \leq g$ and $v(x)$ (which is not monic in general) with $\deg(v(x)) < \deg(u(x))$.

When developing formulas for implementing genus g arithmetic, we are largely concerned with the frequent case that arises where both reduced divisors $D_1 = [u_1, v_1]$, $D_2 = [u_2, v_2]$ are used to find the unique reduced divisor equivalent to $D_1 + D_2$. To find the unique reduced divisor equivalent to $D_1 + D_2$, one performs two steps (Cantor's algorithm):

(1) Determine a semi-reduced divisor D representing the sum of D_1 and D_2.

(2) Transform the semi-reduced divisor D into a reduced one, i.e compute D' such that D' is the unique reduced divisor equivalent to D.

2.2 Security of Hyperelliptic Curves

In hyperelliptic curve cryptography, finding a suitable hyperelliptic curve is an important fundamental problem [16]. Security of HEC is based on the difficulty of solving the discrete logarithm problem in Jacobian of HEC. The HCDLP in $J(C; \mathbb{F}_q^n)$ is: given two divisors D_1, D_2 defined on $J(C, \mathbb{F}_q^n)$ over $\mathbb{F}_q^n$, to determine integer m such that $D_2 = mD_1$, provided such an integer m exists.

To establish a secure hyperelliptic curve, its Jacobian should satisfy the following conditions:

(1) Adleman et al. [4] found a subexponential time algorithm to solve the DL in the Jacobian of HEC of a big genus over a finite field. Curves of higher genera (preferably $g \leq 4$) are, therefore, not suitable for cryptographic use ($2g+1 < \log q^n$).

(2) If the group order is large but divisible by only small primes, the DLP can be broken by Pohlig-Hellman attack. It is claimed that this largest prime factor should be at least 160 bits in length.

(3) To prevent the attack of Frey [11] which uses Tate pairing generation of MOV attacks, the large prime factor of $J(C; \mathbb{F}_q^n)$ should not divide $(q^n)^k - 1$, here $k < (\log q^n)^2$.

(4) To prevent the attack generated by Ruck [19], the Jacobian of a hyperelliptic curve over the large prime field should not have p-order subgroup.

(5) To harden a cryptographic primitive against simple side-channel attacks, we make the observable information independent of the secret scalar. This can be achieved by applying Montgomery's ladder for scalar multiplication.

3 Proposed Mutual Authentication Protocol

A mutual authentication protocol is necessary to resist the attacks when a malicious user pretends as an authorized one and duplicate, modify, insert or delete the data during transmission. We propose a mutual authentication protocol based on HECDSA [8] in wireless network which gives authentication and non-repudiation. The novelties of the protocol are as follows:

- Our proposed mutual authentication protocol based on HECDSA in wireless network is suitable for constrained devices as it uses genus 2 HEC on 80-bit finite fields which achieve the same security level as 160-bit ECC.
- Any remote user can obtain service from other users without registering each time with the KDC. They can transfer data after mutual authentication.
- New session key is established for each particular session to protect data which resists replay attack in wireless network.
- Encryption of transmitted message using asymmetric encryption process saves energy and storage, which is critical for constrained devices.

Notations used

C: A hyperelliptic curve of genus g defined over $\mathbb{F}_p$
p : A large prime number
q : A large prime divisor of $p-1$.
P: A Base point on the Hyperelliptic Curve
D: The semi reduced divisor of the HEC
D': The unique reduced divisor of the HEC
PR_A, PU_A: Private and Public key of A respectively
PR_B, PU_B: Private and Public key of B respectively
ID_A: Identity of A
ID_B: Identity of B
M: Input Message
(r,s): Signature pair
M': Received message
(r',s'): Received Signature pair
$H(.)$: One-way hash function with fixed length output
K: Common Secret key
K_s: Session key
K_a: Premaster key shared between Users and KDC
T_s: Session time
$\oplus$: Group addition between Jacobian elements

Our proposed protocol is described below:

During the initialization phase, Key Distribution Center (KDC) generates a random hyperelliptic curve C defined over $\mathbb{F}_p$. Then KDC computes semi reduced Divisor D and the unique reduced divisor D' of the selected curve using Cantor's algorithm. In this protocol we use the unique reduced divisor (D') and semi reduced divisor (D) distinctly.

KDC also computes a point $P = (x_1,y_1) \in C(\mathbb{F}_p)$ which is a base point on the curve, a large prime number p and a prime divisor q such that q divides $p-1$. $\mathbb{F}_p$ contains the representation of all field elements of order n. Finally the following system parameters ($\mathbb{F}_p$, C, D', p, q, D, n) are generated by KDC.

During the network deployment phase, all Users send request message to KDC for getting registered in the network. After registration KDC assigns a unique ID to each User and send ID with a Premaster key K_a to each User. KDC also makes a list of all Users with their ID.

After the network deployment phase, KDC publish this list encrypted by the Premaster key K_a to all Users of the network with the system parameters ($\mathbb{F}_p$, C, D', p, q, D, n). We also assume after deployment of the Users (nodes), they become static.

Now User A wants to communicate with User B. So User A sends a request message to User B containing ID_A and a nonce N_1.Once the message received, if User B wants to communicate with A, it first verify the ID from the list. If it matches, then B sends an accept message to User A containing ID_B and nonce N_1.

Now User A and User B will communicate after the mutual authentication.

Our proposed mutual authentication scheme is shown in Figure 1.

In this scheme the session key are generated using following three major steps:

Step 1

User B chooses a random challenge d_B, where $1 \leq d_B \leq n-1$, then it calculates

$Q_B = d_B \times D' = [u_B, v_B]$ using scalar multiplication in genus 2 HECC described in [8].

Then User B generates private key $PR_B \in_R N$ [choose a positive prime at random in N] and public key $PU_B = [PR_B]D$. Here PU_B is represented using Mumford representation which is of the form $[u_B, v_B]$.

Finally User B sends (Q_B, PU_B) to User A.

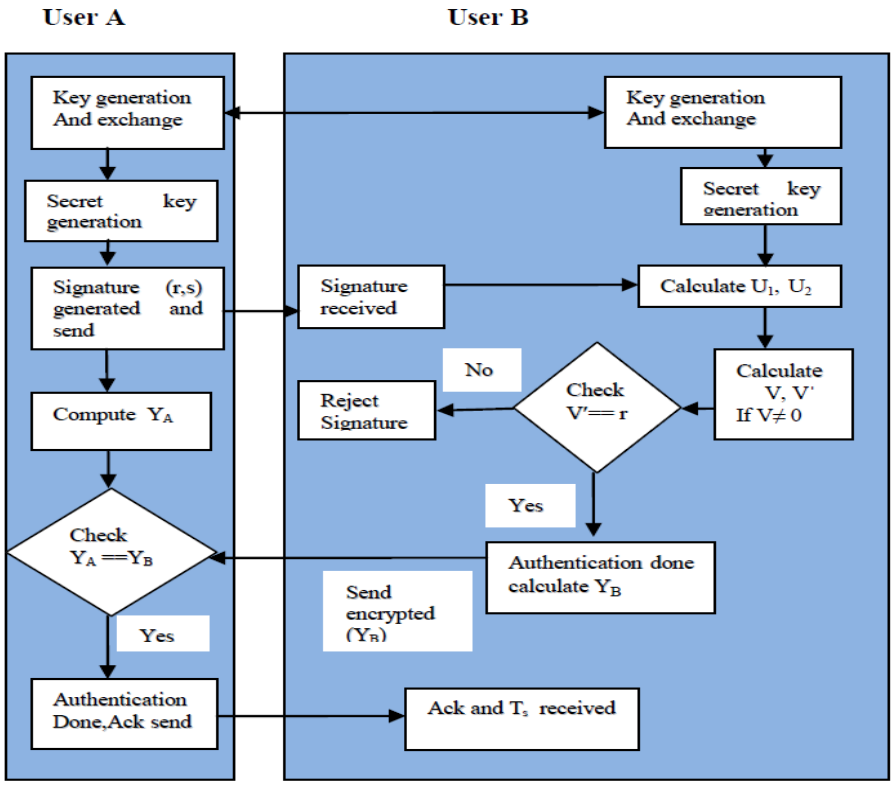


Figure 1: Mutual authentication between two users

Step 2

User A chooses a random challenge d_A, where $1 \leq d_A \leq n-1$, then computes

$Q_A = d_A \times D' = [u_A, v_A]$ using scalar multiplication in genus 2 HECC.

Then A computes $z = Q_A \oplus Q_B$ for mutual authentication (by performing Divisor addition in affine coordinate).

The user A generate private key $PR_A \in_R N$ [choose a positive prime at random in N] and public key $PU_A = [PR_A]\, D$.

After that User A compute the secret key $K = d_A \times Q_B = [u_i, v_i]$.

In addition, User A calculates $r = (\sum_{i=0}^{e-1} L(u_i) q^i) \bmod p$ where

e is an integer with $e \leq g$ and assuming the finite field

elements are ordered such that $0 \leq L(u_i) < q$ [for mapping between Jacobian $J(\mathbb{F}_p)$ and finite field $GF(p)$].

User A also computes $s = [r^{-1}(H(M) - [PR_A]\, r)] \bmod p$.

Finally (r, s) becomes the signatures pair and A transfers signature pairs with (Q_A, PU_A) to User B.

Step 3

User B computes $\beta = Q_A \oplus Q_B$ by performing Divisor addition in affine coordinate.

He also computes secret key $K = d_B \times Q_A = [u_i, v_i]$. If the protocol works correctly, both the Users generate the same value of K.

This can be proved by the simple mathematical calculation shown below:

$K = d_B \times Q_A = d_B \times d_A \times D' = d_A \times d_B \times D' = d_A \times Q_B = [u_i, v_i]$

User B also calculates $w = (s')^{-1} \bmod p$ where

(r', s') is the received signature.

After that User B calculates $U_1 = (H(M')\, w) \bmod p$ and $U_2 = (r'w) \bmod p$.

In addition, B calculates $V = [U_1]\, D \oplus [U_2]\, PU_A = [u_f, v_f]$. If $V = [1,0]$, it implies that the signature is incorrect and User B rejects the signature with message.

Otherwise User B calculates

$V' = (\sum_{i=0}^{e-1} L(u_{f,i}) q^i) \bmod p$

[for mapping between Jacobian $J(\mathbb{F}_p)$ and finite field $GF(p)$].

If $(V' == r)$, it implies that the signature is correct, so User B authenticates User A and User B can be confirmed that User A has actually established the same secret key.

Then User B computes $Y_B = H(\beta) + u_i$ and sends encrypted Y_B (encryption done by using secret key K) to User A. User A decrypt the packet and find Y_B.

In order to authenticate User B, User A will compute $Y_A = H(z)+ u_i$ and then User A will verify the value of Y_A by checking that ($Y_A == Y_B$).

Table 1: Experimental Results of genus 2 HECC (Prime Field)

G-2 Prime Field Group order	Divisor Generation (ms)	Public Key PU_A (ms)	Public Key PU_B (ms)	User-A Secret key K (ms)	User-B Secret key K (ms)	Signature Generation (ms)	Signature Verification (ms)
(2^152)	10	16	15	11	15	84	28
(2^158)	11	16	15	16	15	72	30
(2^162)	10	15	14	10	9	60	31
(2^166)	11	16	16	15	16	56	32

Table 2: Comparison of our experimental result with existing literature

Reference	Curves	Group Order	Signature Generation (ms)	Signature Verification (ms)
N. Jansma et al. [12]	ECC (genus-1)	2^163	150	230
M. Aydos et al. [5]	ECC (genus-1)	2^160	46	92
Our results	HECC (genus-2)	(2^152)	84	28
		(2^158)	72	30
		(2^162)	60	31
		(2^166)	56	32

If they match, then User A authenticates User B and User A can be confirmed that User B has actually established the same secret key with him. User A then sends an acknowledgement with the session time T_s to User B.

Finally, A and B agree on the common session key K_s where $K_s = H(ID_A|| ID_B||K)$

If all the above steps are executed correctly, both sides will agree on the session Key K_s. Once the protocol run completes successfully, both sides may use K_s to encrypt messages (using ElGamal method) with timestamp for subsequent session traffic in order to create a confidential communication channel. After each valid session a new session key will generate.

4 Security Analysis

In this section, we discuss the security of our proposed mutual authentication protocol based on HECDSA. The proposed protocol will be considered to be a secure authenticated protocol, if it satisfies the following properties:

Man in the middle attack: It can be considered as an active attack. In this protocol, no useful information about the secret key K is revealed during a successful run. If an attacker E intercepts the message packet containing (Q_B , PU_B), E then receives PU_B and Q_B from B. However, this means that E must calculate K but E cannot compute the value of K because E does not know the value of d_A or the value of d_B. This problem is called Computational Diffie-Hellman Problem (CDHP). So, E will not be able to compute K. Thus this protocol resists the man in-the-middle attack.

Small subgroup attack: If hyperelliptic curve C has enough prime factors, the attacker could determine the secret scalar modulo of all these primes and recover a large part of the secret by using Chinese remaindering. To avoid this attack, we check D has order l where l is prime. For checking this, we first check that $[l]D=0$ and computing $[h]D$ for $h=c/p_i$, for all prime divisors p_i of c and checking that the result is not 0 [8].

Known-key attack: In our proposed protocol, both users generate new PU_A and PU_B in every new session, and in addition the secret key K is generated with every new session also. Another important aspect of our protocol is that the session key is calculated independently on both sides and protected by the secure hash function. Thus our proposed protocol is secure against known key attacks assuming that the hyperelliptic curve discrete logarithm problem is intractable.

Perfect forward secrecy: In perfect forward secrecy, even if the user's ID is compromised, it never allows the adversary to determine the session key for past sessions and decrypt them. In our protocol, it is based upon the assumption that the discrete logarithm problem is intractable and on the value of the secret key K. Even if the attacker knew the correct Q_B, the attacker still cannot compute the previous session keys because K_s is derived from the secret key K which is generated from the value of d_A and d_B . Thus the property of perfect forward secrecy is satisfied by our proposed protocol.

Replay attack: Replay attack involves passive capture of data and its subsequent retransmission to show unauthorized effect. Any unauthorized malicious user can send duplicate data repeatedly to the receiver which is already sent. Our protocol protects replay attack as it depends upon timestamp values. In our protocol, after each valid session time T_s which is unknown to malicious user, a new session key will generate for encryption, so that replay attack is not possible.

5 Implementation Results

We have implemented our proposed protocol of HEC (genus 2) on different prime fields using *jdk1.6.2*. The timings of basic operations have been measured on a PC with Intel Core i3 CPU 540 @3.07 GHz and Windows 7 operating system having jdk1.6. For HECC implementation, we have considered the genus 2 hyperelliptic curve $C: y^2 = x^5 + x^3 + 1$ over the finite field $\mathbb{F}_p$. Considering $D = (u(x), v(x))$, $D' = (u'(x), v'(x)) \in J(C)$, we generate $D = (x^2 + 6x + 33, 22x + 47)$ and $D' = (x^2 + 13x + 12, 46x + 59)$. Next we have implemented our proposed scheme using Hyperelliptic curves of different group order and main operations like key Generation, Signature Generation / Verification timing of the proposed mutual authentication protocol are listed in the Table 1.

The proposed protocol is suitable for constrained devices as it requires less key size and has low storage requirement for user side. The protocol has also low computational load (4 point multiplication + 2 secret key encryption / decryption + 1 signature generation / verification + 3 SHA-1 operation) on each side.

It can be seen that the proposed protocol of HEC (genus 2) is efficient as the timings of our signature generation / verification compares favorably with the timings of ECC available in existing literature as shown in the Table 2.

As HECC (genus 2) of 80-bit operand lengths provide same security level with ECC of 160-bit, it can be stated that HECC is more suitable for implementation in the constrained platforms in wireless networks.

6 Conclusions

HECC is well suited for secure communication in wireless network for constrained devices as HEC operand size is only a fractional amount of the EC operand size and almost all the standard discrete logarithm based protocols such as the Diffie-Hellman and EIGamal can be planted to HEC. We have proposed in this paper a mutual authentication protocol based on HECDSA for a secure access in constrained devices which allows both the entities to verify each other's authenticity. It is seen that the proposed protocol of HECC is efficient as the timings of our signature generation / verification compares favorably with the timings of ECC available in existing literature. As HECC (genus 2) of 80-bit operand lengths provide same security level with ECC of 160-bit, in our view, HECC is more suitable for implementation in the constrained platforms in wireless networks.

REFERENCES

[1] R. M. Avanzi, "Aspects of hyper-elliptic curves over large prime fields in software implementations," Cryptographic Hardware and Embedded Systems, LNCS vol. 3156, pp. 148-162, 2004.

[2] M. A. Azim and A. Jamalipour. "An efficient elliptic curve cryptography based authenticated key agreement protocol for wireless LAN security", International Workshop on High Performance Switching and Routing (HPSR'05), pp. 376-380, 2005.

[3] P. E. Abi-char, A. Mhamed, and B. E. Hassan, "A secure authenticated key agreement protocol based on elliptic curve cryptography", IEEE International Symposium on Information Assurance and Security, vol. 57, pp. 89-94, 2007.

[4] L. Adleman, J. DeMarrais, and M. Huang, "A subexponential algorithm for discrete. logarithms over the rational subgroup of the Jacobians of large genus hyperelliptic curves over finite fields", Algorithmic Number Theory (ANTS-1), LNCS 877, pp. 28-40, 1994.

[5] M. Aydos, T. Yanık, and C. K. Koc, "High-speed implementation of an ECC-based wireless authentication protocol on an ARM microprocessor", IEE Proceedings: Communications, vol.148, no. 5, pp. 273–279, 2001.

[6] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks", Proceedings of the IEEE Security and Privacy Symposium, pp. 197-213 , 2003.

[7] D. G. Cantor, "Computing in the Jacobian of a hyperelliptic curve", Mathematics of Computation, vol. 48, pp. 95-101, 1987.

[8] H. Cohen and G. Frey, "Handbook of Elliptic and Hyperelliptic Curve Cryptography", Chapman & Hall/CRC Press, 2006.

[9] K. Chatterjee and D. Gupta, "Evolution of Hyperelliptic Curve Cryptosystems", in proceedings of ICDCIT 2010, LNCS 5966, pp.206-211, Springer -Verlag Berlin Heidelberg 2010.

[10] L. Eschenauer and V. Gligor. "A key management scheme for distributed sensor networks", Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS'02), pp. 41-47, 2002.

[11] G. Frey and H. Ruck, "A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves", Mathematics of Computation, vol. 62, pp. 865-874, 1994.

[12] N. Jansma and B. Arrendondo, "Performance Comparison of Elliptic Curve and RSA Digital Signatures", Technical Report, University of Michigan, 2004. (http://www.nicj.net/files/498termpaper.pdf)

[13] N. Koblitz, "Hyperelliptic cryptosystems", Journal of Cryptology, vol. 1, no. 3, pp. 139–150, 1989.

[14] L. Law, A. Menezes, M. Qu, J. Solinas, and S.Vanstane, "An efficient protocol for authenticated key agreement", Designs, Codes and Cryptography, vol. 28, pp. 361-377, 1998.

[15] T. Landstra, S. Jagannathan, and M. Zawodniok, "Energy-efficient hybrid key management protocol for wireless sensor networks", International Journal of Network Security, vol. 9, no. 2, pp. 121-134, 2009.

[16] A. Menezes, Y. Wu, and R. Zuccherato, "An elementary introduction to hyperelliptic curves", Technical Report

CORR 96-19, Department of C&O, University of Waterloo, Ontario, Canada, November 1996. (http://www.cacr.math.uwaterloo.ca/techreports/1997/tech-reports97.html)

[17] J. Pelzl, T. Wollinger, J. Guajardo, and C. Paar, "Hyperelliptic curve cryptosystems: Closing the performance gap to elliptic curves", Cryptology ePrint Archieve, Report 026, pp. 351-365, 2003. (http://eprint.iacr.org/)

[18] J. Pelzl, T. Wollinger, and C. Paar, "Elliptic & hyperelliptic curves on embedded µP", ACM Transactions on Embedded Computing Systems, vol. 3, no. 3, pp. 509-533, 2004.

[19] H. G. Ruck "On the discrete logarithms in the divisor class group of curves", Mathematics Computation, vol. 68, pp. 805-806, 1999.

[20] L. Uhsadel, A. Poschmann, and C. Paar, "An efficient general purpose elliptic curve cryptography module for ubiquitous sensor networks", Workshop on Software Performance Enhancement for Encryption and Decryption (SPEED'07), pp. 95-104, 2007.

[21] R. Watro, D. Kong, S. F. Cuti, C. Gardiner, C. Lynn, and P. Kruus. "TinyPK: Securing sensor networks with public key technology", Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN'04), pp. 59–64, New York, USA, 2004.

[22] L. Yongliang, W. Gao1, H. Yao, and X. Yu, "Elliptic curve cryptography based wireless authentication protocol", International Journal of Network Security, vol. 5, no. 3, pp. 327–337, 2007.

Kakali Chatterjee has done M.Tech from Centre for Development of Advanced Computing, a R&D and Academic Centre of Govt. of India. She is currently pursuing Ph.D. at Delhi College of Engineering (Faculty of Technology), University of Delhi, India. Her field of interest is Information Security and Cryptography.

Asok De received Ph.D. from IIT Kharagpur (India) and his field of interest is Microwave Antennas and Communication Systems. He is Professor in Delhi Technological University (formerly Delhi College of Engineering). Presently he is working as Principal, Ambedkar Institute of Advanced Communication Technologies & Research, Delhi. He has published many research papers in reputed International Journals.

Daya Gupta is a Professor and Head of Computer Engineering Department of Delhi Technological University, India. She has done Ph.D. in Computer Engineering from Delhi University. Her field of interest is Software Engineering, Information Security etc. She has published many research papers in reputed International Journals.

On Dynamic Cumulative Residual Inaccuracy Measure

HC Taneja and Vikas Kumar *

Abstract—**An alternate measure of entropy based on distribution function rather than the density function of a random variable X, called the cumulative residual entropy (CRE) was proposed by Rao et al. (2004). In this communication the concept of CRE has been extended to cumulative residual inaccuracy (CRI) and then to a dynamic version of it. A characterization problem for the proposed dynamic inaccuracy measure has been studied under proportional hazard model. Three specific lifetime distributions exponential, Pareto and the finite range have been characterized using the proposed dynamic inaccuracy measure.**

Keywords: entropy, cumulative residual entropy, inaccuracy, hazard rate , mean residual life function, proportional hazard model.

1 Introduction

Let X and Y be two non-negative random variables representing time to failure of two systems with p.d.f. respectively $f(x)$ and $g(x)$. Let $F(x) = P(X \leq x)$ and $G(Y) = P(Y \leq y)$ be failure distributions, and $\bar{F}(x) = 1 - F(x)$, $\bar{G}(x) = 1 - G(x)$ be survival functions of X and Y respectively . Shannon's (1948) measure of uncertainty associated with the random variable X and Kerridge measure of inaccuracy (1961) are given by

$$H(f) = -\int_0^\infty f(x) \log f(x) dx \ . \tag{1}$$

and

$$H(f;g) = -\int_0^\infty f(x) \log g(x) dx \ . \tag{2}$$

respectively. In case $g(x) = f(x)$, then (2) reduces to (1).

The measures (1) and (2) are not applicable to a system which has survived for some unit of time. Ebrahimi (1996) considered the entropy of the residual lifetime $X_t = [X - t|X > t]$ as a dynamic measure of uncertainty given by

$$H(f;t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} log \frac{f(x)}{\bar{F}(t)} dx \ . \tag{3}$$

*Department of Applied Mathematics, Delhi Technological University, (Formerly Delhi College of Engineering) Bawana Road, Delhi-110042, India. Email: hctaneja@rediffmail.com

Extending the dynamic measure of information, a dynamic measure of inaccuracy, refer to Taneja et al. (2009) is given as

$$H(f,g;t) = -\int_t^\infty \frac{f(x)}{\bar{F}(t)} log \frac{g(x)}{\bar{G}(t)} dx \ . \tag{4}$$

Rao et al. (2004) introduced an alternate measure of entropy called cumulative residual entropy(CRE) of a random variable X defined as

$$\xi(F) = -\int_0^\infty \bar{F}(x) \log \bar{F}(x) dx \ , \tag{5}$$

where $\bar{F}(x) = 1 - F(x)$ is the survival function.

This measure is based on cumulative distribution function (CDF) rather than probability density function, and is thus, in general more stable since the distribution function is more regular because it is defined in an integral form unlike the density function which is defined as the derivative of the distribution. Some general results regarding this measure have been studied by Rao (2005), Drissi et al. (2008) and Navarro et al. (2009).

Asadi and Zohrevand (2007) have defined the dynamic cumulative residual entropy (DCRE) as the cumulative residual entropy of the residual lifetime $X_t = [X - t|X > t]$. This is given by

$$\xi(F;t) = -\int_t^\infty \frac{\bar{F}(x)}{\bar{F}(t)} \log \frac{\bar{F}(x)}{\bar{F}(t)} dx \ . \tag{6}$$

In this communication in Section 2, we define a cumulative residual inaccuracy measure. In Section 3, we propose a dynamic cumulative residual inaccuracy measure. In Section 4, we prove that dynamic measure determines the lifetime distribution functions uniquely, and characterize three specific lifetime distributions in this context.

2 Cumulative Residual Inaccuracy

If $\bar{F}(x)$ and $\bar{G}(x)$ are survival functions of lifetime random variables X and Y respectively, then the cumulative residual inaccuracy (CRI) is defined as

$$\xi(F;G) = -\int_0^\infty \bar{F}(x) \log \bar{G}(x) dx \ . \tag{7}$$

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

When these two distributions coincide, the measure (7) reduces to the cumulative residual entropy (5).

If the two random variables X and Y satisfy the proportional hazard model (PHM), refer to Cox (1959) and Efron (1981), that is, if $\lambda_G(x) = \beta\lambda_F(x)$, or equivalently

$$\bar{G}(x) = [\bar{F}(x)]^{\beta}, \tag{8}$$

for some constant $\beta > 0$, then obviously the cumulative residual inaccuracy (7) reduces to a constant multiple of the cumulative residual entropy (5).

Example 2.1 Let X be a non-negative random variable with p.d.f.

$$f_X(x) = \begin{cases} 2x & ; \quad \text{if } 0 \le x < 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

and survival function $\bar{F}(x) = 1 - F(x) = (1 - x^2)$, and let the random variable Y be uniformly distributed over $(0, 1)$, then its density and survival functions are respectively given by

$$g_Y(x) = 1 \ \textit{and} \ \bar{G}_Y(x) = 1 - x\ , \ \ 0 < x < 1.$$

Substituting these in (7) , we obtain the cumulative residual inaccuracy as

$$\xi(F, G) = \frac{7}{18}.$$

Example 2.2 Let a non-negative random variable X be uniformly distributed over (a, b), $a < b$, with density and distribution functions respectively given by

$$f(x) = \frac{1}{b-a} \ \textit{and} \ F(x) = \frac{x-a}{b-a}, \ \ a < x < b.$$

If the random variables X and Y satisfy the proportional hazard model (PHM), then the distribution function of the random variable Y is

$$\bar{G}(x) = [\bar{F}(x)]^{\beta} = \left[\frac{b-x}{b-a}\right]^{\beta} \ \ a < x < b, \ \beta > 0.$$

Substituting these in (7) and simplifying we obtain the cumulative inaccuracy measure as

$$\xi(F; G) = \frac{\beta(b-a)}{4}\ . \tag{9}$$

3 Dynamic Cumulative Residual Inaccuracy

In life-testing experiments normally the experimenter has information about the current age of the system under consideration. Obviously the CRI measure (7) defined above is not suitable in such a situation and should be modified to take into account the current age also. Further, if X is the lifetime of a component, which has already survived upto time t, then the random variable $X_t = [X - t|X > t]$ called the residual lifetime random variable has the survival function

$$\bar{F}_t(x) = \begin{cases} \frac{\bar{F}(x)}{\bar{F}(t)} & ; \quad \text{if } x > t \\ 1 & ; \quad \text{otherwise} \end{cases}$$

and similarly for $\bar{G}_t(x)$. Thus the cumulative inaccuracy measure for the residual lifetime distribution is given by

$$\xi(F, G; t) = -\int_t^{\infty} \bar{F}_t(x) \log \bar{G}_t(x) dx \tag{10}$$

$$= -\int_t^{\infty} \frac{\bar{F}(x)}{\bar{F}(t)} \log \frac{\bar{G}(x)}{\bar{G}(t)} dx\ . \tag{11}$$

Obviously when $t = 0$, then (11) becomes (7).

Example 3.1 Let X be a non-negative random variable with p.d.f.

$$f_X(x) = \begin{cases} 2x & ; \quad \text{if } 0 \le x < 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

and the survival function $\bar{F}(x) = 1 - F(x) = (1 - x^2)$, and let the random variable Y be uniformly distributed over $(0, 1)$ with density and survival functions given respectively by

$$g_Y(x) = 1 \ \textit{and} \ \bar{G}_Y(x) = 1 - x\ , \ \ 0 < x < 1.$$

Substituting these values in (11) , we obtain the dynamic cumulative residual inaccuracy measure as

$$\xi(F, G; t) = \left(\frac{9(1-t) - 2(1-t)^2}{18(1+t)}\right).$$

Example 3.2 Let X and Y be two non-negative random variables with survival functions $\bar{F}(x) = (x+1)e^{-x}, \ x > 0$ and $\bar{G}(x) = e^{-2x}, \ x > 0$. Substituting these values in (11), we obtain

$$\xi(F, G; t) = \left(\frac{6+2t}{t+1}\right).$$

Taking limit as $t \to 0$, we obtain

$$\lim_{t \to 0} \xi(F, G; t) = \xi(F, G) = 6.$$

4 Characterization Problem

The general characterization problem is to determine when the proposed dynamic inaccuracy measure (11) characterizes the distribution function uniquely. It is known that the hazard rate $\lambda_F(t) = \frac{f(t)}{\bar{F}(t)}$ and the mean residual life function $m_F(t) = \frac{\int_t^{\infty} \bar{F}(x) dx}{\bar{F}(t)}$ characterize the distribution and the relation between the two is given by

$$\lambda_F(t) = \frac{1 + m_F'(t)}{m_F(t)}\ . \tag{12}$$

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

We study the characterization problem under the proportional hazard model (8).

Theorem 4.1 Let X and Y be two non negative random variables with survival functions $\bar{F}(x)$ and $\bar{G}(x)$ satisfying the proportional hazard model (8). Let $\xi(F,G;t) < \infty, \forall\, t \geq 0$ be an increasing function of t, then $\xi(F,G;t)$ uniquely determines the survival function $\bar{F}(x)$ of the variable X.

Proof. Rewriting (11) as

$$\xi(F,G;t) = -\frac{1}{\bar{F}(t)} \int_t^\infty \bar{F}(x) \log \bar{G}(x) dx + m_F(t) \log \bar{G}(t) \ , \tag{13}$$

where $m_F(t)$ is the mean residual life function. Substituting (8) into (13) gives

$$\xi(F,G;t) = -\frac{\beta}{\bar{F}(t)} \int_t^\infty \bar{F}(x) \log \bar{F}(x) dx + \beta m_F(t) \log \bar{F}(t) \ .$$

Differentiating this w.r.t. t both sides, we obtain

$$\begin{aligned} \xi'(F,G;t) &= \beta \log \bar{F}(t)[1 + m_F'(t)] - \beta \lambda_F(t) m_F(t) \\ &\quad -\beta \lambda_F(t) \int_t^\infty \frac{\bar{F}(x)}{\bar{F}(t)} \log \bar{F}(x) dx \ , \end{aligned} \tag{14}$$

where $\lambda_F(t)$ is hazard rate function. Substituting (12) and (13) in (14) we obtain

$$\xi'(F,G;t) = \lambda_F(t)\{\xi(F,G;t) - \beta m_F(t)\} \ . \tag{15}$$

Let F_1, G_1 and F_2, G_2 be two sets of the probability distribution functions satisfying the proportional hazard model, that is, $\lambda_{G_1}(x) = \beta \lambda_{F_1}(x)$, and $\lambda_{G_2}(x) = \beta \lambda_{F_2}(x)$, and let

$$\xi(F_1, G_1; t) = \xi(F_2, G_2; t) \ \forall \ \ t \geq 0 \ . \tag{16}$$

Differentiating it both sides w.r.t. t, and using (15), we obtain

$$\lambda_{F_1}(t)\{\xi(F_1,G_1;t) - \beta m_{F_1}(t)\} = \lambda_{F_2}(t)\{\xi(F_2,G_2;t) - \beta m_{F_2}(t)\} \tag{17}$$

If for all $t \geq 0$, $\lambda_{F_1}(t) = \lambda_{F_2}(t)$, then $\bar{F}_1(t) = \bar{F}_2(t)$ and the proof will be over, otherwise, let

$$A = \{t : t \geq 0, and \ \ \lambda_{F_1}(t) \neq \lambda_{F_2}(t)\} \tag{18}$$

and assume the set A to be non empty . Thus for some $t_0 \in A$, $\lambda_{F_1}(t_0) \neq \lambda_{F_2}(t_0)$. Without loss of generality suppose that $\lambda_{F_2}(t_0) > \lambda_{F_1}(t_0)$. Using this, (17) for $t = t_0$ gives

$$\xi(F_1, G_1; t_0) - \beta m_{F_1}(t_0) > \xi(F_2, G_2; t_0) - \beta m_{F_2}(t_0),$$

which implies that

$$m_{F_1}(t_0) < m_{F_2}(t_0),$$

a contradiction. Thus the set A is empty set and this concludes the proof.

Next, we give the characterization theorem.

Theorem 4.2 Let X and Y be two non-negative continuous random variables satisfying the PHM (8). If X is with mean residual life $m_F(t)$, then the dynamic cumulative residual inaccuracy measure

$$\xi(F,G;t) = c\, m_F(t), \ \ c > 0 \tag{19}$$

if, and only if
(i) X follows the exponential distribution for $c = \beta$,
(ii) X follows the Pareto distribution for $c > \beta$,
(iii) X follows the finite range distribution for $0 < c < \beta$.

Proof. First we prove the 'if' part.
(i) If X has exponential distribution with survival function $\bar{F}(x) = \exp(-\theta x), \theta > 0$, then the mean residual life function $m_F(t) = \frac{1}{\theta}$. The dynamic cumulative residual inaccuracy measure (11) under PHM is given as

$$\xi(F,G;t) = \frac{\beta}{\theta} = c m_F(t),$$

for $c = \beta$.

(ii) If X follows Pareto distribution with p.d.f.

$$f(x) = \frac{a b^a}{(x+b)^{a+1}}, \quad a \ > 1, \ \ b > \ 0,$$

then the survival function is

$$\bar{F}(x) = 1 - F(x) = \left(1 + \frac{x}{b}\right)^{-a} = \frac{b^a}{(x+b)^a} \ ,$$

and the mean residual life is

$$m_F(t) = \frac{\int_t^\infty \bar{F}(x) dx}{\bar{F}(t)} = \frac{t+b}{a-1} \ . \tag{20}$$

The dynamic cumulative inaccuracy measure (11), under PHM is given by

$$\xi(F,G;t) = \frac{\beta a (t+b)}{(a-1)^2} = c m_F(t) \ ,$$

for $c = \frac{\beta a}{a-1} > \beta$.

(iii) In case X follows finite range distribution with p.d.f.

$$f(x) = a(1-x)^{a-1}, \quad a \ > 1, \ \ 0 \leq x \leq \ 1,$$

then the survival function is

$$\bar{F}(x) = 1 - F(x) = (1-x)^a,$$

and the mean residual life is

$$m_F(t) = \frac{1-t}{a+1}.$$

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The inaccuracy measure (11) under PHM is given by

$$\xi(F,G;t) = \frac{\beta a(1-t)}{(a+1)^2} = cm_F(t) \ ,$$

for $c = \frac{\beta a}{a+1} < \beta$.
This proves the 'if' part.

To prove the only if part, consider (19) to be valid. Using (13) under PHM, it gives

$$-\frac{\beta}{\bar{F}(t)} \int_t^\infty \bar{F}(x) \log \bar{F}(x) dx + \beta m_F(t) \log \bar{F}(t) = cm_F(t) \ .$$

Differentiating it both sides w.r.t. t, we obtain

$$\begin{aligned} \frac{c}{\beta} m'_F(t) &= m'_F(t) \log \bar{F}(t) - \lambda_F(t) m_F(t) + \log \bar{F}(t) \\ &\quad -\lambda_F(t) \frac{1}{\bar{F}(t)} \int_t^\infty \bar{F}(x) \log \bar{F}(x) dx, \\ &= m'_F(t) \log \bar{F}(t) - \lambda_F(t) m_F(t) + \log \bar{F}(t) \\ &\quad +\lambda_F(t) [\frac{c}{\beta} m_F(t) - m_F(t) \log \bar{F}(t)] \ . \end{aligned}$$

From (12) put $m'_F(t) = \lambda_F(t) m_F(t) - 1$ and simplify, we obtain

$$\lambda_F(t) m_F(t) = \frac{c}{\beta} \ ,$$

which implies

$$m'_F(t) = \frac{c}{\beta} - 1 \ .$$

Integrating both sides of this w.r.t. t over $(0,\ x)$ yields

$$m_F(x) = (\frac{c}{\beta} - 1)x + m_F(0). \qquad (21)$$

The mean residual life function $m_F(x)$ of a continuous non-negative random variable X is linear of the form (21) if, and only if the underlying distribution is exponential $for\ c = \beta$, Pareto $for\ c > \beta$, or finite range $for\ 0 < c < \beta$, refer to Hall and Wellner (1981). This completes the theorem.

Conclusion By considering the concept of cumulative residual inaccuracy and extending it to its dynamic version, we have characterized certain specific life-time distributions functions like exponential, Pareto and the finite range distributions which play a vital role in reliability modeling. Also the results reported generalize the existing results in context with cumulative residual entropy.

References

[1] Asadi, M., Zohrevand, Y. (2007). On the dynamic cumulative residual entropy. Journal of Stat. Planning and Inference 137,1931-1941.

[2] Cox, DR. (1959). The analysis of exponentially distributed lifetimes with two type of failure, J. Roy. Statist. Soc. Ser. B, 21: 411-421.

[3] Drissi, N., Chonavel, T., Boucher, J.M. (2008). Generalized cumulative residual entropy for distributions with unrestricted supports. Re. Lett. Sign. Proc.

[4] Ebrahimi, N. (1996). How to measure uncertainty about residual lifetime? Sankhya A 58: 48-57, .

[5] Efron, B. (1981). Censored data and the bootstrap , J. Am. Statist. Assoc, 76: 312-319.

[6] Hall, W.J., Wellner, J.A. (1981). Mean residual life. In Statistics and Related Topics (Csrg, et al. eds), North-Holland, Amsterdam, 169-184.

[7] Kerridge, D.F. (1961). Inaccuracy and inference. J. Roy. Statist. Soc. Ser. B 23: 184-194.

[8] Navarro, J., Aguila, Y.D., Asadi, M. (2009). Some new results on the cumulative residual entropy. Journal of Stat. Planning and Inference.

[9] Rao, M., Chen, Y., Vemuri, B.C., Wang, F. (2004). Cumulative residual entropy: a new measure of information . IEEE Transactions on Information Theory 50, 1220-1228.

[10] Rao, M., (2005). More on a new concept of entropy and information. Journal of Theoretical Probability 18,967-981.

[11] Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, pp. 279-423 and 623-656.

[12] Taneja, H.C, Kumar, V. Srivastava, R. (2009). A dynamic measure of inaccuracy between two residual lifetime distributions. International Mathematical Forum 4(25):1213-1220.

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Int J Theor Phys
DOI 10.1007/s10773-012-1287-3

Particle Creation in Higher Dimensional Space-time with Variable G and Λ

C.P. Singh · A. Beesham

Received: 14 February 2012 / Accepted: 28 July 2012

Abstract In this paper we consider higher dimensional FRW type cosmological models with time-varying gravitational constant $G = G(t)$ and decaying vacuum energy $\Lambda = \Lambda(t)$ in context of the open thermodynamical system of particle creation. The particle creation rate is determined by variable G and Λ. An Exact solution of the gravitational field equations is obtained by assuming the exact power-law solution of the scale factor of the universe. We evaluate the various cosmological parameters and discuss their physical significance in detail for radiation and matter-dominated phases of the universe. The solutions are also relevant for quintessence (accelerating) universe. We also discuss the relevance of the solutions for 5-dimensional Kaluza-Klein FRW type model with and without particle creation in radiation and matter-dominated phases as special case. We obtain the solution where the gravitational constant increases with time during no particle creation. The cosmological constant Λ, however, decreases with time in all the cases.

Keywords Cosmology · Exact solutions · Particle creation · Vacuum energy density

1 Introduction

The discovery during the 1920's of a systematic relation between redshift and distance removed any interest in static Einstein universe as a realistic model. Nevertheless, the existence of a cosmological constant Λ remains a logical possibility. Cosmologists have thoroughly explored the dynamics of expanding universe with a cosmological constant [1, 2].

C.P. Singh (✉)
Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi 110 042, India
e-mail: cpsphd@rediffmail.com

A. Beesham
Department of Mathematical Sciences, University of Zululand, Private Box. X1oo1, Kwa-Dlangezwa 3886, South Africa
e-mail: abeesham@pan.uzulu.ac.za

Published online: 11 August 2012

Recent observations by Perlmutter et al. [3] and Riess et al. [4] have strongly favour a significant and a positive value of Λ. The present estimate of cosmological constant is very small. One way to resolve this problem is to assume that Λ is not a pure constant, but rather decreases continuously with cosmic expansion. Recent observations in particle physics and cosmology indicate that Λ-term ought to be a dynamical quantity rather than a simple constant. The dynamics of the scale factor in Friedmann-Robertson-Walker (FRW) models with a variable Λ have been discussed by Overduin and Cooperstock [5]. The cosmological constant as a function of cosmic time have been studied by many authors. Berman et al. [6] and Berman [7, 8] have studied that the relation $\Lambda \propto t^{-2}$ plays an important role in cosmology. Bertolami [9] has studied a cosmological model with a time dependent cosmological term. Later, Özer and Taha [10, 11] and others [12–15] have obtained cosmological models of the universe, which are free from cosmological problems, by considering a varying cosmological constant in the case of closed universe. A number of papers [16–19] have appeared which describe various cosmological issues, such as, the age problem, observational constraints on Λ, structure formation, and gravitational lenses. Peebles and Ratra [20, 21] have studied the dark energy content of the universe considering a varying Λ, which may be realized using scalar field. Sahni and Starobinsky [22] have discussed the observational and theoretical aspects of a small positive cosmological constant term. Padmanabhan [23] has discussed several aspects of the cosmological constant problem in the framework of string theory.

The question of time variation is more interesting in the case of the gravitational coupling constant G since several theories of gravity have been suggested in which G decreases with cosmic time. Dirac [24, 25] proposed a theory of variable gravitational constant G based on the Large Number Hypothesis (LNH). He postulated that G varies inversely with cosmic time. If the gravitational constant G varies, then general relativity needs to be replaced with some other field theory of gravitation. However, Dirac did not specify what this field theory would be like, so his cosmological models remained incomplete. Dirac's theory inspired a number of attempts to formulate a field theory in which the effective constant of gravitation is some function of scalar field. Brans and Dicke [26] proposed the most interesting and complete scalar-tensor theory of gravitation where the gravitational constant G is replaced by the reciprocal of a scalar field. Under most circumstances, G can be regarded as a genuine constant after a time, which is distant enough from the big bang. The implication of a time-varying G will effect only in the early epoch or when the overall history and evolution of the universe is considered. Hence, with the approximation of zero Λ and constant G, we can immediately go over to Einstein's field equations.

Lau [27] and Lau and Prokhovnik [28] have proposed a generalized theory of gravitation incorporating time-dependent G and Λ terms. Thus a modified theory was proposed treating G and Λ as coupling scalars within the Einstein equations and hence the principle of equivalence demands that only metric tensor g_{ij} and not G and Λ must enter the equation of motion of particles and photons. In this way the usual conservation law holds. Taking the divergence of the Einstein equations and using the Bianchi identities we obtain an equation that controls the variation of G and Λ. These are modified field equations that allow to take into account a variable G and Λ. Nevertheless this approach has some drawbacks, for example, it can not be derived from Hamiltonian, although there are several advantages in this approach. In the last few years, a number of works in cosmology with variable G and Λ in context of a perfect fluid as a source in isotropic [29–34] as well as in anisotropic models [35–40] of the universe in the framework of Einstein theory have been considered.

The latest development of super-string theory and super gravitational theory have created interest among the researchers to consider higher dimensional space-times for studying

the evolution of the universe [41]. The most famous five dimensional theory proposed by Kaluza [42] and Klein [43], was the first theory in which extra dimensions is used to couple the gravity and electromagnetism. Many authors [44, 45] have extended the Kaluza-Klein five dimensional theory to higher dimensions. Overduin and Wesson [46] have presented an excellent review of higher dimensional unified theory in which the cosmological and astrophysical implications of extra dimensions have been discussed. A number of authors [47–52] have studied the cosmological models in higher dimensional space-time.

The problem of particle creation in the universe was first discussed by Prigogine et al. [53, 54]. Particle creation in cosmology and its phenomenological description has recently attracted a lot of interest (see, [55, 56]). The particle creation comes from introducing supplementary pressure (or bulk viscous type) terms in the expression of energy momentum tensor. The possibility that in the early universe particle creation and entropy generation processes were the result of the decaying of the false vacuum (with vacuum energy density interpreted as a cosmological constant Λ), in the presence of a variable gravitational coupling constant G, has been considered in detail by Harko and Mak [57, 58].

It is the purpose of the present paper to consider an alternative interpretation of cosmology with variable G and Λ, namely the possibility of phenomenologically description of particle creation and entropy generation in higher dimensional FRW model. The divergence of the energy momentum tensor is non-zero in the model with variable G and Λ and by supposing that G and Λ are independent field equations implies a matter creation process with the particle number creation rate determined by the variation of the G and Λ.

The paper is organized as follows. In Sect. 2 we present $(m+2)$-dimensional Friedmann-Robertson-Walker (FRW) line element and the Einstein field equations with variable G and Λ as a theory describing matter and entropy creation. A theory of thermodynamics of particle creation is discussed in Sect. 3. Exact solutions of the field equations are found in Sect. 4. Some particular solutions are obtained in Sect. 5. In Sect. 6 we discuss the cosmological solution in details.

2 Model and Field Equations

Let us consider $(m+2)$-dimensional homogeneous, isotropic and flat model represented by Friedmann-Robertson-Walker (FRW) space-time in the unit $c=1$

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + r^2\left(dx_m^2\right)\right], \tag{1}$$

where $a(t)$ represents the scale factor and

$$dx_m^2 = d\theta_1^2 + \sin^2\theta_1 d\theta_2^2 + \cdots + \sin^2\theta_1 \sin^2\theta_2 \cdots \sin^2\theta_{m-1} d\theta_m^2. \tag{2}$$

We assume that the universe is filled with perfect fluid whose energy momentum tensor is given by

$$T_{ij} = (\rho + p)u_i u_j - p g_{ij}, \tag{3}$$

where ρ is the energy density, p is the pressure of the cosmic fluid and u_i is the $(m+2)$ velocity vector such that $u_i u^j = 1$.

The Einstein's field equations with time dependent gravitational constant G and cosmological constant Λ are given by

$$R_{ij} - \frac{1}{2} g_{ij} R = -8\pi G T_{ij} - \Lambda g_{ij}, \tag{4}$$

where R_{ij}, g_{ij}, and R are the Ricci tensor, metric tensor and Ricci scalar, respectively. Using (1)–(3), the field equations (4) yield

$$\frac{m(m+1)}{2}\left(\frac{\dot{a}}{a}\right)^2 = 8\pi G(t)\rho + \Lambda(t), \tag{5}$$

$$m\frac{\ddot{a}}{a} + \frac{m(m-1)}{2}\left(\frac{\dot{a}}{a}\right)^2 = -8\pi G(t)p + \Lambda(t), \tag{6}$$

where an overhead dot denotes derivative with respect to cosmic time t.

Further, (5) and (6) lead to the continuity equation

$$\dot{\rho} + (m+1)(\rho + p)\frac{\dot{a}}{a} + \left(\rho\frac{\dot{G}}{G} + \frac{\dot{\Lambda}}{8\pi G}\right) = 0. \tag{7}$$

3 Thermodynamics of Particle Creation

We suppose that the particle content of the early universe is formed from a non-interacting co-moving relativistic fluid having a particle number density n and obeying the following equations of state

$$p = \omega\rho, \tag{8}$$

$$n = n_0\left(\frac{\rho}{\rho_0}\right)^{\frac{1}{1+\omega}}, \tag{9}$$

where ω is the equation of state (EoS) parameter lying in the interval $-1 < \omega \leq 1$. However, Supernovae (SNe) Ia data, cosmic microwave background (CMB) and galaxy clustering statistics [59] show that EoS parameter is lying in the range $-1.3 < \omega < -0.79$ for accelerating universe. Here, $n_0 \geq 0$ and $\rho_0 > 0$ are the present values of particle number density and energy density, respectively. The particle number density is supposed to satisfy the balance equation

$$\dot{n} + (m+1)nH = \psi(t)n, \tag{10}$$

where $\psi(t)$ is the particle creation rate of dimension of t^{-1}. ψ can have either sign, so that $\psi > 0$ in the case of particle source, $\psi < 0$ in the case of a particle sink (decay) and $\psi = 0$ gives no particle creation.

The entropy S generated during particle creation at temperature T can be obtained from the equation

$$T\frac{dS}{dt} = \frac{d(\rho V)}{dt} + p\frac{dV}{dt}, \tag{11}$$

where $V = a^{(m+1)}$ is the spatial co-moving volume. In a cosmological fluid where the density and pressure are functions of the temperature only, i.e., $\rho = \rho(T)$ and $p = p(T)$, the entropy is given by [60]

$$S = \frac{(\rho + p)a^{(m+1)}}{T} = \frac{(1+\omega)\rho a^{(m+1)}}{T}. \tag{12}$$

Maartens [61] has suggested that the Gibbs integrability condition shows explicitly that one can not independently specify an equation of state for the pressure and temperature. If the

equation of state for pressure is barotropic, i.e., $p = p(\rho)$, then the equation of state for the temperature should be barotropic, i.e., $T = T(\rho)$ and may be written as

$$T \propto \exp \int \frac{dp}{(\rho + p)}. \tag{13}$$

Using (8), Eq. (13) reduces to

$$T = T_0 \rho^{\frac{\omega}{(1+\omega)}}, \tag{14}$$

where T_0 is a proportionality constant. Thus, for adiabatic particle creation the cosmic temperature scales as $T \sim \rho^{\omega/(1+\omega)}$.

Comparing (7) and (10) with the help of (8) and (9) we find

$$\psi(t) = -\frac{1}{(1+\omega)G(t)}\left[\frac{\dot{\Lambda}(t)}{8\pi\rho(t)} + \dot{G}(t)\right]. \tag{15}$$

In order that (15) describes particle creation the condition $\psi(t) \geq 0$, is required. This inequality leads to the following restriction imposed on the time variations of $G(t)$ and $\Lambda(t)$

$$\frac{\dot{\Lambda}(t)}{8\pi\rho(t)} + \dot{G}(t) \leq 0. \tag{16}$$

From (11) and (7) we get

$$T\frac{dS}{dt} = -\frac{1}{G(t)}\left[\frac{\dot{\Lambda}}{8\pi\rho(t)} + \dot{G}(t)\right]\rho(t)a^{(m+1)} \tag{17}$$

or, equivalently

$$\frac{dS}{dt} = \frac{(1+\omega)\rho(t)a^{(m+1)}}{T}\psi(t). \tag{18}$$

From (17) and (18), the entropy as a function of the particle creation rate is given by

$$S(t) = S_0 e^{\int \psi(t)dt}, \tag{19}$$

where $S_0 \geq 0$ is a constant of integration.

In the case of no particle production, i.e., $\psi(t) = 0$, we obtain the usual particle conservation law of the standard cosmology. Assuming that the conservation law for material field is hold, the continuity Eq. (7) decouples to the following equations:

$$\dot{\rho} + (m+1)(1+\omega)H\rho = 0, \tag{20}$$

$$\dot{\Lambda}(t) + 8\pi\rho\dot{G}(t) = 0. \tag{21}$$

In this case, the entropy $S(t) = S_0$= constant and thus the change in entropy is zero.

4 Cosmological Solutions

In order to solve the field equations we consider the exact power-law expansion of the universe

$$a = a_0 t^{\alpha}, \tag{22}$$

where a_0 is a positive constant and α is a positive real number which determines the expansion of scale factor in different phases of the evolution of the universe, e.g., $\alpha = 1$ determines the marginal inflation ($a \propto t$), $\alpha = 1/2$ for radiation-dominated phase ($a \propto t^{1/2}$), $\alpha = 2/3$ for

matter-dominated phase ($a \propto t^{2/3}$) and $\alpha = 4/3$ determines the accelerated phase ($a \propto t^{4/3}$) of the universe. Therefore, the exact power-law solution (22) describes the expansion of the scale factor in decelerating and accelerating universes. The history of the universe may also be described by relation (8) where EoS parameter ω is related to α as

$$\omega = \frac{2}{3\alpha} - 1. \tag{23}$$

Therefore, $\alpha = 1, 1/2, 2/3$ and $4/3$,implies that $\omega = -1/3, 1/3, 0$ and $-1/2$ for marginal inflation, radiation, matter and quintessence models, respectively. The deceleration parameter q is related to H as

$$q = -\frac{a\ddot{a}}{\dot{a}^2} = -\frac{\ddot{a}}{aH^2}, \tag{24}$$

where $H = \dot{a}/a$ is the Hubble parameter. For any physically relevant model, the Hubble parameter H and deceleration parameter q are some of the most important observational quantities in cosmology. The first quantity sets the present time scale of the evolution while the second one is telling us that present stage is speeding up instead of slowing down as expected before the Supernovae type Ia observations. The sign of q determines whether or not the model inflate. For $q \geq 0$ the universe is expanding with decelerated/marginal inflation rate whereas $q < 0$ determines the accelerating universe. From (22) and (24) we get

$$q = \frac{(1-\alpha)}{\alpha}, \tag{25}$$

which is constant. It is observed that $q \geq 0$ for $0 < \alpha \leq 1$ and $q < 0$ for $\alpha > 1$. Therefore, the deceleration parameter shows signature flipping on the constraint on α. Now a days some authors (see, Pradhan et al. [62]) have assumed some other suitable form of scale factor which gives the time-dependent deceleration parameter. But, How much realistic is such type of model? A detail investigate is needed. It may be noted that though the current observations of SNe Ia and CMB favor accelerating models ($q < 0$), it can not be altogether ruled out the decelerating ones which are also consistent with these observations [63].

A relations between q and ω is given by

$$q = \frac{(1+3\omega)}{2}. \tag{26}$$

Therefore, $\omega = -1/3, 1/3, 0$ and $-1/2$ gives $q = 0, 1, 1/2$ and -0.25, respectively. The Hubble parameter is given by

$$H = \frac{\alpha}{t}, \qquad q = \frac{1}{Ht} - 1. \tag{27}$$

We now look the cosmological solutions with and without particle creation, respectively in the following subsections.

4.1 Particle Creation

In this case $\psi(t) > 0$, using (8), Eq. (7) becomes

$$\dot{\rho} + (m+1)(1+\omega)\rho H = -\left(\frac{\dot{G}}{G}\rho + \frac{\dot{\Lambda}}{8\pi G}\right). \tag{28}$$

Using (5) and (27), Eq. (28) gives the energy density and cosmological constant, respectively as

$$\rho = \frac{m\alpha}{8\pi(1+\omega)G}\frac{1}{t^2}, \tag{29}$$

$$\Lambda = \frac{m\alpha}{2(1+\omega)}\big[\alpha(m+1)(1+\omega)-2\big]\frac{1}{t^2}. \tag{30}$$

The above and thereafter solutions are valid for $\omega \neq -1$. Substituting (29) and (30) in to (6), we find the equation is identically satisfy for any arbitrary $G(t)$. The energy density and cosmological constant are decreasing functions of time and infinite at $t = 0$ and tend to zero as $t \to \infty$. It is noted that the above solutions are permitted with an arbitrary gravitational constant G. The form (30) of the time variation of the cosmological constant is similar to that proposed by several authors [19, 30, 32, 35, 57, 64].

The vacuum energy density (ρ_Λ) is given by

$$\rho_\Lambda(t) = \frac{\Lambda}{8\pi G} = \frac{m\alpha}{16\pi G(1+\omega)}\big[\alpha(m+1)(1+\omega)-2\big]\frac{1}{t^2}. \tag{31}$$

The density parameter of matter and vacuum contributions are respectively obtained as

$$\Omega_{matt.} = \frac{\rho}{\rho_c} = \frac{2}{(m+1)(1+\omega)\alpha}, \tag{32}$$

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} = \left[1 - \frac{2}{(m+1)(1+\omega)\alpha}\right], \tag{33}$$

where $\rho_c = \frac{m(m+1)H^2}{16\pi G}$. The density parameter of matter and vacuum in terms of q are given by

$$\Omega_{matt.} = \frac{2}{(m+1)(1+\omega)}q + \frac{2}{(m+1)(1+\omega)}, \tag{34}$$

$$\Omega_\Lambda = \left(1 - \frac{2}{(m+1)(1+\omega)}\right) - \frac{2}{(m+1)(1+\omega)}q. \tag{35}$$

Therefore, the total density parameter Ω_T is calculated as

$$\Omega_T = \Omega_{matt.} + \Omega_\Lambda = 1. \tag{36}$$

The solution is similar to the usual Einstein gravity and is favored by the inflationary scenario. According to high redshift supernovae and CMB, the preliminary results suggest that the universe may be accelerating with a dominant contribution to its energy density coming in the form of cosmological constant. Recent measurement have given wide range of values of $\Omega_{matt.0}$ and $\Omega_{\Lambda 0}$, the present cosmic matter and vacuum energy density parameters. Observations of SNe Ia combined with the total energy density constraints from CMB [65] and combined gravitating lens and stellar dynamical analysis [66] lead to $\Omega_{matt.0} \sim 0.3$ and $\Omega_{\Lambda 0} \sim 0.7$.

We now calculate the particle production rate, which is given by

$$\psi(t) = \frac{1}{(1+\omega)}\big[(m+1)(1+\omega)\alpha - 2\big]\frac{1}{t}, \quad \omega \neq -1. \tag{37}$$

Thus, it is observed that the particle creation rate is decreasing function of time. At the end the particle number density obeys the approximate relation $\dot{n} + (m+1)Hn \approx 0$, which leads to an approximate particle number density $n \sim a^{-(m+1)}$.

The particle number density and entropy production during the particle creation era are obtained as

Table 1 Values of various cosmological parameters for radiation and matter-dominated phases with particle creation in $(m+2)$-dimensions

Values	$\omega = 1/3, \alpha = 1/2$	$\omega = 0, \alpha = 2/3$
$a(t)$	$a_0 t^{1/2}$	$a_0 t^{2/3}$
$\rho(t)$	$\frac{3m}{64\pi G} t^{-2}$	$\frac{m}{12\pi G} t^{-2}$
$\Lambda(t)$	$\frac{m(m-2)}{8} t^{-2}$	$\frac{2m(m-2)}{9} t^{-2}$
q	1	$\frac{1}{2}$
$\Omega_{matt.}$	$\frac{3}{(m+1)}$	$\frac{3}{(m+1)}$
Ω_Λ	$\frac{(m-2)}{(m+1)}$	$\frac{(m-2)}{(m+1)}$
$\psi(t)$	$\frac{(m-2)}{2} t^{-1}$	$\frac{2(m-2)}{3} t^{-1}$
S	$S_0 t^{\frac{(m-2)}{2}}$	$S_0 t^{\frac{2(m-2)}{3}}$
n	$\sim t^{-3/2}$	$\sim t^{-2}$
T	$\sim t^{-\frac{1}{2}}$	T_0

$$n = n_0 t^{-\frac{2}{(1+\omega)}}, \tag{38}$$

$$S = S_0 t^{\frac{(m+1)(1+\omega)\alpha - 2}{(1+\omega)}}. \tag{39}$$

From (39), it follows that $S \sim nV$, where $V = a^{(m+1)}$ is the comoving volume scale factor. The total mass M of matter inside the comoving volume determined by the scale factor V is $M \sim \rho a^{(m+1)}$, which is not a constant but also has a power -law time variation.

$$M \sim t^{(m+1)\alpha - 2}. \tag{40}$$

From (14) and (29), we obtain

$$T = T_0 \left(\frac{m\alpha}{8\pi(1+\omega)G} \right)^{\frac{\omega}{(1+\omega)}} t^{-\frac{2\omega}{(1+\omega)}}. \tag{41}$$

We observe that the particle number density and temperature are independent of dimensions of space-time and are decreasing functions of time where as the entropy and mass depend on dimensions.

For different phases of the universe, namely, radiation ($\omega = 1/3$) and matter-dominated ($\omega = 0$), the values of scale factor $a(t)$, matter energy density $\rho(t)$, the vacuum energy density $\Lambda(t)$, deceleration parameter q, density parameter for matter $\Omega_{matt.}$, density parameter for vacuum Ω_Λ, rate of particle creation $\psi(t)$, entropy production S, particle number density n and temperature T for $(m+2)$ dimensional space-times are given in Table 1.

From Table 1, it is observed that $\rho(t)$, $\Lambda(t)$, $\psi(t)$ and n decrease with time whereas the entropy production increases with time in both the phases. The temperature decreases with time during the radiation phase and it is constant in matter-dominated phase. The total density parameter $\Omega_T = 1$ during these two phases of the evolution of the universe.

It may also be obtained the various cosmological parameters from solutions (29)–(41) for the quintessence (accelerated) phase where $\omega = -1/2$, $\alpha = 4/3$ and $q = -0.25$.

4.2 No Particle Creation

We now consider the case in which both G and Λ vary with time in such a way that the Eqs. (20) and (21) hold. Equation (21) represents a coupling between vacuum and gravity and that

the vacuum decays to strengthen the gravitating interaction that will induce an acceleration of the expansion of the universe. In the case of no particle creation, (20) gives

$$\rho = \frac{\rho_0}{a^{(m+1)(1+\omega)}}, \tag{42}$$

where ρ_0 is a positive constant. Using (22) into (42) we get

$$\rho = At^{-\alpha(m+1)(1+\omega)}, \tag{43}$$

where $A = \rho_0/a_0^{(m+1)(1+\omega)} > 0$ is another constant. Using (5), (21) and (43), one gets the evolution of the cosmological and gravitational constants, which are given by

$$\Lambda(t) = \frac{m\alpha}{2(1+\omega)}\big[\alpha(m+1)(1+\omega) - 2\big]\frac{1}{t^2}, \quad (\omega \neq -1). \tag{44}$$

$$G(t) = \frac{m\alpha}{8\pi A(1+\omega)}t^{(m+1)(1+\omega)\alpha-2}, \quad (\omega \neq -1). \tag{45}$$

The particle number density is given by

$$n = n_1 t^{-4(m+1)}, \tag{46}$$

where n_1 is a constant of integration. From the above equations, it is observed that the energy density and cosmological term are decreasing while the gravitational constant G is increasing during the expansion of the universe for $(m+1)(1+\omega)\alpha > 2$. In most variable G cosmologies [67], G is a decreasing function of time. But the possibility of an increasing G has also been suggested by several authors [68–71]. The density parameter of matter and vacuum contributions are given by (32) and (33).

The rate of change of G is given by

$$\frac{\dot{G}}{G} = \big[(m+1)(1+\omega)\alpha - 2\big]t^{-1}. \tag{47}$$

It is notice that the rate of change of G decreases linearly with time and independent on the space-time dimension.

For radiation and matter-dominated phases, the values of the various parameters, like $a(t)$, ρ, $\Lambda(t)$, $G(t)$, q, $\Omega_{matt.}$, Ω_Λ, $\psi(t)$, S and n for $(m+2)$-dimensional space times are given in Table 2.

Table 2 Values of various cosmological parameters for radiation and matter-dominated phases without particle creation in $(m+2)$-dimensions

Values	$\omega = 1/3, \alpha = 1/2$	$\omega = 0, \alpha = 2/3$
$a(t)$	$a_0 t^{1/2}$	$a_0 t^{2/3}$
$\rho(t)$	$At^{-\frac{2(m+1)}{3}}$	$At^{-\frac{2(m+1)}{3}}$
$\Lambda(t)$	$\frac{m(m-2)}{8}t^{-2}$	$\frac{2m(m-2)}{9}t^{-2}$
$G(t)$	$\frac{3m}{64\pi A}t^{\frac{2(m-2)}{3}}$	$\frac{3m}{12\pi A}t^{\frac{2(m-2)}{3}}$
q	1	$\frac{1}{2}$
$\Omega_{matt.}$	$\frac{3}{(m+1)}$	$\frac{3}{(m+1)}$
Ω_Λ	$\frac{(m-2)}{(m+1)}$	$\frac{(m-2)}{(m+1)}$
$\psi(t)$	0	0
S	*const.*	*const.*
n	$\sim t^{-\frac{(m+1)}{2}}$	$\sim t^{-\frac{2(m+1)}{3}}$

It is observed from the Table 2 that the cosmological parameters are dependent on dimensions of space time in both phases of the universe. The rate of production $\psi(t)$ is zero, entropy is constant and G increases with time in both phases of the evolution of the universe. The solution reduces to the familiar four dimensional space time for $m = 2$ with $\Lambda = 0$. The present value of matter density parameter gives $\Omega_T = 1$, which is inconsistent with many observational tests.

It may also be obtained the various cosmological parameters from solutions (43)–(46) for the quintessence (accelerated) phase where $\omega = -1/2$, $\alpha = 4/3$ and $q = -0.25$.

5 Particular Solution: Kaluza-Klein Type Cosmology

We now study the evolution of particle creation for five dimensional ($m = 3$) Kaluza-Klein FRW type cosmology. For $m = 3$, the solutions are listed in Tables 3 and 4 with and without particle creation, respectively for radiation and matter-dominated phases.

The other cosmological parameters have same value as mentioned in Tables 1 and 2 for this space-time. During no particle creation the gravitational parameter G increases with cosmic time t in both phases. It may also be obtained the various cosmological parameters from solutions for the quintessence (accelerated) phase where $\omega = -1/2$, $\alpha = 4/3$ and $q = -0.25$.

Table 3 Values of various cosmological parameters for radiation and matter-dominated phases with particle creation in 5-dimensions

Values	$\omega = 1/3, \alpha = 1/2$	$\omega = 0, \alpha = 2/3$
$\rho(t)$	$\frac{9}{64\pi G}t^{-2}$	$\frac{1}{4\pi G}t^{-2}$
$\Lambda(t)$	$\frac{3}{8}t^{-2}$	$\frac{2}{3}t^{-2}$
$\Omega_{matt.}$	$\frac{3}{4}$	$\frac{3}{4}$
Ω_Λ	$\frac{1}{4}$	$\frac{1}{4}$
$\psi(t)$	$\frac{1}{2}t^{-1}$	$\frac{2}{3}t^{-1}$
S	$S_0 t^{\frac{1}{2}}$	$S_0 t^{\frac{2}{3}}$
n	$\sim t^{-2}$	$\sim t^{-2}$

Table 4 Values of various cosmological parameters for radiation and matter-dominated phases without particle creation in 5-dimensions

Values	$\omega = 1/3, \alpha = 1/2$	$\omega = 0, \alpha = 2/3$
$\rho(t)$	$At^{-\frac{8}{3}}$	$At^{-\frac{8}{3}}$
$\Lambda(t)$	$\frac{3}{8}t^{-2}$	$\frac{2}{3}t^{-2}$
$G(t)$	$\frac{9}{64\pi A}t^{\frac{2}{3}}$	$\frac{3}{4\pi A}t^{\frac{2}{3}}$
$\Omega_{matt.}$	$\frac{3}{4}$	$\frac{3}{4}$
Ω_Λ	$\frac{1}{4}$	$\frac{1}{4}$
$\psi(t)$	0	0
n	$\sim t^{-2}$	$\sim t^{-\frac{8}{3}}$

6 Conclusion

In this paper we have studied the evolution of particle creation with varying gravitational constant G and cosmological term Λ in higher dimensional space-times. We have assumed an exact power-law solution for the scale factor of the universe, which leads to decelerating and accelerating phases of the universe. For expanding and decelerating universe the parameter α is constrained to the condition $0 < \alpha \leq 1$ and for accelerating universe it is $\alpha > 1$. The cosmological consequences of the model are shown to be very interesting. The variation of various physical and geometrical parameters have been obtained for different phases of the universe. The solutions are also relevant for quintessence model where $\alpha = 4/3$, $\omega = -1/2$ and $q = -0.25$. All the solutions are valid for $\omega \neq -1$. The analytical expressions of the cosmological parameters indicate a clear dependence of dynamical properties on dimensionality of the space-time. Many models in the literature can be retrieved from this model with choice of the dimensions, for example, 5-dimension Kaluza-Klein FRW type models as discussed in Sect. 5. A possible explanation for the small value of Λ can be based on the idea that the vacuum energy density is not constant, but decays as the universe expands. It is interesting to note that the gravitational constant G is increasing function of time in five or higher dimensional space-time in the case of no particle creation. In 5-dimension Kaluza-Klein FRW type models, Λ varies inversely as square of cosmic time t during particle creation, which is good agreement with observations and matches its natural units. We have also studied the entropy production, particle number density and temperature in these models for radiation and matter-dominated phases which have been shown in Tables 1–4.

The present model predicts the possibility that the particle creation associated with $\Lambda(t)$ and $G(t)$ can occur in the present day universe, as considered by Dirac [24, 25]. Therefore, we have interpreted the Einstein field equations with variable G and Λ in higher dimensional space-times describing particle creation. A decaying cosmological constant and a constant G always produces the particle creation whereas a decaying Λ and varying G or $\Lambda = 0$ with constant G produces no particle creation during the evolution of the universe.

Acknowledgements The authors express their sincere thanks to the referees for their constructive suggestions to improve the manuscript.

References

1. Bergman, P.G.: Int. J. Theor. Phys. **1**, 25 (1968)
2. Linde, A.O.: JETP Lett. **19**, 183 (1974)
3. Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999)
4. Riess, A.G., et al.: Astron. J. **116**, 1009 (1998)
5. Overduin, J.M., Cooperstock, F.I.: Phys. Rev. D **58**, 043506 (1998)
6. Berman, M.S., Som, M.M., Gomide, F.M.: Gen. Relativ. Gravit. **21**, 287 (1989)
7. Berman, M.S.: Int. J. Theor. Phys. **29**, 567 (1990)
8. Berman, M.S.: Nuovo Cimento B **105**, 239 (1990)
9. Bertolami, O.: Nuovo Cimento B **93**, 36 (1986)
10. Özer, M., Taha, M.O.: Phys. Lett. B **171**, 363 (1986)
11. Özer, M., Taha, M.O.: Nucl. Phys. B **287**, 776 (1987)
12. Weinberg, S.: Rev. Mod. Phys. **61**, 1 (1989)
13. Abdel-Rahaman, A.-M.M.: Phys. Rev. D **45**, 3497 (1992)
14. Chen, W., Wu, Y.S.: Phys. Rev. D **42**, 4728 (1992)
15. Freedman, W.L.: Phys. Scr. T **85**, 37 (2000)
16. Dolgov, A.D.: Phys. Rev. D **55**, 5881 (1997)
17. Vishwakarma, R.G.: Gen. Relativ. Gravit. **197**, 33 (2001)
18. Vishwakarma, R.G.: Class. Quantum Gravity **19**, 4747 (2002)

19. Beesham, A.: Phys. Rev. D **48**, 3539 (1999)
20. Peebles, P.J.E., Ratra, B.: Astrophys. J. **325**, L17 (1988)
21. Peebles, P.J.E., Ratra, B.: Phys. Rev. D **37**, 3406 (1988)
22. Sahni, V., Starobinsky, A.: Int. J. Mod. Phys. D **9**, 373 (2000)
23. Padmanabhan, T.: Phys. Rep. **380**, 235 (2003)
24. Dirac, P.A.M.: Nature **139**, 323 (1937)
25. Dirac, P.A.M.: Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci. **165**, 199 (1938)
26. Brans, C., Dicke, R.H.: Phys. Rev. **124**, 925 (1961)
27. Lau, Y.K.: Aust. J. Phys. **38**, 547 (1985)
28. Lau, Y.K., Prokhovnik, S.J.: Aust. J. Phys. **39**, 339 (1986)
29. Beesham, A.: Nuovo Cimento B **96**, 17 (1986)
30. Kalligas, D., Wesson, P., Everitt, C.W.F.: Gen. Relativ. Gravit. **24**, 351 (1992)
31. Abdel-Rahaman, A.-M.M.: Gen. Relativ. Gravit. **22**, 655 (1990)
32. Abdussattar Vishwakarma, R.G.: Class. Quantum Gravity **14**, 945 (1997)
33. Arbab, A.I.: Gen. Relativ. Gravit. **29**, 391 (1998)
34. Singh, C.P., Kumar, S., Pradhan, A.: Class. Quantum Gravity **24**, 455 (2007)
35. Beesham, A.: Gen. Relativ. Gravit. **26**, 159 (1994)
36. Kalligas, D., Wesson, P., Everitt, C.W.F.: Gen. Relativ. Gravit. **27**, 645 (1995)
37. Pradhan, A., Chakraborty, I.: Gravit. Cosmol. **7**, 55 (2001)
38. Saha, B.: Mod. Phys. Lett. A **16**, 1287 (2001)
39. Saha, B.: Astrophys. Space Sci. **302**, 83 (2006)
40. Singh, G.P., Kotambar, S., Srivastava, D., Pradhan, A.: Rom. J. Phys. **53**, 607 (2008)
41. Weinberg, S., et al.: Physics in Higher Dimensions. World Scientific, Singapore (1986)
42. Kaluza, T.: Sitz. Preuss. Akad. Wiss. Phys.-Math. K **1**, 966 (1921)
43. Klein, O.: Z. Phys. **37**, 895 (1926)
44. Witten, E.: Phys. Lett. B **144**, 351 (1984)
45. Appelquist, T., Chodos, A., Freund, P.G.O.: Modern Kaluza-Klein Theories. Addison-Wesley, Reading (1987)
46. Overduin, J.M., Wesson, P.S.: Phys. Rep. **283**, 303 (1997)
47. Chatterjee, S., Bhui, B.: Mon. Not. R. Astron. Soc. **247**, 57 (1990)
48. Banerjee, A., Bhui, B.: Astron. Astrophys. **232**, 305 (1990)
49. Singh, G.P., Kotambkar, S.: Gen. Relativ. Gravit. **33**, 621 (2001)
50. Singh, G.P., Kotambkar, S.: Gravit. Cosmol. **9**, 206 (2003)
51. Singh, G.P., Kotambkar, S., Pradhan, A.: Int. J. Mod. Phys. D **12**, 853 (2003)
52. Khadekar, G.S., Kamdi, V., Pradhan, A., Otarod, S.: Astrophys. Space Sci. **310**, 141 (2007)
53. Prigogine, I., Geheniau, J., Gunzig, E., Nardone, P.: Proc. Natl. Acad. Sci. USA **85**, 7428 (1988)
54. Prigogine, I., Geheniau, J., Gunzig, E., Nardone, P.: Gen. Relativ. Gravit. **21**, 767 (1989)
55. Sudharsan, R., Johri, V.B.: Gen. Relativ. Gravit. **26**, 41 (1994)
56. Harko, T., Mak, M.K.: Astrophys. Space Sci. **253**, 161 (1997)
57. Harko, T., Mak, M.K.: Gen. Relativ. Gravit. **31**, 849 (1999)
58. Harko, T., Mak, M.K.: Class. Quantum Gravity **16**, 2741 (1999)
59. Tegmark, M., et al.: Astrophys. J. **606**, 70 (2004)
60. Weinberg, S.: Gravitation and Cosmology. Wiley, New York (1972)
61. Maartens, R.: Class. Quantum Gravity **12**, 1455 (1995)
62. Pradhan, A., Jaiswal, R., Jotania, K., Khare, R.K.: Astrophys. Space Sci. **337**, 401 (2012)
63. Vishwakarma, R.G.: Mon. Not. R. Astron. Soc. **345**, 545 (2003)
64. Berman, M.S.: Phys. Rev. D **43**, 1075 (1991)
65. Rebelo, R.: Nucl. Phys. B, Proc. Suppl. **114**, 3 (2003)
66. Koopmans, L.V.E., et al.: Astrophys. J. **599**, 70 (2003)
67. Abdel Rahaman, A.-M.M.: Gen. Relativ. Gravit. **22**, 665 (1990)
68. Weinberg, S.: Gravitation and Cosmology. Wiley, New York (1971)
69. Abdel Rahaman, A.-M.M.: Phys. Rev. D **45**, 3497 (1992)
70. Levit, L.S.: Lett. Nuovo Cimento **29**, 23 (1980)
71. Sistero, R.F.: Gen. Relativ. Gravit. **23**, 1265 (1991)

Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol 3, No.7, 2012

Realization Of An 8-bit Pipelined Microprocessor in Verilog HDL

Jayant Chowdhary* (Corresponding Author) Vivek Garg Tushar Negi Shreya Jain
Delhi Technological University,Shahbad Daulatpur,Main Bawana Road,Delhi-42
*Email of the corresponding author: jayantc11@gmail.com

Abstract
Pipelining is a technique of decomposing a sequential process into sub-operations, with each sub process being divided segment that operates concurrently with all other segments. A pipeline may be visualized as a collection of processing segments through which binary information flows. Each segment performs partial processing segments dictated by the way the task is partitioned. The result obtained in one segment is transferred to subsequent segments in each step. The final result is obtained after the data has passed through all segments.This paper develops a code for the implementation of an 8-Bit microprocessor which implements instruction pipelining. After synthesis, an FPGA realization may be obtained . Simulation using Xilinx and ModelSim also produces favourable results which showcase the speedup (in terms of time) to carry out a program as compared to a non-pipelined version of this microprocessor.
Keywords: Pipelining, Segments,sysnthesis,realization,FPGA,microprocessor

1.Introduction

Instruction pipelining
An instruction pipeline reads consecutive instructions from memory while previous instructions are being executed in other segments. This causes the instruction 'fetch' and 'execute' phases to overlap and perform simultaneous operations.
An instruction can generally be decomposed into the following steps:

1. **FI- Fetch instruction**: In this segment an instruction is fetched from memory.
2. **DA- Decode the instruction and calculate the effective address**: This instruction gets its input from the FI segment and the instruction is decoded. This step does not require a clock cycle. The operation which requires a clock cycle is the calculation of effective address.
3. **FO- Fetch operand**: In this step the operand is fetched from the memory.
4. **EX- Execute and store**: In this step the instruction is executed and the result is also stored in an appropriate memory location.

A Space time diagram for an instruction pipeline can be made as follows:

Instruction	Step1	Step2	Step3	Step4	Step5	Step6	Step7
1	FI	DA	FO	EX			
2		FI	DA	FO	EX		
3			FI	DA	FO	EX	
4				FI	DA	FO	EX

As we can see each segment is simultaneously busy processing some part of an instruction.

2. Processor Realization

CPU ORGANIZATION AND LAYOUT :

1) 8 bit cpu
2) Architecture: VON NEUMAN
3) Behavioural modelling only
4) Different Program Memory and Data Memory
5) 16 words of Program memory
6) 4096 words of data memory

Register Organization:

1)Program Counter: PC – 6 bits
2)Instruction register: IR - 16 bits
3)Address register: AR - 12 bits

4)Temporary address storage: ad - 6 bits (6 LSBs of AR)
5)Memory check register: memocheck -1 bit (for register reference and memory reference istructions)
6)Current state register: current_state – 2 bits (to check the present state- fetch/decode/execute/store)
7)Next state register: next_state- 1 bit (to hold the value of next operation to be carried out- fetch/decode/execute/return)
8)OPcode register: opcode – 3 bits
9)Instruct register: instruct -4 bits (in case instruction type is arithmetic/logical then this specifies type of arithmetic/logical op to be carried out)
10)Registers RA,RB,rstore - 4 bits (to hold the address of source and destination registers respectively)
11)Temporary register: W- 8 bits (to hold the value of the end result of any arithmetic/ logical operation and incase of data MOV operation it holds the immediate value of the data to be transferred)

3.Instruction Format:

1 bit	3 bits	4 bits	4 bits	4 bits
I	Opcode	Instruct	RA(index)	RB(index)

I : 1- Memory reference
0-Register reference
Opcode: operational code which tells us what type of operation is to be carried out on the data present at source registers ,given by the index in RA RB. Opcode list:
000 - HLT: The cpu goes into an infinite loop terminating the program
001 - MVI: Move immediate value to destination register register(indicated by rstore), {RA,RB} is the 8-bit concatenated word used as immediate data, rstore: address of destination register
000 (I=1) - STA: store contents of R0 (virtual accumulator) in the memory address given by ad;
001 (I=1) - LDA: load the contents of memory specified by ad into the accumulator (R0)
010 (I=1) - JMP: jump to the location specified by 6 bits in ad
010 – operation: opcode for all arithmetic /logical instructions further classified by 4 bits instruct field
MOV=0000; Move Contents Of Register[RB] To Register[RA]
ADD=0001; Add Contents Of Register[RB] with Contents Of Register[RA] & save result in R[A]
ADC=0010; ADD With Carry, Add contents of register[RB] with contents Of Register[RA] & save result In R[A]
SBB=0011; Subtract With Borrow
SUB=0100; Subtract contents of Register[RB] with contents of Register[RA] & save result in R[A]
INC=0110;Increment contents of Register[RA] & save result in R[A]
DEC=0111; Decrement contents of Register[RA] & save result in R[A]
AND=1001;Logically AND contents of Register[RB] with contents of Register[RA] & save result in R[A]
OR=1010; Logically OR contents of Register[RB] with contents of Register[RA] & save result in R[A]
XOR=1011; Logically XOR contents of Register[RB] with contents of Register[RA] & save result in R[A]
CMP=1000; Complement contents of Register[RA]
SHR=1100; Shift right contents of Register[RA] by 1 bit
SHL=1101; Shift left contents of Register[RA] by 1 bit

4.Program Code

```
module micropipeline(clk,rst);
input clk;
input rst;
reg [0:3]PC;
reg [0:15] imem [0:15];      //instruction memory
reg [0:15] IR;               // instruction register
reg [0:7] datareg;           //data register
reg [0:15] IRW;
reg [0:7] mem [0:255];
reg memocheck;
reg [0:1] current_state;   //stages
reg [0:1] next_state;      //for keeping track
reg [0:11] AR;             //Address register
reg [0:2] opcode;
reg [0:3] instruct;
reg [0:3] RA,RB,rstore;
```



```
reg [0:7] regfile [0:15];     //16 register(R0-R7)
reg [0:7] W;                              //temporary register
reg [0:2] count;
reg fetch,decode,execute,store;
reg [0:100]TR;
reg moveimm,movein,operation,sta,halt,halt1;
reg loadmemory,memorystore;
reg load;
wire carry;
wire [0:7] temp;
reg carry1;
reg [0:7] tink;
reg car;
initial
begin
PC=0;
TR=0;
fetch=1;
decode=0;
execute=0;
store=0;
moveimm=0;
movein=0;
operation=0;
halt=0;
sta=0;
loadmemory=1;
memorystore=1;
load=0;
next_state=4'b1000;
halt1=0;
IRW=0;
datareg[0]=0;
          imem[0]={16'b0001000111001111};
imem[1]={16'b0001001011011000};
imem[2]={16'b0011000000010010};
imem[4]={16'b0000000000000000};
imem[3]={16'b0000000000000000};
imem[5]={16'b0000000000000000};
imem[6]={16'b0000000000000000};
imem[7]={16'b0000000000000000};
end
always@(clk or rst)
begin
TR<=TR+1;
if(rst)
begin
PC<=0;
TR<=0;
fetch<=1'b1;
decode<=1'b0;
execute<=1'b0;
store<=1'b0;
sta<=1'b0;
halt<=1'b0;
end
else
      begin
if(store)
            begin
if(loadmemory&memorystore)
                        begin
regfile[rstore]<=W;
end
if(halt1)
                        begin
fetch<=#1 1'b0;
                              decode<=#1 1'b0;
                              execute<=#1 1'b0;
store<=#1 1'b0;
end
end
if(execute)
            begin
if(halt)
                  begin
halt1<=1'b1;
end

if(moveimm)
                  begin
W<=IRW[8:15];
                              rstore<=IRW[4:7];
                              moveimm<=1'b0;
end
if(movein)
begin
W<=datareg[IRW[12:15]];
rstore<=IRW[8:11];
movein<=1'b0;
end
if(operation)
begin
case(IRW[4:7])
4'b0000: begin
W<=regfile[RA]+regfile[RB];
end

4'b0001: begin

W<=regfile[RA]+regfile[RB]+1;
end
```



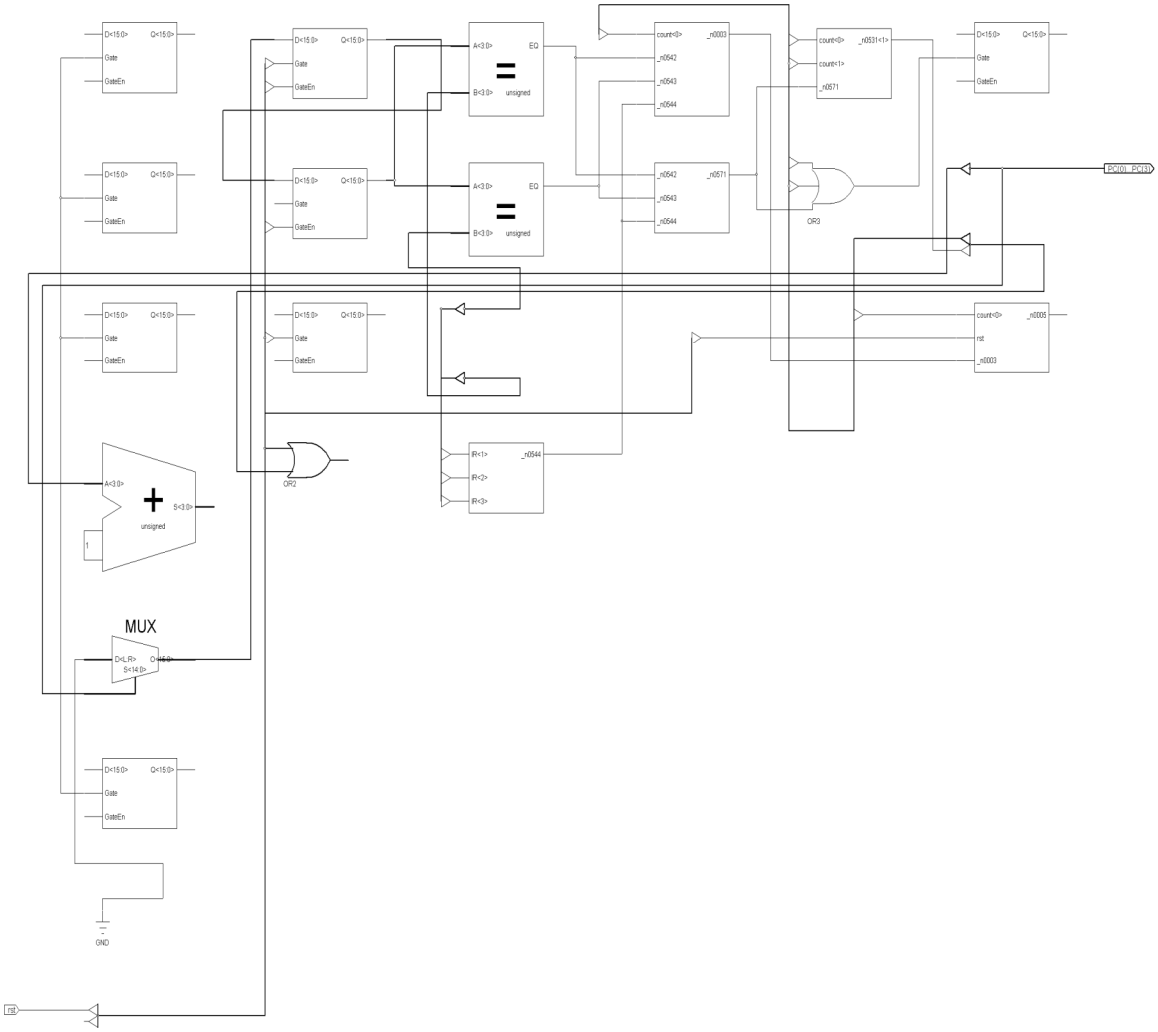
```
4’b0010: begin
W<=regfile[RA]-regfile[RB]-1;
end

4'b0011:begin
W<=regfile[RA]-regfile[RB];
end

4'b0100: begin
W<=regfile[RA]+1;
end

4'b0101: begin
W<=regfile[RA]-1;
end
4'b0110: begin
W<=regfile[RB];
end

4'b1000: begin


W<=regfile[RA]&regfile[RB];
end

4'b1001: begin
W<=regfile[RA]|regfile[RB];end
4'b1010: begin
W<=regfile[RA]^regfile[RB];
end

4'b1011: begin
W<=regfile[RA];          // has to include ~
End
                         endcase
// alu operation
rstore<=4'b0000;
if(count==2'b00)
operation<=1'b0;
end
if(sta)
begin
memorystore<=1'b0;
mem[IRW[8:15]]<=regfile[4'b0000];
sta<=1'b0;
end
if(load)
begin
loadmemory<=1'b0;
regfile[4'b0000]<=mem[IRW[8:15]];
load<=1'b0;
end
store<=1'b1;
end
if(decode)
begin
            IRW<=IR;
case(IR[0])
0:begin
case(IR[1:3])
3'b000: begin
halt<=1'b1;
end
3'b001:
begin
moveimm<=1'b1;
end          //the instruction is mvi
3'b010: begin
movein<=1'b1;
end
3'b011:begin
operation<=1;
RA<=IR[8:11];
RB<=IR[12:15];
end
3'b100:begin
sta<=1'b1;
end
default:begin
end
endcase
end
1:begin
load<=1'b1;
end
endcase
execute<=1'b1;
end
if(fetch)
begin
IR<=imem[PC];
PC<=PC+1;
decode<=1'b1;
end
end
if((IRW[4:7]==IR[8:11]|IRW[4:7]==IR[12:15])&IR[1:
3]==3'b011)
begin
count<=2'b01;
fetch<=#2 1'b0;
decode<=#2 1'b0;
```



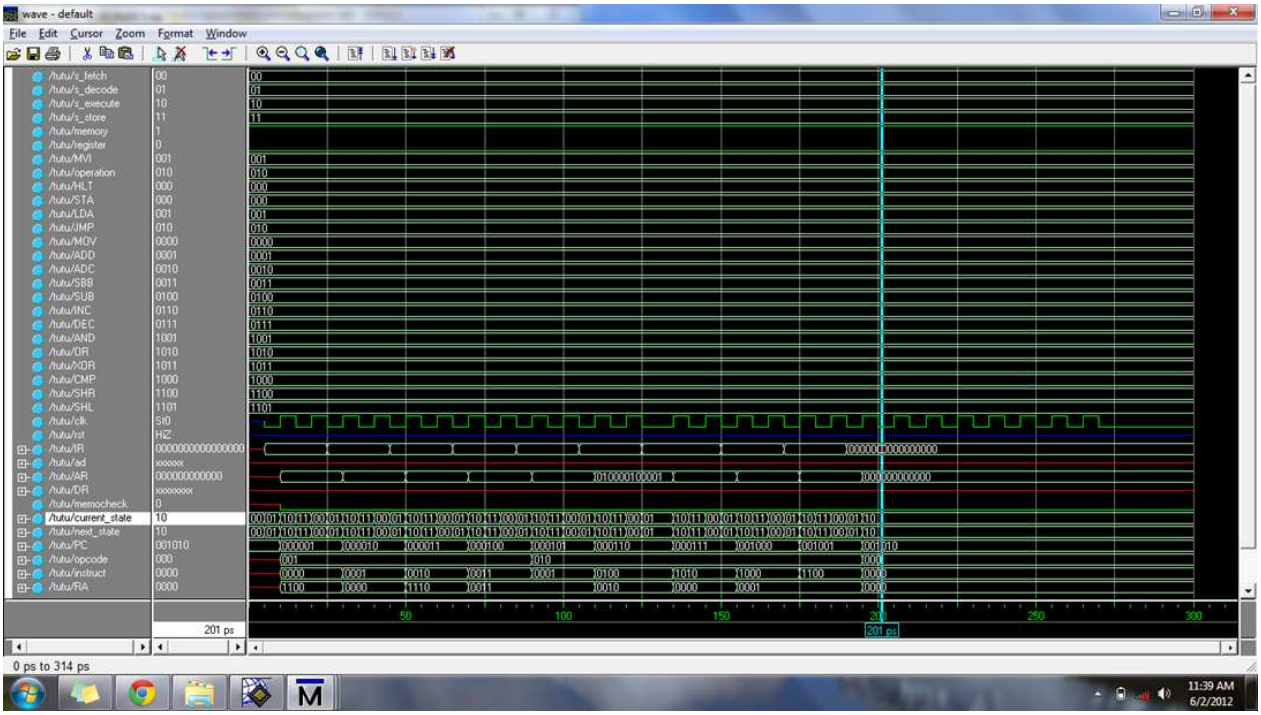
```
end
if(count==2'b01)
begin
count<=2'b10;
execute<=1'b0;
end
if(count==2'b10)
begin
count<=2'b00;
fetch<=1'b1;
decode<=1'b1;
end
end
endmodule
```

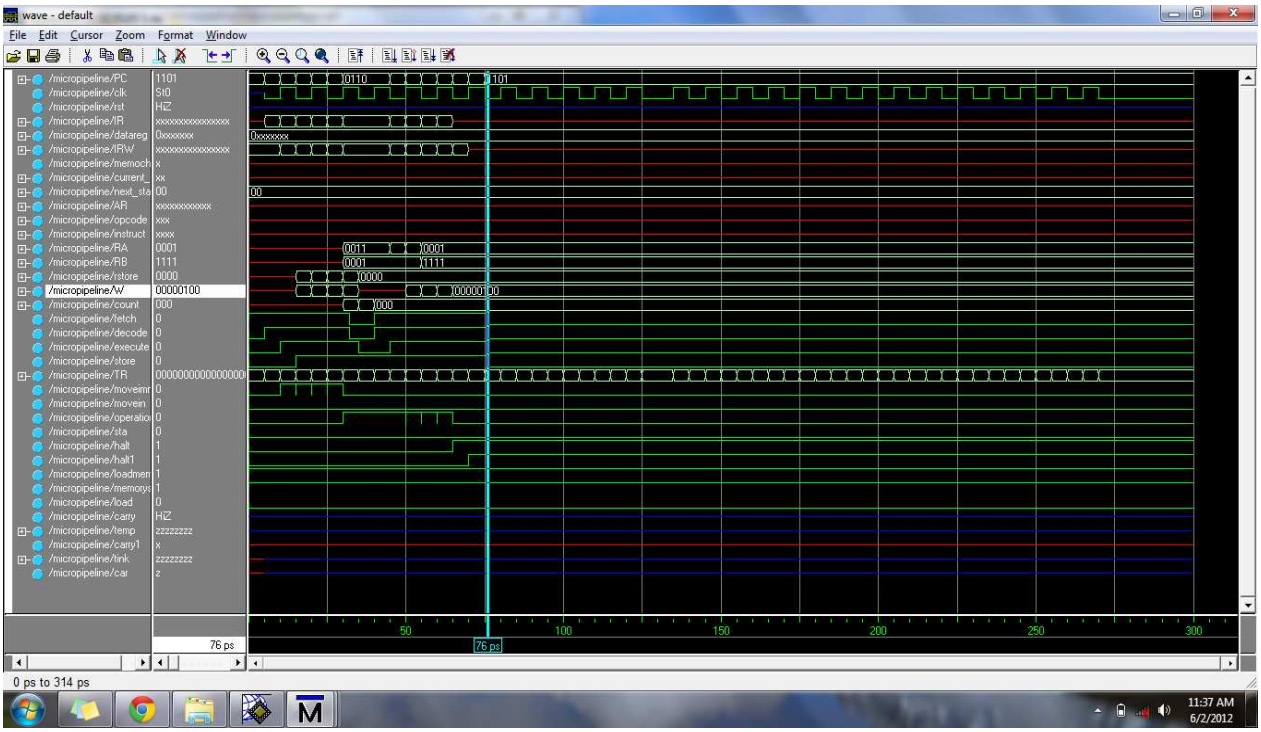
5.Results

The following was the result of synthesis of the code on Xilinx and on comparing the timing diagram with that of a non-pipelined microprocessor running the same program we find that there is a considerable speedup of 125ps.The synthesis and timing diagrams of pipelined and non pipelined microprocessors are shown below:



Given above is the timing diagram of a non pipelined microprocessor built on specs to the pipelined one. A 4 instruction program takes 201 ps as shown to get executed. Given below is the timing diagram of the same program





using the pipelined microprocessor presented in this paper. It can be seen the program execution time has reduced considerably to 76ps.

5.Conclusion

This paper presents a basic pipelined microprocessor which has been written in Verilog HDL. This is simulation is unconventional in a way such that it showcases the beauty which goes into implementing pipelining of instructions in a microprocessor without going into too many complexities of the same.Moreover,this paper may be very effectively used as a tool in the field of education to introduce students to computer simulations of microprocessors. Students will definitely benefit from building upon this. The purpose of implementing an efficient microprocessor who's working and intricacies are easy to understand has been successfully achieved.

6.Acknowledgments

We would like to thank Mr. Kunwar Singh (Associate Professor,Department of Electrical Engineering,Delhi Technological University) and Dr.Suman Bhowmick(Associate Professor,Department of Electrical Engineering,Delhi Technological University) who never ceased in providing their valuable guidance till this project was completed.

References

1. *Verilog HDL (2nd Edition),Samir Palnitkar*
2. *Computer System Architecture (3rd Edition) M Morris Mano,Prentice Hall (1993)*
3. *Computer Systems Organization and Architecture ,John D. Carpinelli*
4. Ronald, J. Tocci, Widmer, N. Moss, G. (1998), "Digital Systems Principles and Application", Prentice-Hall
5. International Inc., New Jersey, 182-341.
6. *Digital Design and Verilog HDL Fundamentals ,Joseph Canavagh,CRC Press(1st Edition 2008)*
7. *Verilog for Digital Design Frank Vahid , Roman Lysecky ,Wiley*

Digital Design (4th Edition)M.Morris Mano, Prentice Hall; 4 edition (December 25, 2006)

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a fast manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar

Role of e-Agriculture in Rural Development in Indian Context

Deka Ganesh Chandra
DGE&T, Ministry of Labour & Employment, New Delhi-1, India
ganeshdeka2000@gmail.com

Dutta Borah Malaya
Dept of Computer Engineering, Delhi Technological University, Delhi -42, India
malayadutta@dce.ac.in

Abstract— **e-Agriculture involves the conceptualization, design, development, evaluation and application of innovative ways to use information and communication technologies (ICTs) in the rural domain, with a primary focus on agriculture.People around the Globe from few years from now will be carrying a handheld computer connected to the Web to get the information about the World at their fingertips. e-Agriculture is an emerging field focused on the enhancement of agricultural and rural development through improved information and communication processes.**We have 28 States and 7 Union Territories 626 District **6, 38,596 villages and** 70% of the Indian population lives in rural areas **as per the Census of India 2001.**Agriculture in the backbone of Indian economy and 68% of India population is mainly dependent on Agriculture for their livelihood. **In this paper, we have discussed the problems & prospects of e-Agriculture in Rural Development in Indian context.**

Keywords- e-Agriculture, ICT,* Kishan Call Centre(KCC) & *Common Services Centers & Precision Agriculture.

I. INTRODUCTION

Indian Food Processing Market is US$ 105 billion currently and it is likely to increase 3 fold by 2015 and touch US$ 310 billion. India's share in world trade in food and agricultural products may touch US$ 20 billion by 2015 from the current $ 7.5 billion. In order to match the global demand for affordably priced food by 2050, annual food production must increase by more than 1% annually, and an estimated 80 percent of the increase will have to come from growth in yields.

Agriculture sector in India is facing lots of Challenges. 50% of the population engaged in agriculture are still illiterate and just 5% have completed Higher Secondary education. The Education & Incomes of agricultural labourers very low. During 1982 families operating farms below 1 hectare was 56% which has become 70% in 2003[1].

Information and Communication Technologies (ICT) plays a vital role in Development & Economic growth of the Developing countries of the World. Political, Cultural, Socio-Economic developmental & behavioral decisions today rests on the ability to access, gather, analyze and utilize information and knowledge. ICT is the conduits that transmit information and Knowledge to individual to widen their choices for Economic and social empowerment.

ICT also has an impact on a country's development and growth. Recent World Bank study shows that a 10% increase in mobile phone subscribers is associated with a 0.8% increase in economic growth while 10% increase in high-speed internet connections is related with a 1.3% increase in economic growth. ICT infrastructure development attracts foreign direct investment, generates fiscal revenues, and creates employment opportunities. Local information technology service industries generate exports, improve a firm's productivity, and offer equalizing job opportunities, especially for youth and women.

Mobile phones and other ICTs can provide a broad range of public and social services to the poor in remote areas and they have become an essential as well as an essential utility for the poor. Farmers in remote villages use mobile phones to access the most current crop prices and migrant workers use mobile banking services to transfer money to relatives back home. Increasingly, land registration, education, health care and voting are being conducted electronically using ICT. ICT is the conduits that transmit information and knowledge to individual to widen their choices for Economic and social empowerment. By integrating technology into development, more effective and speedy solutions can be found for sustainable human development & economic growth.

e-Agriculture involves application of information and communication technologies (ICTs) in an innovative ways to use with a primary focus on agriculture in the rural development domain.

FAO proposes the following definition: "e-Agriculture is an emerging field in the intersection of agricultural informatics, Agricultural development and entrepreneurship, referring to agricultural services, technology dissemination, and information delivered or enhanced through the Internet and related technologies. More specifically, it involves the conceptualization, design, development, evaluation and application of new (innovative) ways to use existing or emerging information and communication technologies (ICTs)".

The role of e-Agriculture [3] is:

- Ensure the systematic dissemination of information using ICTs on agriculture, animal husbandry, fisheries, forestry and food, in order to provide ready access to comprehensive, up-to-date and detailed knowledge and information, particularly in rural areas.
- Public-private partnerships should seek to maximize the use of ICTs as an instrument to improve production (quantity and quality).

The rest of this paper is organized into 6 sections. Section-2 is about Precision Agriculture in Section-3 we have discussed about the Indian Agricultural Scenario Section-4 is about Kisan Call Center & Community Service Center(CSC) Section-5 is about the some of the Challenges of India Agriculture Sector Section-6 ts the Conclusions & Future scope.

II. PRECISION AGRICULTURE

Precision Agriculture is the term used for application of sophisticated technologies in Agriculture to improve the quality as well the quantity of production. Precision Agriculture is about harnessing computer and satellite technologies by farmers for:

- Cutting costs
- Improving yields and
- Protect the environment

It also uses the e-commerce for marketing the sale of agricultural products over the Internet and extranets.

Precision agriculture is described as:"a system to manage farm resources better. Precision farming is an information technology based management system now possible because of several technologies currently available to agriculture. These include global positioning systems, geographic information systems, yield monitoring devices, soil, plant and pest sensors, remote sensing, and variable rate technologies for application of inputs."[3]

Applications of e-Agriculture has geared up for maximizing profits in developed countries but in many developing countries farmers' access to information is improved through grass root level initiatives of using ICTs as well as distance education modalities to enhance the knowledge base among service providers.

In precision agriculture or site-specific farming, farmers are using ICTs and other technologies to obtain more precise information about agricultural resources which allow them to identify, analyze, and manage the spatial and temporal variability of soil and plants for optimum profitability, sustainability, and protection of the environment [2].

Precision Agriculture is basically the advanced e-agriculture application. It makes use of 5 major components of technology:

- Geographical Information Systems (GIS) for analysis and management of spatial data and mapping;
- Remote Sensing (RS) to identify and
- Global Positioning Systems (GPS) to locate and define spatial features or activities that contributes to the quality of site-specific practices;
- Variable Rate Technology (VRT) allowing targeted, site-specific input applications; and
- Yield monitoring for recording crop productivity as an historical database for crop management [4]

III. INIAN AGRICULTURAL SCENARIO

7 out of 10 of the world's poor people still lives in rural areas. Agricultural sector's growth has a crucial role to play in improving the incomes of poor people, by providing farm jobs and stimulating off-farm employment.In order to match the global demand for affordably priced food by 2050, annual food production must increase by more than 1% annually, and an estimated 80 percent of the increase will have to come from growth in yields. Moreover, productivity-led increases in food and agricultural production will increase not only farm incomes, but will also stimulate backward and forward linkages in the rural economy and lead to a reduction in poverty.

Successes in transforming agriculture in India were based on state support to credit, inputs and irrigation infrastructure, which the market had failed to provide. However, the wrong policy choices can block the transmission of higher prices to producers, stifle incentives and discourage supply response [6].

Indian Food Processing Market is US$ 105 billion currently. It may increase 3 fold by 2015 and touch US$ 310 billion. India's share in world trade in food and agricultural products may touch US$ 20 billion by 2015 from the current US$ 7.5 billion.

The following table shows the world rank of India in the production of following commodities during the Year 2008:

IV. ICT IN AGRICULTURE KISAN CALL CENTRE

Asymmetry of Information between farmer and farmer, village and village, region and region and the country as a whole versus other countries is a big challenge for Indian Agriculture sector and ICT is the feasible solution.

More than 5 lakh Indian villages are already under the coverage of telecommunication network. To make the best utilization of the fast growing ICT sector Department of Agriculture and Cooperation appointed TCIL, a Govt. of India Company to start the Kisan Call Centre (KCC). KCC scheme is operational since 21st January 2004. These KCC can be accessed on a common toll free number 1551 anywhere in India.

There are 13 KCC; each centre is allocated a cluster of states. 116 Agriculture Graduate are posted in these call centers. The quarries received from farmers attended by these call centre executives (Agriculture Graduate) are replied in the local language. The quarries/problems which could not be cleared at level-I (call centre executives) are forwarded to 123 experts located in different parts of the country at State Agriculture Universities, ICAR institutes,

State Department of Agriculture, Horticulture and other developments for answering at Level –II.[7].

TABLE I. WORLD RANK OF INDIA IN FOOD PRODUCTION

Rank	Food Production Rank	
	Commodity	Production(MT)
1	Buffalo milk, whole, fresh	60900000
2	Rice, paddy	148260000
3	Cow milk, whole, fresh	44100000
4	Wheat	78570200
5	Sugar cane	348187900
6	Vegetables freshness	31402000
7	Cotton lint	3787000
8	Potatoes	34658000
9	Chillies and peppers, dry	1269850
10	Bananas	26217000
11	Mangoes, mangos teens, guavas	13649400
12	Groundnuts, with shell	7168000
13	Hen eggs, in shell	3060000
14	Onions, dry	13565000
15	Tomatoes	10303000
16	Indigenous Buffalo Meat	1407462
17	Soybeans	9905000
18	Chick peas	5748600
19	Indigenous Cattle Meat	896721
20	Millet	11340000

TABLE II. LIST OF KISAN CALL CENTRE IN INDIA

Sl No	Kisan Call Centre	
	Location	States covered
1	Mumbai	Maharashtra, Goa,Daman, Diu
2	Kanpur	Uttar Pradesh, Uttranchal
3	Kochi	Kerala, Lakshadweep
4	Bangalore	Karnataka
5	Chennai	T.N.,Andaman/ Nicobar
6	Hyderabad	Andhra Pradesh
7	Chandigarh	Chandigarh, J&K, HP, Punjab
8	Jaipur	Rajasthan
9	Indore	Madhya Pradesh, Chattisgarh
10	Kolkata	W Bengal, Bihar, Orissa, Jharkhand
11	Kolkata	N.E. States
12	Delhi	Delhi/ Haryana
13	Ahmedabad	Gujarat and Dadra & Nagar Haveli

Considering the India Agricultural scenario the KCC is the instant ICT solution for the farmers. The reasons are as following:

- Real time interaction in local languages
- Simply a mobile phone is sufficient for communication with KCC executives.
- Almost sufficient to meets the local requirement

A. ICT for masses :Common Services Centers (CSC)

In May 2006 the Government approved the CSC is a strategic cornerstone of the National e-Governance Plan (NeGP) as part of its commitment in the National Common Minimum Programme to introduce e-governance on a massive scale.

By March 2011, Department of Information & Technology (DIT), Government of India, has planned to rollout out 100,000+ Common Services Centers (CSCs) across the country particularly in rural areas [8].

High quality and cost-effective video, voice and data content and services in the areas of e-governance, education, health, telemedicine, entertainment as well as other private services would be provide by these CSCs. The main objectives of these CSCs are to offer web-enabled e-governance services to rural communities in providing the following services:

- Application forms
- Certificates and utility payments such as electricity, telephone and water bills.

Under the Scheme, favorable atmosphere would be created for private sector and NGOs active participation for effective implementation of the CSC Scheme thereby becoming a partner of the government in the development of rural India. The proposed CSC scheme under the Private Public Partnership (PPP) model will be consisting of a 3-tier structure:

- Village Level Entrepreneur(VLE) will be at the bottom of the structure also called the CSC operator
- The 2nd layer above the VLE will be the Service Centre Agency (SCA) which will be responsible for a cluster of 500-1000 CSCs; and
- The topmost layer called State Designated Agency (SDA) identified by the State Government will be responsible for managing the implementation above the entire State.

V. CHALLENGES

The following are the some of the challenges of the Indian agriculture:

- Agriculture sector in India is facing lots of Challenges. 50% of the population engaged in agriculture are still illiterate and just 5% have completed Higher Secondary education. The Education & Incomes of agricultural labourers very low. During 1982 families operating farms below 1 hectare was 56% which has become 70% in 2003.
- Lack of proper planning for utilization of funds available under the foreign aid is also a major issue. As per the CAG report during the Financial Year 2009-10 the total

amount of unutilized fund was Rs 9,557 Crore for "Agriculture & Rural Development".

- The real challenge for the India Agriculture is the information asymmetry between farmer and farmer, village and village, region and region and the country as a whole versus other countries in which e-Agriculture is the feasible solution.
- The 100,000+ Common Services Centers (CSCs) has a big role to plan in e-Agriculture as well as in bridging the digital divide in India. But the non-availability of the contents in local languages is a big handicap considering the huge percentage of illiterate population involved in Agriculture sector.
- Connectivity especially high speed Internet connectivity in rural areas is also another big challenge for e-Agriculture. Although the KCC is providing the first level instant solution for Agriculture related issues but it not sufficient for the growing demand of e-Agriculture like Precession Agriculture.

VI. Conclusion And Future Scope

- The real challenge for the India Agriculture is the information asymmetry between farmer and farmer, village and village, region and region and the country as a whole versus other countries in which e-Agriculture is the feasible solution.
- Main failure of Indian Agriculture sector has been the inability to reduce the dependence of the workforce on agriculture Half of those engaged in agriculture are still illiterate and just 5% have completed Higher Secondary education Incomes and education are of course least among agricultural labourers Families operating farms now suffer from much smaller holdings (70% below 1 hectare in 2003 compared to 56% in 1982).Farming members in such families are twice as likely to be illiterate as non-farming members Workforce in agriculture that lacks non-farm skills and is also ageing and getting feminized.
- Information and Communication Technologies (ICTs) play a key role in development and economic growth of the Nation. ICT is Cross Cutter, which can be used for collaboration of various services for single window solutions for the citizens. ICT must reach the grass root level for bridging the Digital Divide.
- ICT is already on roll in e-commerce, but e-Agriculture is yet to find its feet. The best example of e-commerce application for the masses in India is India Railways Catering & Tourism Services online ticketing website https://www.irctc.co.in & Core Banking Solutions (CBS) of various banks.
- The integration of e-commerce & e-Agriculture is one of the best options keeping in view the successful implementation of many e-commerce solutions as of now.
- The CSCs are likely to serve as Information Kiosks for easy & quick access for Information on Agriculture. The project in the PPP model specifically in the NE States without Grants/Subsidies will not be economically viable. In some cases it will be difficult to find Village Level Entrepreneur (VLE) to operate a CSC in remote areas.

e-Agriculture is an emerging field in the intersection of agricultural informatics, Agricultural development and entrepreneurship, referring to agricultural services, technology dissemination, and information delivered or enhanced through the Internet and related technologies. There are lots of scopes for Study, Research & Development in e-Agriculture.

References

[1] Shailendra Singh "Agribusiness Entrepreneurs: Competency Approach", Proc. National Conference on Entrepreneurial Opportunities in Agri-Business: Challenges & Prospects Organized by Banaras Hindu University, Faculty of management studies,2008.

[2] Jefferson County Cooperative Extension Services. 2000. Precision agriculture-sitespecificfarming.<URL: http://jefferson.ifas.ufl.edu/ag_pages/PA-SST_page.htm>.

[3] Rains, G. C. and Thomas, D. L. 2000. Precision farming: an introduction.

[4] <URL http://pubs.caes.uga.edu/caespubs/pubcd/B1186.htm>

[5] World Summit on the Information Society, Geneva 2003- Tunis 2005, Plan of Action 21(E).

[6] World Agriculture towards 2015/2030: An FAO perspective, Edited by Jelle Bruinsma

[7] Website of Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India.

2012 17th Opto-Electronics and Communications Conference (OECC 2012) Technical Digest
July 2012, Busan, Korea

Segmented Cladding Fiber Design for Femtosecond Laser Pulse Delivery at 1550-nm and 1064-nm Wavelengths

Babita*, Vipul Rastogi*, and Ajeet Kumar**
*Department of Physics, Indian Institute of Technology Roorkee, Roorkee 247 667, INDIA
**Department of Applied Physics, Delhi Technological University, New Delhi, INDIA

Abstract

We numerically demonstrate near distortion-free propagation of 100-fs, over 50-kW peak power pulses at 1550-nm wavelength with 1825-μm^2 mode area, and 250-fs, 15-kW pulses at 1064-nm wavelength with 1800 μm^2 mode area, through segmented cladding fiber.

I. Introduction

For many applications it is desirable to deliver ultra-short pulses through an optical fiber over a distance of few meters to some specific location. In past few years substantial progress has been achieved in ultra-short pulse (USP) laser beam delivery using optical fibers. Propagation of such high peak power pulses in small-core conventional single-mode fibers leads to self phase modulation (SPM) and unrecoverable pulse distortion. In large-core multimode fibers intermodal dispersion significantly distorts ultra-short pulses. To preserve the pulse during propagation through the fiber, one needs to compensate for or minimize dispersive and nonlinear effects. Dispersion can be compensated by suitable frequency chirp on input pulses. One approach to reduce nonlinearity is to increase mode area by increasing the core size and correspondingly decreasing the numerical aperture (NA) while maintaining single-mode operation. While decrease in NA yields to extremely high bending loss because of weak mode confinement, increasing core radius without maintaining single-mode operation leads to multimode cross-coupling. Recently, specialty fibers such as photonic crystal fibers (PCFs) [1] or photonic band gap fibers, OmniGuide fibers, multi-mode high dispersion fibers, higher-order mode fibers [2], and rigid glass rod fibers have drawn attention for delivery of USPs. In this paper we propose a design of segmented cladding fiber (SCF) for delivery of high energy fs-pulses through the fundamental mode of the fiber. The SCF has been successfully fabricated in polymer and silver halide glass [3, 4]. One of the important features of the SCF is that its mode area does not vary much with the wavelength. Here we utilize this feature of SCF to design a fiber for high peak power ultra-short laser pulse delivery through LP_{01} mode at 1550 and 1064-nm wavelengths. The design has low susceptibility to mode coupling, leaky higher-order modes except LP_{11} and sufficiently low bend loss. We have numerically demonstrated the distortion-free propagation of 100-fs, over 50-kW peak power laser pulses through 4-m length of the fiber with mode area of about 1825 μm^2 at 1550-nm wavelength and 250-fs, 15-kW peak power laser pulses at 1064-nm wavelength with a mode area of about 1800 μm^2.

This work has been partially supported by research project 'Development of silica based segmented cladding fiber' sponsored by Department of Science and Technology, New Delhi, India.

II. Fiber Design and Method of Analysis

Transverse cross-section of the fiber with modal field of LP_{01} at 1550-nm wavelength is shown in Fig. 1(a). Modal properties of the fiber have been analyzed by radial effective index method (REIM) in conjunction with transfer matrix method (TMM) [5, 6]. Effective area of fundamental mode A_{eff} and dispersion coefficient D of the fiber at different wavelengths for 60% duty cycle are plotted in Fig. 1(b). We can see that the A_{eff} does not vary significantly over the entire wavelength range.

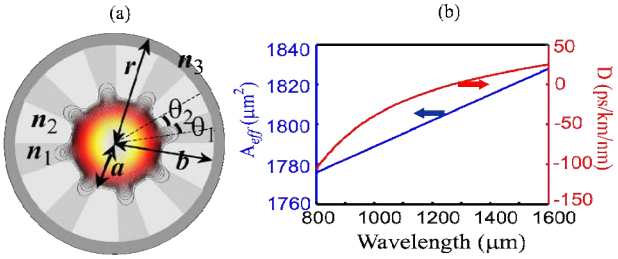


Fig.1. (a) Transverse cross-section of SCF with field profile of LP_{01} mode at 1550 nm wavelength (b) Variation of A_{eff} and dispersion coefficient D with wavelength.

III. Results and Discussion

Propagation of the pulse through the fiber can be characterized by ratio of dispersion length, L_D and the nonlinear length L_{NL} [7]. When L_D / L_{NL} < 1, pulse evolution along the fiber is dominated by dispersion and pulse will get broaden in absence of any chirp otherwise it stretches or compresses in time as it propagates through the fiber depending on the sign of initial chirp. In order to avoid nonlinearities, one needs to limit peak power to make L_{NL} larger than L_D. One approach to do this is by increasing the A_{eff}. We have employed the SCF design to achieve this.

978-1-4673-0978-3/12/$31.00 ©2012 IEEE

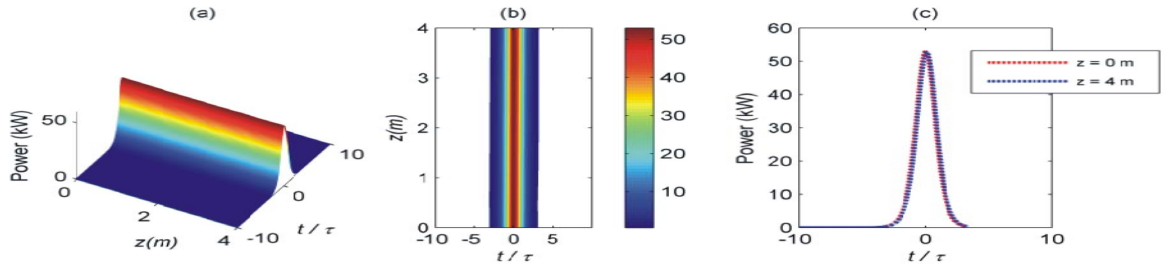


Fig. 2. (a) Propagation of a sech pulse in the SCF at 1550-nm wavelength (b) corresponding contour plot (c) input and output temporal profiles of pulses

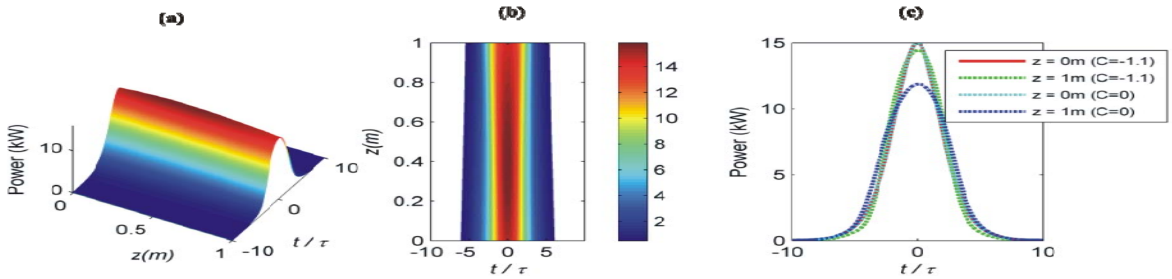


Fig. 3. (a) Propagation of a sech pulse in the SCF at 1064-nm wavelength (b) corresponding contour plot (c) input and output temporal profiles of the chirped and unchirped pulses

To study pulse propagation at 1550-nm wavelength through the fiber we have launched the following secant-hyperbolic pulse and studied its propagation dynamics by solving the nonlinear Schrödinger equation by split step Fourier method [7].

$$U(z=0,T) = P_0^{1/2} \times \mathrm{sech}(T/\tau)^{1+iC} \qquad (1)$$

where C is chirp parameter and τ is 1/e half width. In the simulation of pulse propagation through the fiber, we have considered second order Group velocity dispersion (GVD), third order GVD, SPM, and transmission loss of 0.2 dB/km. Since the pulses are of the order of 100 fs, we have also taken into account self-steepening and Raman scattering. The value of Raman time constant for silica has been taken as 3 fs. Evolution of pulse, corresponding contour plot and input-output profiles after 4-m propagation through the fiber are shown in Figs. 2(a), 2(b) and 2(c) respectively. We can see near distortion-free propagation of unchirped sech pulse.

We have also studied propagation of 250-fs chirped sech pulse at 1064-nm wavelength with a peak power of 15 kW through the same fiber. We have considered input pulse with $C = -1.1$, which is sufficient to counterbalance the total chirping, induced due to dispersion and nonlinearity. The evolution of the sech pulse through the fiber is shown in Fig. 3(a) with corresponding contour plot in Fig. 3(b). The input and output temporal profiles for a 1-m long fiber for $C = 0$ and $C = -1.1$ are shown in Fig. 3(c). We can see near distortion-free propagation of the pulse over 1-m distance of propagation.

IV. Conclusions

We have designed a large-mode-area SCF for high-peak power fs-duration pulse delivery. The fiber has mode areas of 1825 μm^2 at 1550-nm wavelength and 1800 μm^2 at 1064-nm wavelength. We have shown that the fiber can deliver 100-fs sech and Gaussian pulses at 1550-nm wavelength and 250-fs pulses at 1064-nm wavelength without significant distortion. The proposed design should be useful in the area of femtochemistry, biomedical applications and micromachining.

References

[1] W. Göbel, A. Nimmerjahn, and F. Helmchen, "Distortion-free delivery of nanojoule femtosecond pulses from a Ti:sapphire laser through a hollow-core photonic crystal fiber," *Opt. Lett*., vol. 29, pp. 1285-1287, 2004.

[2] S. Ramachandran, J. W. Nicholson, S. Ghalmi, M. F. Yan, P. Wisk, E. Monberg, and F. V. Dimarcello, "Light propagation with ultralarge modal areas in optical fibers," *Opt. Lett*., vol. 31, pp. 1797-1799, 2006.

[3] J. Duan, C. Teng, K. Han, W. Wu, Q. Zhang, and K.S. Chiang, "Fabrication of segmented cladding fiber by bicomponent spinning," *Polym Eng Sci*., vol. 49, pp. 1865-1870 (2009).

[4] A. Millo, I. Naeh, Y. Lavi, and A. Katzir, "Silver-halide segmented cladding fibers for the middle infrared," *Appl. Phys. Lett*., vol. 88, pp. 251101-251103 (2006).

[5] V. Rastogi and K. S. Chiang, "Analysis of segmented-cladding fiber by the radial-effective-index method," *J. Opt. Soc. Am. B*, vol. 21, pp. 258-265, 2004.

[6] K. Thyagarajan, S. Diggavi, A. Taneja, and A. K. Ghatak, "Simple numerical technique for the analysis of cylindrically symmetric refractive-index profile optical fibers," *Appl. Opt.* , vol. 30, pp. 3877–3879 , 1991.

[7] G. P. Agrawal, *Nonlinear Fiber Optics*. Academic: San Diego., 2001, Chaps. 2-5.

Sentiment Analysis on Twitter

Akshi Kumar and Teeja Mary Sebastian

Department of Computer Engineering, Delhi Technological University
Delhi, India

Abstract

With the rise of social networking epoch, there has been a surge of user generated content. Microblogging sites have millions of people sharing their thoughts daily because of its characteristic short and simple manner of expression. We propose and investigate a paradigm to mine the sentiment from a popular real-time microblogging service, Twitter, where users post real time reactions to and opinions about "everything". In this paper, we expound a hybrid approach using both corpus based and dictionary based methods to determine the semantic orientation of the opinion words in tweets. A case study is presented to illustrate the use and effectiveness of the proposed system.

Keywords: Microblogging, Twitter, Sentiment Analysis

1. Introduction

Ongoing increase in wide-area network connectivity promise vastly augmented opportunities for collaboration and resource sharing. Now-a-days, various social networking sites like Twitter[1], Facebook[2], MySpace[3], YouTube[4] have gained so much popularity and we cannot ignore them. They have become one of the most important applications of Web 2.0 [1]. They allow people to build connection networks with other people in an easy and timely way and allow them to share various kinds of information and to use a set of services like picture sharing, blogs, wikis etc.

It is evident that the advent of these real-time information networking sites like Twitter have spawned the creation of an unequaled public collection of opinions about every global entity that is of interest. Although Twitter may provision for an excellent channel for opinion creation and presentation, it poses newer and different challenges and the process is incomplete without adept tools for analyzing those opinions to expedite their consumption.

More recently, there have been several research projects that apply sentiment analysis to Twitter corpora in order to extract general public opinion regarding political issues [2]. Due to the increase of hostile and negative communication over social networking sites like Facebook and Twitter, recently the Government of India tried to allay concerns over censorship of these sites where Web users continued to speak out against any proposed restriction on posting of content. As reported in one of the Indian national newspaper [3] "Union Minister for Communications and Information Minister, Kapil Sibal, proposed content screening & censorship of social networks like Twitter and Facebook". Instigated by this the research carried out by us was to use sentiment analysis to gauge the public mood and detect any rising antagonistic or negative feeling on social medias. Although, we firmly believe that censorship is not right path to follow, this recent trend for research for sentiment mining in twitter can be utilized and extended for a gamut of practical applications that range from applications in business (marketing intelligence; product and service bench marking and improvement), applications as sub-component technology (recommender systems; summarization; question answering) to applications in politics. This motivated us to propose a model which retrieves tweets on a certain topic through the Twitter API and calculates the sentiment orientation/score of each tweet.

The area of Sentiment Analysis intends to comprehend these opinions and distribute them into the categories like positive, negative, neutral. Till now most sentiment analysis work has been done on review sites [4]. Review sites provide with the sentiments of products or movies, thus, restricting the domain of application to solely business. Sentiment analysis on Twitter posts is the next step in the field of sentiment analysis, as tweets give us a richer and more varied resource of opinions and sentiments that can be about anything from the latest phone they bought, movie they watched, political issues, religious views or the individuals state of mind. Thus, the foray into Twitter as the corpus allows us to move into different dimensions and diverse applications.

2. Related Work

Applying sentiment analysis on Twitter is the upcoming trend with researchers recognizing the scientific trials and its potential applications. The challenges unique to this problem area are largely attributed to the dominantly

[1]www.twitter.com
[2]www. facebook.com
[3]www.myspace.com
[4]www.youtube.com

informal tone of the micro blogging. Pak and Paroubek [5] rationale the use microblogging and more particularly Twitter as a corpus for sentiment analysis. They cited:

- Microblogging platforms are used by different people to express their opinion about different topics, thus it is a valuable source of people's opinions.
- Twitter contains an enormous number of text posts and it grows every day. The collected corpus can be arbitrarily large.
- Twitter's audience varies from regular users to celebrities, company representatives, politicians, and even country presidents. Therefore, it is possible to collect text posts of users from different social and interests groups.
- Twitter's audience is represented by users from many countries.

Parikh and Movassate [6] implemented two Naive Bayes unigram models, a Naive Bayes bigram model and a Maximum Entropy model to classify tweets. They found that the Naive Bayes classifiers worked much better than the Maximum Entropy model could. Go et al. [7] proposed a solution by using distant supervision, in which their training data consisted of tweets with emoticons. This approach was initially introduced by Read [8]. The emoticons served as noisy labels. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM). Their feature space consisted of unigrams, bigrams and POS. The reported that SVM outperformed other models and that unigram were more effective as features. Pak and Paroubek [5] have done similar work but classify the tweets as objective, positive and negative. In order to collect a corpus of objective posts, they retrieved text messages from Twitter accounts of popular newspapers and magazine, such as "New York Times", "Washington Posts" etc. Their classifier is based on the multinomial Naïve Bayes classifier that uses N-gram and POS-tags as features. Barbosa et al. [9] too classified tweets as objective or subjective and then the subjective tweets were classified as positive or negative. The feature space used included features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words.

Mining for entity opinions in Twitter, Batra and Rao[10] used a dataset of tweets spanning two months starting from June 2009. The dataset has roughly 60 million tweets. The entity was extracted using the Stanford NER, user tags and URLs were used to augment the entities found. A corpus of 200,000 product reviews that had been labeled as positive or negative was used to train the model. Using this corpus the model computed the probability that a given unigram or bigram was being used in a positive context and the probability that it was being used in a negative context. Bifet and Frank [11] used Twitter streaming data provided by Firehouse, which gave all messages from every user in real-time. They experimented with three fast incremental methods that were well-suited to deal with data streams: multinomial naive Bayes, stochastic gradient descent, and the Hoeffding tree. They concluded that SGD-based model, used with an appropriate learning rate was the best.

Agarwal et al. [12] approached the task of mining sentiment from twitter, as a 3-way task of classifying sentiment into positive, negative and neutral classes. They experimented with three types of models: unigram model, a feature based model and a tree kernel based model. For the tree kernel based model they designed a new tree representation for tweets. The feature based model that uses 100 features and the unigram model uses over 10,000 features. They concluded features that combine prior polarity of words with their parts-of-speech tags are most important for the classification task. The tree kernel based model outperformed the other two.

The Sentiment Analysis tasks can be done at several levels of granularity, namely, word level, phrase or sentence level, document level and feature level [13]. As Twitter allows its users to share short pieces of information known as "tweets" (limited to 140 characters), the word level granularity aptly suits its setting. Survey through the literature substantiates that the methods of automatically annotating sentiment at the word level fall into the following two categories: (1) dictionary-based approaches and (2) corpus-based approaches. Further, to automate sentiment analysis, different approaches have been applied to predict the sentiments of words, expressions or documents. These include Natural Language Processing (NLP) and Machine Learning (ML) algorithms [14]. In our attempt to mine the sentiment from twitter data we introduce a hybrid approach which combines the advantages of both dictionary & corpus based methods along with the combination of NLP & ML based techniques. The following sections illustrate the proposed paradigm.

3. Data Characteristics

Twitter is a social networking and microblogging service that lets its users post real time messages, called tweets. Tweets have many unique characteristics, which implicates new challenges and shape up the means of carrying sentiment analysis on it as compared to other domains.

Following are some key characteristics of tweets:

- *Message Length:* The maximum length of a Twitter message is 140 characters. This is different from previous sentiment classification research that focused on classifying longer texts, such as product and movie reviews.
- *Writing technique:* The occurrence of incorrect spellings and cyber slang in tweets is more often in comparison with other domains. As the messages are quick and short, people use acronyms, misspell, and use emoticons and other characters that convey special meanings.
- *Availability:* The amount of data available is immense. More people tweet in the public domain as compared to Facebook (as Facebook has many privacy settings) thus making data more readily available. The Twitter API facilitates collection of tweets for training.
- *Topics:* Twitter users post messages about a range of topics unlike other sites which are designed for a specific topic. This differs from a large fraction of past research, which focused on specific domains such as movie reviews.
- *Real time:* Blogs are updated at longer intervals of time as blogs characteristically are longer in nature and writing them takes time. Tweets on the other hand being limited to 140 letters and are updated very often. This gives a more real time feel and represents the first reactions to events.

We now describe some basic terminology related to twitter:

- *Emoticons:* These are pictorial representations of facial expressions using punctuation and letters. The purpose of emoticons is to express the user's mood.
- *Target:* Twitter users make use of the "@" symbol to refer to other users on Twitter. Users are automatically alerted if they have been mentioned in this fashion.
- *Hash tags:* Users use hash tags "#" to mark topics. It is used by Twitter users to make their tweets visible to a greater audience.
- *Special symbols*: "RT" is used to indicate that it is a repeat of someone else's earlier tweet.

4. System Architecture

Opinion words are the words that people use to express their opinion (positive, negative or neutral). To find the semantic orientation of the opinion words in tweets, we propose a novel hybrid approach involving both corpus-based and dictionary-based techniques. We also consider features like emoticons and capitalization as they have recently become a large part of the cyber language.

Fig.1gives the architectural overview of the proposed system.

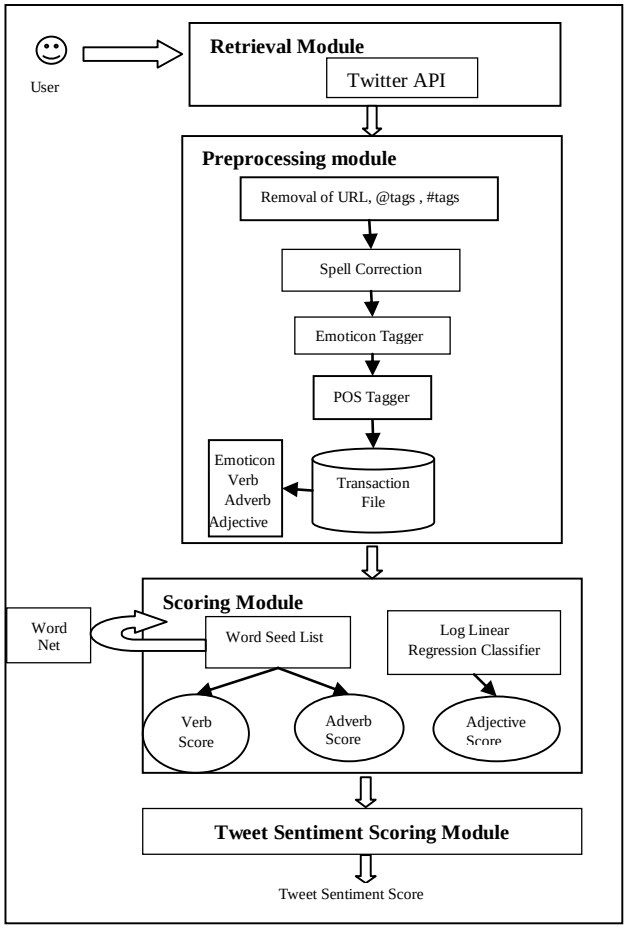


Fig. 1. System Architecture.

To uncover the opinion direction, we will first extract the opinion words in the tweets and then find out their orientation, i.e., to decide whether each opinion word reflects a positive sentiment, negative sentiment or a neutral sentiment. In our work, we are considering the opinion words as the combination of the adjectives along with the verbs and adverbs. The corpus-based method is then used to find the semantic orientation of adjectives and the dictionary-based method is employed to find the semantic orientation of verbs and adverbs. The overall tweet sentiment is then calculated using a linear equation which incorporates emotion intensifiers too.

The following sub-sections expound the details of the proposed system:

4.1 Pre-processing of Tweets

We prepare the transaction file that contains opinion indicators, namely the adjective, adverb and verb along

with emoticons (we have taken a sample set of emoticons and manually assigned opinion strength to them). Also we identify some emotion intensifiers, namely, the percentage of the tweet in Caps, the length of repeated sequences & the number of exclamation marks, amongst others. Thus, we pre-process all the tweets as follows:

a) Remove all URLs (e.g. www.example.com), hash tags (e.g. #topic), targets (@username), special Twitter words ("e.g. RT").
b) Calculate the percentage of the tweet in Caps.
c) Correct spellings; A sequence of repeated characters is tagged by a weight. We do this to differentiate between the regular usage and emphasized usage of a word.
d) Replace all the emoticons with their sentiment polarity (Table 1).
e) Remove all punctuations after counting the number of exclamation marks.
f) Using a POS tagger, the NL Processor linguistic Parser [15], we tag the adjectives, verbs and adverbs.

Table 1: Emoticons

Emoticon	***Meaning***	***Strength***
:D	Big grin	1
BD	Big grin with glasses	1
XD	Laughing	1
\m/	Hi 5	1
:),=),:-)	Happy, smile	0.5
:*	kiss	0.5
:|	Straight face	0
:\	undecided	0
:(	sad	-0.5
</3	Broken heart	-0.5
B(	Sad with glasses	-0.5
:'(	crying	-1
X-(	angry	-1

4.2 Scoring Module

The next step is to find the semantic score of the opinion carriers i.e. the adjectives, verbs and adverbs. As mentioned previously, in our approach we use corpus-based method to find the semantic orientation of adjectives and the dictionary-based method to find the semantic orientation of verbs and adverbs.

4.2.1 Semantic Score of Adjectives

An adjective are a describing word and is used to qualify an object. The semantic orientation of adjectives tend to be domain specific, therefore we use a corpus based approach to quantify the semantic orientation of adjectives in the Twitter domain. Motivated by Hatzivassiloglou and McKeown [16], we ascribe same semantic orientation to conjoined adjectives in most cases and in special cases when the connective is "but", the situation is reversed. Similar to them we apply a log-linear regression model with a linear predictor

$$\eta = w^T x \quad (1)$$

where x is the vector of observed counts in the various conjunction categories(all *and* pairs, all *but* pairs, all attributive *and* pairs, etc.) for the particular adjective pair and w is the vector of weights to be learnt during training. The response y is non-linearly related to η through the inverse logit function

$$y = \frac{e^{\eta}}{1+e^{\eta}} \quad (2)$$

The value y produced denotes the similarity between the words. The seed list of adjectives was taken and assigned semantic scores manually. We also calculated the semantic score of conjoined adjectives by using the manually assigned scores and the similarity value y.

4.2.2 Semantic Score of Adverbs and Verbs

Although, we can compute the sentiment of a certain texts based on the semantic orientation of the adjectives, but including adverbs is imperative. This is primarily because there are some adverbs in linguistics (such as "not") which are very essential to be taken into consideration as they would completely change the meaning of the adjective which may otherwise have conveyed a positive or a negative orientation.

For example;
One user says, "***This is a good book***" and;
Other says, "***This is not a good book***"

Here, if we had not considered the adverb "not", then both the sentences would have given positive review. On the contrary, first sentence gives the positive review and the second sentence gives the negative review. Further, the strength of the sentiment cannot be measured by merely considering adjectives alone as the opinion words. In other words, an adjective cannot alone convey the intensity of the sentiment with respect to the document in question. Therefore, we take into consideration the adverb strength which modify the adjective; in turn modifying the sentiment strength. Adverb strength helps in assessing whether a document gives a *perfect* positive opinion, *strong* positive opinion, a *slight* positive opinion or a *less* positive opinion.

For example;
One user says, ***"This is a very good book"*** and ;
Other says, ***"This is a good book"***

Some groups of verbs also convey sentiments and opinions (e.g. love, like) and are essential to finding the sentiment strength of the tweet. As adverbs and verbs are not

dependent on the domain, we use dictionary methods to calculate their semantic orientation.

The seed lists of positive and negative adverbs and verbs whose orientation we know is created and then grown by searching in WordNet [17]. Based on intuition, we assign the strengths of a few frequently used adverbs and verbs with values ranging from -1 to +1. We consider some of the most frequently used adverbs and verbs along with their strength as given below in table 2:

Table 2: Verb and Adverb Strengths

Verb	***Strength***	***Adverb***	***Strength***
Love	1	complete	+1
adore	0.9	most	0.9
like	0.8	totally	0.8
enjoy	0.7	extremely	0.7
smile	0.6	too	0.6
impress	0.5	very	0.4
attract	0.4	pretty	0.3
excite	0.3	more	0.2
relax	0.2	much	0.1
reject	-0.2	any	-0.2
disgust	-0.3	quite	-0.3
suffer	-0.4	little	-0.4
dislike	-0.7	less	-0.6
detest	-0.8	not	-0.8
suck	-0.9	never	-0.9
hate	-1	hardly	-1

The complete procedure for predicting adverb and verb polarity is given below:

Procedure "*determine_orientation"* takes the target Adverb/ Verb whose orientation needs to be determined and the respective seed list as the inputs.

1. Procedure **determine_orientation** (target_Adverb/ Verb w_i, Adverb/ Verb_ seedlist)
2. begin
3. if (w_i has synonym s in Adverb/ Verb _ seedlist)
4. { w_i's orientation= s's orientation;
5. add w_i with orientation to Adverb/ Verb _ seedlist ; }
6. else if (w_i has antonym a in Adverb/ Verb _ seedlist)
7. { w_i's orientation = opposite orientation of *a*'s orientation;
8. add w_i with orientation to Adverb/ Verb _ seedlist; }
9. end

The procedure *determine_orientation* searches Word Net and the Adverb/ Verb seed list for each target adjective to predict its orientation (line 3 to line 8). In line 3, it searches synonym set of the target Adverb/ Verb from the Word Net and checks if any synonym has known orientation from the seed list. If so, the target orientation is set to the same orientation as the synonym (line 4) and the target Adverb/ Verb along with the orientation is inserted into the seed list (line 5). Otherwise, the function continues to search antonym set of the target Adverb/ Verb from the Word Net and checks if any Adverb/ Verb have known orientation from the seed list (line 6). If so, the target orientation is set to the opposite of the antonym (line 7) and the target Adverb/ Verb with its orientation is inserted into the seed list (line 8). If neither synonyms nor antonyms of the target word have known orientation, the function just continues the same process for the next Adverb/ Verb since the word's orientation may be found in a later call of the procedure with an updated seed list.

Note:

1) *For those adverbs/ verbs that Word Net cannot recognize, they are discarded as they may not be valid words.*
2) *For those that we cannot find orientations, they will also be removed from the opinion words list and the user will be notified for attention.*
3) *If the user feels that the word is an opinion word and knows its sentiment, he/she can update the seed list.*
4) *For the case that the synonyms/antonyms of an adjective have different known semantic orientations, we use the first found orientation as the orientation for the given adjective.*

4.3 Tweet Sentiment Scoring

As adverbs qualify adjectives and verbs, we group the corresponding adverb and adjective together and call it the adjective group; similarly we group the corresponding verb and adverb together and call it the verb group. The adjective group strength is calculated by the product of adjective score (adj_i) and adverb (adv_i) score, and the verb group strength as the product of verb score (vb_i) and adverb score (adv_i). Sometimes, there is no adverb in the opinion group, so the S (adv) is set as a default value 0.5

To calculate the overall sentiment of the tweet, we average the strength of all opinion indicators like emoticons, exclamation marks, capitalization, word emphasis, adjective group and verb group as shown below:

$$S(T) = \frac{(1+(P_c+\log(Ns)+\log(N_x))/3)}{|OI(R)|} * \sum_{i=1}^{|OI(R)|} S(AG_i)+S(VG_i)+N_{ei} * S(E_i) \quad (3)$$

Where,
|OI(R)| denotes the size of the set of opinion groups and emoticons extracted from the tweet,
P_c denotes fraction of tweet in caps,
N_s denotes the count of repeated letters,
N_x denotes the count of exclamation marks,
S (AG_i) denotes score of the i^{th} adjective group,
S (VG_i) denotes the score of the i^{th} verb group,
S (E_i) denotes the score of the i^{th} emoticon
N_{ei} denotes the count of the i^{th} emoticon.

P_c, N_s and N_x represent emphasis on the sentiment to be conveyed so they can be collectively called sentiment intensifiers.
If the score of the tweet is more than 1 or less than -1, the score is taken as 1 or -1 respectively.

5. Illustrative Case Study

To clearly illustrate the effectiveness of the proposed method, a case study is presented with a sample tweet:

<tweet>="@kirinv I hate revision, it's BOOOORING!!! I am totally unprepared for my exam tomorrow :(:(Things are not good...#exams"

5.1 The pre-processing of Tweet

A transaction file is created which contains the preprocessed opinion indicators.

5.1.1 Extracting Opinion Intensifiers

The opinion intensifiers are calculated for the tweet as follows.

1) Fraction of tweet in caps:
   There are a total of 18 words in the sentence out of which one is in all caps. Therefore, P_c=1/18=0.055
2) Length of repeated sequence, N_s=3
3) Number of Exclamation marks, N_x=3

5.1.2 Extracting Opinion Words

After the tweet is preprocessed, it is tagged using a POS tagger and the adjective and verb groups are extracted.

The list of Adjective Groups extracted**:**
AG_1=totally unprepared
AG_2=not good
AG_3=boring

The list of Verb Groups extracted:
VG_1=hate

The list of Emoticons extracted:
$E_{1\,=}$:(
$N_{e1\,=}$ 2

5.2 Scoring Module

Now that we have our adjective group and verb group, we have to find their semantic orientation. Calculation is based on ke

5.2.1 Score of Adjective Group

S (AG_1) = S (totally unprepared) =0.8*-0.5 == -0.4
S (AG_2) = S (not good) =-0.8*1= -0.8
S (AG_3) = S (boring) = 0.5*-0.25 = -0.125

5.2.2 Score of Verb Group

S (VG_1) = S (hate) = 0.5*-0.75 = 0.375

5.3 Tweet Sentiment Scoring

Using the formula defined in equation 3 we can calculate the sentiment strength of the tweet as follows:

$$S(T) = \frac{1.33}{5}\sum_{i=1}^{5} S(AG_i) + S(VG_i) + N_{ei} * S(E_i)$$

$$= \frac{(1.33)}{5} * ((-0.4) + (-0.8) + (-0.125) + (-0.5) + 2*(-0.5))$$
$$= -0.751$$

As we have got a negative value, we can safely classify the tweet as negative.

We applied our approach to a sample set of 10 tweets. The semantic analysis results obtained are depicted in table 3 below.

Table 3: Sample Tweets and semantic orientation

Tweet	***Score***	***Orientation***
@kirinv I hate revision, it's BOOOORING!!! I am totally unprepared for my exam tomorrow :(:(Things are not good...#exams	-0.751	Negative
Criticism of UID laumched yday is extremely unfair. You may hate or even envy Nilekani but can not deny the idea.	0.009	Neutral
"@bigDEElight Keeping it real gone wrong, that was hilarious!! And I wonder how often that actually happens IRL!	0.145	Positive
#iranElection this could get nasty	-0.437	Negative
just getting back from Oaxaca, Mexico by plane	0.125	Positive
I have created a twitter! This is my ONE AND ONLY twitter guys, someone already stole my url. not too happy about it either :(	-0.24	Negative
Happy happy happy :D	0.625	Positive
That was pretty much awesome. :)	0.263	Positive
That other dude sucks!!!	-0.664	Negative
@prncssmojo hey i got a im thingy what is ur screen name?	0	Neutral
Just got home From work. Dam it wuz tough today	-0.281	Negative

The practice result proves that the proposed system has the characteristics of perceiving the semantic orientation of tweets. The results of this work serve as a partial view of the phenomenon. More research needs to be done in order to validate or invalidate these findings, using larger samples.

6. Conclusion

The proliferation of microblogging sites like Twitter offers an unprecedented opportunity to create and employ theories & technologies that search and mine for sentiments. The work presented in this paper specifies a novel approach for sentiment analysis on Twitter data. To uncover the sentiment, we extracted the opinion words (a combination of the adjectives along with the verbs and adverbs) in the tweets. The corpus-based method was used to find the semantic orientation of adjectives and the dictionary-based method to find the semantic orientation of verbs and adverbs. The overall tweet sentiment was then calculated using a linear equation which incorporated emotion intensifiers too. This work is exploratory in nature and the prototype evaluated is a preliminary prototype. The initial results show that it is a motivating technique.

References

[1] L. Colazzo, A. Molinari and N. Villa. "Collaboration vs. Participation: the Role of Virtual Communities in a Web 2.0 world", International Conference on Education Technology and Computer, 2009, pp.321-325.

[2] nlp.stanford.edu/courses/cs224n/2011/reports/patlai.pdf

[3] National Daily, Economic Times: articles.economictimes.indiatimes.com › Collections › Facebook

[4] K. Dave, S. Lawrence, and D.M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews". In Proceedings of the 12th International Conference on World Wide Web (WWW), 2003, pp. 519–528.

[5] A. Pak and P. Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320–1326.

[6] R. Parikh and M. Movassate, "Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques", CS224N Final Report, 2009

[7] A. Go, R. Bhayani, L.Huang. "Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper ,2009

[8] J. Read. "Using emoticons to reduce dependency in machine learning techniques for sentiment classification". In Proceedings of ACL-05, 43nd Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2005

[9] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume, pp. 36-44.

[10] S. Batra and D. Rao, "Entity Based Sentiment Analysis on Twitter", Stanford University,2010

[11] A. Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data", In Proceedings of the 13th International Conference on Discovery Science, Berlin, Germany: Springer,2010, pp. 1–15.

[12] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011 Workshop on Languages in Social Media,2011 , pp. 30–38

[13] A. Kumar. and T. M. Sebastian, "Sentiment Analysis: A Perspective on its Past, Present and Future", International Journal of Intelligent Systems and Applications (IJISA), MECS Publisher, 2012 (Accepted to be published)

[14] A. Kumar and T. M. Sebastian, "Machine learning assisted Sentiment Analysis". Proceedings of International Conference on Computer Science & Engineering (ICCSE'2012), 2012, pp. 123-130.

[15] POS Tagger: http://www.infogistics.com/textanalysis.html

[16] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives". In Proceedings of the Joint ACL/EACL Conference,2004, pp. 174–181

[17] WordNet: http://wordnet.princeton.edu/

Akshi Kumar is a PhD in Computer Engineering from University of Delhi. She has received her MTech (Master of Technology) and BE (Bachelor of Engineering) degrees in Computer Engineering. She is currently working as a University Assistant Professor in Dept. of Computer Engineering at the Delhi Technological University, Delhi, India. She is editorial review board member for '*The International Journal of Computational Intelligence and Information Security'*, Australia, ISSN: 1837-7823; *'International Journal of Computer Science and Information Security'*, USA, ISSN: 1947-5500; '*Inter-disciplinary Journal of Information, Knowledge & Management'*, published by the Informing Science Institute, USA. (ISSN Print 1555-1229, Online 1555-1237) and '*Webology*', ISSN 1735-188X. She is a life member of Indian Society for Technical Education (ISTE), India, a member of International Association of Computer Science and Information Technology (IACSIT), Singapore, a member of International Association of Engineers (IAENG), Hong Kong, a member of IAENG Society of Computer Science, Hong Kong and a member of Internet Computing Community (ICC), AIRCC. She has many publications to her credit in various journals with high impact factor and international conferences. Her current research interests are in the area of Web Search & Mining, Intelligent Information Retrieval, Web 2.0 & Web Engineering.

Teeja Mary Sebastian is doing M.Tech (Master of Technology) in Computer Technology & Application from Delhi Technological University, Delhi, India and has done her B.Tech (with Distinction) also in Computer Science and Engineering; she is currently working as a scholar in the field of Sentiment Analysis.

Int J Syst Assur Eng Manag
DOI 10.1007/s13198-012-0112-y

ORIGINAL ARTICLE

System modeling and analysis of a combined cycle power plant

Nikhil Dev · Samsher · S. S. Kachhwaha

Received: 30 January 2012 / Revised: 30 January 2012

Abstract The performance of a combined cycle power plant (CCPP) and cost of electricity generation per unit is a function of its basic structure (i.e., layout and design), availability (maintenance aspects), efficiency (trained manpower and technically advanced equipments), cost of equipments and maintenance, pollutants emission and other regulatory aspects. Understanding of its structure will help in the improvement of performance, design, maintenance planning, and selection of new power generation systems. A mathematical model using the graph theory and matrix method is developed to evaluate the performance of a gas based CCPP. In the graph theoretic model, a directed graph or digraph is used to represent abstract information of the system using directed edges, which is useful for visual analysis. The matrix model developed from the digraph is useful for computer processing. Detailed methodology for developing a system structure graph, various system structure matrices, and their permanent functions are described for the combined cycle power plant. A top–down approach for complete analysis of CCPP is given.

Keywords CCPP · Digraph · Matrix method · Gas turbine · System structure · Graph theory

N. Dev (✉)
YMCA University of Science & Technology, Faridabad, Haryana, India
e-mail: nikhildevgarg@yahoo.com

Samsher · S. S. Kachhwaha
Delhi Technological University, Delhi, India

1 Introduction

Research in engineering analysis usually starts with an understanding of the physical system, then the adoption of a suitable mathematical model for the system. Combined cycle power plant is a very large and complex system. Performance of its components and subsystems are integrated with each other in such a manner that any of its component or system can't be studied without taking the effect of others. Such as the performance of steam turbine not only depends upon steam turbine efficiency but it is affected by HRSG and water system performance also.

For the complete analysis of CCPP a multi attribute decision making technique is required which can study the effect of one parameter on the others. Further the technique should be capable of integrating the measurable quantity such as cost and efficiency with virtual quantity such as quality. Thermodynamic model of a combined cycle power plant gives solution for optimum efficiency. Thermoeconomic model is useful for the analysis of CCPP on efficiency and cost basis in association with each other. For the complete analysis of CCPP thermodynamic, economic, reliability, human being and environment aspects have to be studied in association with each other.

At present there is no effective mathematical model for studying these aspects in association with each other for CCPP. Simple model, easy to implement, lesser computational cost and flexible with changing environment is required to evaluate the combined cycle power plant performance on design basis or analysis basis. From the literature it is clear that graph theory and matrix approach as a decision making method is relatively new, and offers a generic, simple, easy, and convenient decision making method that involves less computational efforts.

Published online: 04 August 2012

The advanced theory of graphs and their applications are well documented (Harary 1969; Deo 2000; Gandhi and Agrawal 1996; Mohan 2003; Grover et al. 2000; Gandhi et al. 1991; Grover et al. 2006; Mohan et al. 2008; Rao 2004; Rao and Gandhi 2001; Rao and Padmanabhan 2006; Wani and Gandhi 1999). Graph theory has been extensively applied in numerous disciplines of science and technologies.

Venkataswamy and Agrawal (1997) applied graph theoretical approach to evaluate and analyze the quality of the automotive vehicle by considering the characteristics of the vehicle. Mohan et al. (2006) calculated real-time efficiency index (RTEI) for a steam power plant which is the ratio of the values of variable permanent system structure function (VPF) in realtime (RT) situation to its achievable design value.

Tamassia (1999) survey selected research trends in graph drawing, and overview some recent results of the author and his collaborators. Rao (2006) presents a methodology for material selection for a given engineering component using graph theory and matrix approach. Mohan et al. (2003) develop a mathematical model using graph theory and matrix method to evaluate the performance of a steam power plant. For developing a system structure graph (SSG), six systems for boiler viz. air system, water system, combustion chamber/furnace, flue gas system and superheated steam system and their interactions are considered. The methodology converts a real life steam power plant into a block representation and then to a graph theoretic representation.

Yadav et al. (2011) carried out the quality evaluation of a Gas turbine system by considering its different characteristics which govern the qualitative aspects of the gas turbine system adopting graph theoretic approach. Garg et al. (2006) developed a deterministic quantitative model based on graph theoretical methodology to compare various technical and economical features of wind, hydro and thermal power plants and is used to evaluate and rank the power plants in ascending or descending order in accordance with the value of their suitability index.

During literature survey, author did not come across any work using Graph Theory for assessing the CCPP performance at component or system level.

Graph theory is a logical and systematic approach useful for modeling and analyzing various kinds of systems and problems in numerous fields of science and technology. If the graph/digraph is complex, it becomes difficult to analyze it visually. This can be done by computer through the use of the matrix method. It is a three stage unified systems approach.

- Modeling of system and subsystem in terms of nodes and edges gives a structural representation in the form of directed graph. This representation is suitable for visual analysis and gives a better understanding of interrelationships among system and subsystems.
- For further analysis, digraph representation is converted to matrix form, which makes it suitable for computational analysis. In the matrix value of each element assigned is based on the inheritance of element itself and its impact on the other performance parameters. However the matrix representation is not unique as changing the labeling of nodes can change it.
- Matrix model is modified according to the suitability of graph theory and results in permanent function model, which is in the expression form. Simplified permanent function expression is represented in terms of a single numerical index which is the indication of system performance.

Digraph representation, matrix representation and permanent function can be developed for the quality, cost, reliability, environmental and efficiency characteristics of a combined cycle power plant. Selection of a new power plant is a complex function of the factors effecting the efficiency, reliability and cost of the plant. Factors related to the efficiency, reliability and cost may be identified either with the help of thermodynamic model of power plant or with the help of experts in this area.

This paper presents a mathematical model using graph theoretic systems approach that enables the prediction of the performance of an CCPP by taking into account the performance of its six systems, that is, the air compressor, combustion chamber, gas turbine, HRSG, water system and steam turbine and the interactions between various systems.

2 Graph theoretic approach

There are number of factors responsible for effecting combined cycle power plant performance. The performance of plant depends upon the inheritance of these factors and the amount of interactions present between them, which conventional representations are unable to analyse. These interactions may be direction dependent or independent. The network showing these factors and interactions is proposed to study the CCPP performance and is called a digraph representation. If interactions are not direction dependent, the CCPP performance is represented by an undirected graph; if direction dependent, it is called a digraph representation.

The graph theoretic representation is suitable for visual analysis, it can be computer processed and can be expressed as a mathematical entity, where as the conventional representations, like block diagrams, cause and effect diagrams and flow charts, although providing visual analysis, do not depict interactions among factors and are not suitable for further analysis and cannot be processed or

expressed in mathematical form, whereas, the digraph is the starting point for further analysis in the graph theoretical methodology. The six systems, their critical elements and their sub elements are used to evaluate the CCPP performance for an index known as the CCPP performance index.

CCPP performance index $= f(\textbf{six systems})$

The endeavor here is to co-relate these six systems, their quantification based on their performance and interdependencies of six systems. Based on the above quantification it is proposed to find the performance of CCPP.

As stated earlier, it is achieved through a structural approach called graph theoretical methodology. The graph theoretical methodology consists of the digraph representation, the matrix representation and the permanent function representation. The digraph is the visual representation of the characteristics and their interdependence. The matrix converts the digraph into mathematical form. The permanent function is a mathematical model that helps to determine index. The main features of CCPP Performance Index (CPI) are as follows:

1. This index is a means to evaluate the factors and sub-factors affecting the combined cycle power plant performance.
2. Inheritance of each factor contributing in combined cycle power plant performance is represented by a single numerical value or index. Values of index indicate the relative importance of each factor and suggest to improve the weak factors.
3. Index value may be used for the comparison of combined cycle power plant performance under varying sets of factors and sub-factors.
4. Performance of two power plants may be compared on design or performance basis and it may help in deciding selection criteria for the new plant.

2.1 Digraph representation

A CCPP performance digraph is prepared to present the effect of one system on the others in terms of nodes and edges. Let nodes represent system and edges represent their interactions. It represents CCPP systems (S_i's) through its nodes and dependence of factors (c_{ij}'s) through its edges. S_i indicates the inheritance of systems and c_{ij} indicates degree of dependence of *j*th system on the *i*th factor. In a digraph c_{ij} is represented as a directed edge from node *i* to node *j*. The digraph permits to show the proposed CCPP systems and interactions between systems.

In particular six systems identified form the CCPP performance digraph. If all the six systems of CCPP are linked to each other then digraph for the CCPP will be as shown in Fig. 1.

This graph theoretic representation permits the incorporation or deletion of any interconnection or system in order to make it closer to a real life combined cycle power plant based on different design and principles in any given situation. This presentation is very suitable for visual analysis but is not very suitable for computer processing. Moreover if the system is large, its corresponding graph is complex and this complicates its understanding visually. In view of this, it is necessary to develop a representation of the CCPP that can be understood, stored, retrieved and processed by the computer in an efficient manner.

2.2 Matrix representation

Since, digraph is a visual representation; it helps in analysis to a limited extent only. To establish an expression for system performance, the digraph is represented in matrix form, which is convenient in computer processing also.

Many matrix representations for example adjacency and incidence matrix available in literature (Deo 2000; Harary 1969). The adjacency matrix is a square matrix and is selected for this purpose. An adjacency matrix is a means of representing which vertices (or node) of a graph are adjacent to which other vertices. Specifically, the adjacency matrix of a finite graph S on n vertices is the $n \times n$ matrix where the non-diagonal entry a_{ij} is the number of edge from vortex i to vortex j, and the diagonal entry a_{ii}, depending on the convention, is either once or twice the number of edges (loops) from vortex i to itself. Using this, the SSG of CCPP is represented in the matrix form.

The structural expansion, a characteristic of the CCPP system is obtained and is useful for its exhaustive analysis. It being a matrix and multinomial has been derived subsequently based on connectivity only, neglecting the directional properties. The SSG can be represented by either a variable characteristic system structure matrix or a variable permanent system structure matrix (VPSSM) (Mohan et al. 2004). Out of the two, VPSSM is chosen because it comprises of only the positive signs in its expansion which ensures that no information about characteristic values (like reliability, efficiency, etc.) is lost.

Variable permanent process performance matrix (VPPPM) for the CCPP is defined and developed below and is obtained by the following expression

$$E_N = [D_N] + [F_N]$$

where F_N is a variable process interconnection matrix, represented by c_{ij}'s as off-diagonal elements that represent the interconnection among the processes.

D_N is a diagonal matrix, with diagonal elements S_i representing variable performance characteristic value of

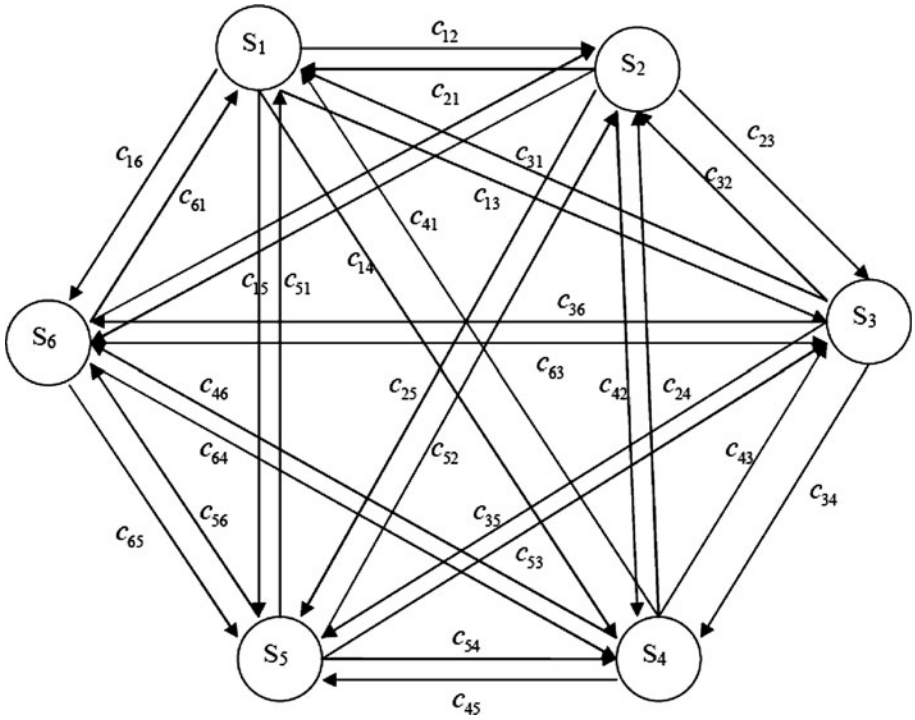


Fig. 1 Digraph showing six attributes of combined cycle power plant and their interdependencies in the system (attributes: air compressor system (S_1), combustion chamber system (S_2), gas turbine system (S_3), heat recovery steam generator system (S_4), steam turbine system (S_5), water system (S_6))

the ith processes. In case we have to write variable permanent system structure matrix (E_N), for a general case with n systems in that case it would be represented as

interaction among them that is c_{ij}'s for a general case having n systems. Permanent of this is called Variable System Structure Permanent Function, abbreviated as VPF-

$$E_N = [D_N] + [F_N] = \begin{array}{c} \begin{matrix} i & j & k & -- & m & n & \text{Systems} \end{matrix} \\ \begin{bmatrix} S_i & c_{ij} & c_{ik} & -- & c_{im} & c_{in} \\ c_{ji} & S_j & c_{jk} & -- & c_{jm} & c_{jn} \\ c_{ki} & c_{kj} & S_k & -- & c_{km} & c_{kn} \\ -- & -- & -- & -- & -- & -- \\ c_{mi} & c_{mj} & c_{mk} & -- & S_m & c_{mn} \\ c_{ni} & c_{nj} & c_{nk} & -- & c_{nm} & S_n \end{bmatrix} \begin{matrix} i \\ j \\ k \\ -- \\ m \\ n \end{matrix} \end{array} \quad (1)$$

The diagonal elements represent the contribution of subsystems in CCPP performance and the off diagonal elements represent interdependencies of each element in the matrix. The contribution can be expressed quantitatively and is explained later in this paper.

2.3 Permanent representation

It may be noted that above matrix represents different values of the characteristic of the system S_i's and the n. It is the characteristic of the system as it contains a number of terms which are its invariants. It is a standard function and is used in combinatorial mathematics.

Use of this concept in the system characteristic of CCPP will help in representing structural information of CCPP from combinatorial considerations. Moreover in using this matrix (that is VPSSM), no negative sign will appear in the expression, and thus avoids loss of any information. VPF-n of the matrix (1) is written as in sigma form as:

$$
\begin{aligned}
Per(E_N) = {} & \prod S_i + \sum_i \sum_j \sum_k \cdots \sum_m \sum_n c_{ij}^2 S_k S_l S_m S_n \ldots + 2 \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij} c_{jk} c_{ki}\right) S_l S_m S_n \ldots \\
& + 2 \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij} c_{jk} c_{kl} c_{li}\right) S_m S_n \ldots + \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij}^2 c_{ij}^2\right) S_m S_n \ldots \\
& + 2 \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij} c_{jk} c_{kl} c_{lm} c_{mi}\right) S_n \ldots + 2 \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij} c_{jk} c_{ki}\right) c_{lm}^2 S_n \ldots \\
& + \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij}^2\right) \left(c_{kl}^2\right) \left(c_{mn}^2\right) \ldots + 4 \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij} c_{jk} c_{ki}\right) \left(c_{lm} c_{mn} c_{ln}\right) \ldots \\
& + 2 \sum_i \sum_j \sum_k \cdots \sum_m \sum_n \left(c_{ij} c_{jk} c_{kl} c_{li}\right) \left(c_{mn}^2\right) \ldots
\end{aligned} \tag{2}
$$

The permanent of matrix (i.e. Eq. 1) is a mathematical expression in symbolic form. Equation 2 contains n! terms. The Eq. 2 contains terms arranged in N + 1 groups, where N is number of elements. The physical significance of various grouping is explained as under:

- The first term (grouping) represents a set of N unconnected elements, i.e. S_1 S_2 $S_3 \ldots S_N$.
- The second grouping is absent in absence of self-loops.
- Each term of the third grouping represents a set of two-element loop (i.e. c_{ij} c_{ji}) and is the resultant dependence of characteristics i and j unconnected elements.
- Each term of the fourth grouping represents a set of 3 elements loop (c_{ij} c_{jk} c_{ki} or its pair c_{ik} c_{kj} c_{ji}) is measure of remaining N-3 unconnected elements.
- The fifth grouping contains two subgroups. The terms of first sub grouping consist of 2 elements loop (i.e. $c_{ij}c_{ji}$ and $c_{kl}c_{lk}$) and component (S_i). The terms of second grouping are a product of 4- element loop (i.e. c_{ij} c_{jk} c_{kl} c_{li}) or its pair (i.e. c_{il} c_{lk} c_{kj} c_{ji}) and component (i.e. S_i).
- The terms of sixth grouping are also arranged in two sub-groupings. The terms of first sub-grouping are a product of a two element loop (i.e. c_{ij} c_{ji}) and a 3 element loop (i.e. c_{kl} c_{lm} c_{mk}) or its pair (i.e. c_{km} c_{ml} c_{lk}). The second sub-grouping consists of a 5 component loop (i.e. c_{ij} c_{jk} c_{kl} c_{lm} c_{mi}) or its pair (c_{im} c_{ml} c_{lk} c_{kj} c_{ji}).

2.4 Quantification of S_i's and c_{ij}'s

Quantification of the S_i's and c_{ij}'s (i.e. the diagonal and off-diagonal elements of VPM-CCPP) is required for evaluation of VPF-CCPP. The performance of each element (i.e. S_i's) is evaluated considering each S_i as a subsystem and graph theoretic approach is applied in each system. The various subfactors affecting the S_i are identified. The corresponding subsystem digraph is identified, the subsystem permanent matrix is written and the subsystem permanent function is evaluated. This will lead to inheritance of subsystem S_i. Similarly the other S_i's are evaluated considering subfactors affecting each S_i. In order to avoid complexity at subsystem level, values for inheritance may be taken from Table 1.

Obtain the values of the attributes and analyze their level of inter-dependencies on a normalized scale of 1–9. The inherent attribute value (i.e. S_i's) are generally calculated from the standard tests or retrieving the experimental data. Whenever, the quantitative data is not available, and then a criterion of ranked values by judgments over a scale of 1–9 is generally adopted.

Quantification of c_{ij}'s i.e. the off-diagonal elements representing interdependencies, are to be assigned numeric values. This dependence between the elements at system level or subsystem level cannot be measured directly. However, values can be assigned through proper interpretation by experts or experimental data.

The relative importance between the two attributes is assigned a value on a scale of 0–5 and is arranged into classes as mentioned in Table 2.

Table 1 Quantification of factors affecting combined cycle power plant performance

S. no.	Qualitative measure of factors at sub-system level affecting combined cycle performance	Assigned value of factors at sub-system level S_i
1	Exceptionally low	1
2	Very low	2
3	Low	3
4	Below average	4
5	Average	5
6	Above average	6
7	High	7
8	Very high	8
9	Exceptionally high	9

Table 2 Quantification of interdependencies/off diagonal elements

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

3 Development of graph theoretic model of combined cycle power plant

The graph theoretical methodology combines various attributes relevant to a thermal power plant into a single measure so that a comprehensive ranking of the two or more plants could be made on the basis of their performance. Combined cycle power plant being very large system has to be divided into small systems and no system should be independent. Inheritance of every system and their interdependency decides the value of CCPP performance index. In the following section system structure modeling of CCPP and its Graph Theoretic analysis is described.

3.1 System structure of combined cycle power plant

In this part system modeling of combined cycle power plant and its analysis is described. Combined cycle power plant considered for the present analysis is shown in Fig. 2.

The air at the ambient temperature is compressed by the air compressor and directed to the combustion chamber. The compressed air mixes with the natural gas from the fuel supply system to produce hot combustion gas in the combustor. The hot combustion gas is delivered to the gas turbine where the power is generated. The exhaust gas passes through a heat recovery steam generator where water is converted to high pressure steam. The high pressure steam from the boiler drives the steam turbine. The spent steam from the turbine flows into the condenser. The steam is separated in the boiler drum and supplied to the super heater section and the boiler condenser section. The super heated steam produced in the super heater then enters into the turbine through the turbine stop valve. After expansion in the turbine the exhaust steam is condensed in the condenser.

For the graph theoretic analysis combined cycle power plant is divided into following six systems:

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)

The system is modeled considering its structure on the basis of reliability, efficiency and cost. 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 = 1, 2, 3, 4, 5, 6) connecting the two vertices S_i and S_j.

In actual system all six system doesn't affect each other. The graph theoretic representation [S, c] of vertex and edge sets of the six system combined cycle power plant, called system structure graph (SSG) is shown in Fig. 3.

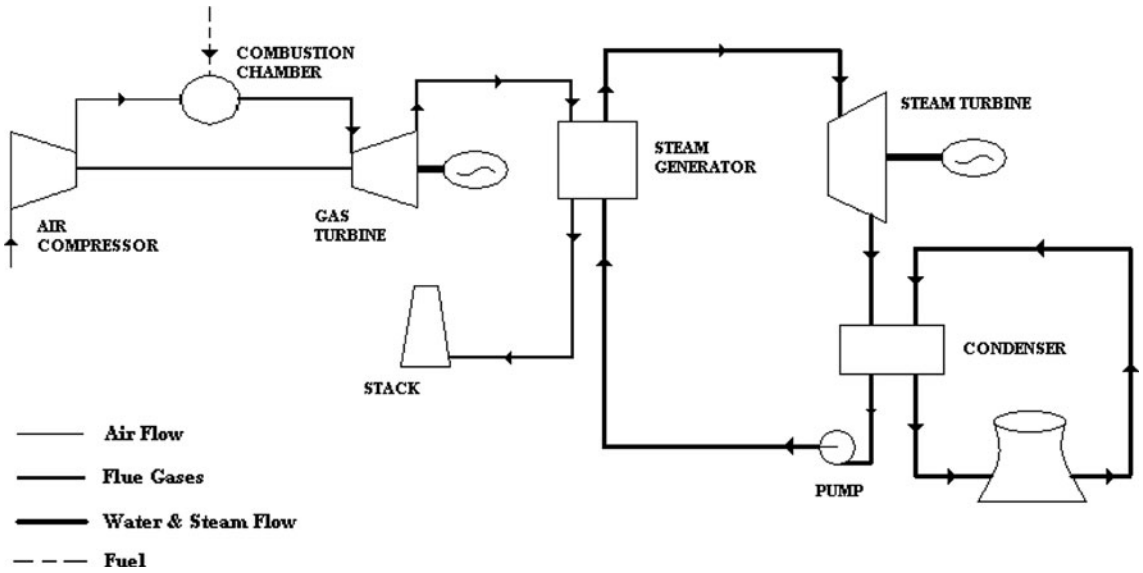


Fig. 2 Schematic flow diagram of combined cycle power plant

This is based upon the working of combined cycle power plant as per the following:

1. The ambient air comes to the compressor after being filtered by air filters. Compressor and turbine are attached with a shaft. So the power to compress the air comes to the compressor from the turbine. This is represented by the edge c_{31}. S_1 is the compression system.
2. After compression air goes to the combustion chamber. Fuel is added in the combustion chamber. This is represented by edge c_{12}.
3. A blade is cooled by being made hollow so that a coolant air can circulate through it. Coolant air is obtained directly from the compressor, thus bypassing the combustion chamber. Edge c_{13} represents the bypassing of cooling air.
4. Fuel supplied to the combustion chamber is generally compressed natural gas (CNG). For the analysis calorific value (CV) of the fuel is considered irrespective of the composition. Fuel supply is taken a part of combustion chamber. Outlet temperature of combustion chamber system [S_2] depends upon thermal stress limit of gas turbine blade material. Highest temperature of flue gas coming out from combustion chamber is controlled by changing air–fuel (A/F) ratio. Combustion product flows to gas turbine as shown by edge c_{23}.
5. Depending upon the temperature of flue gas, HRSG [S_4] may be used for (i) partial heating (regeneration) of the compressed air leaving the compressor (c_{42}), (ii) feed water heating of the steam cycle in a closed type feed water heater (c_{46}), or (iii) generating steam in a dual or multipressure steam cycle.
6. Flue gases coming out of combustion chamber and entering to HRSG system [S_4] are shown by the edge c_{34}.
7. As gas turbines are not yet built in sizes as large as steam turbines, combined cycle are often built in combination of more than one gas turbine plus one steam turbine. The steam turbine output is greater than gas turbine output by up to 8:1. High temperature and high pressure steam flows from HRSG system to steam turbine system [S_5]. This interaction is shown by edge c_{45}.
8. From steam turbine low pressure and low temperature steam comes to the cooling tower (c_{65}) and after condensation it goes to HRSG after passing the pump.
9. DM water, at initial start up is filled in the boiler water walls and drum and thereafter during normal operation only makeup water for the evaporation and condensation losses is given in the condenser hot well. Feed water is injected to control the temperature of superheated and reheated steam as an attemperation spray (c_{65}).
10. DM water from the water circuit [S_6] is fed to HRSG as feed water represented by edge c_{64}.

The system structure graph of Fig. 3 represents the internal structure of the CCPP at system level. It clearly shows different systems and their interconnections in the CCPP as discussed above.

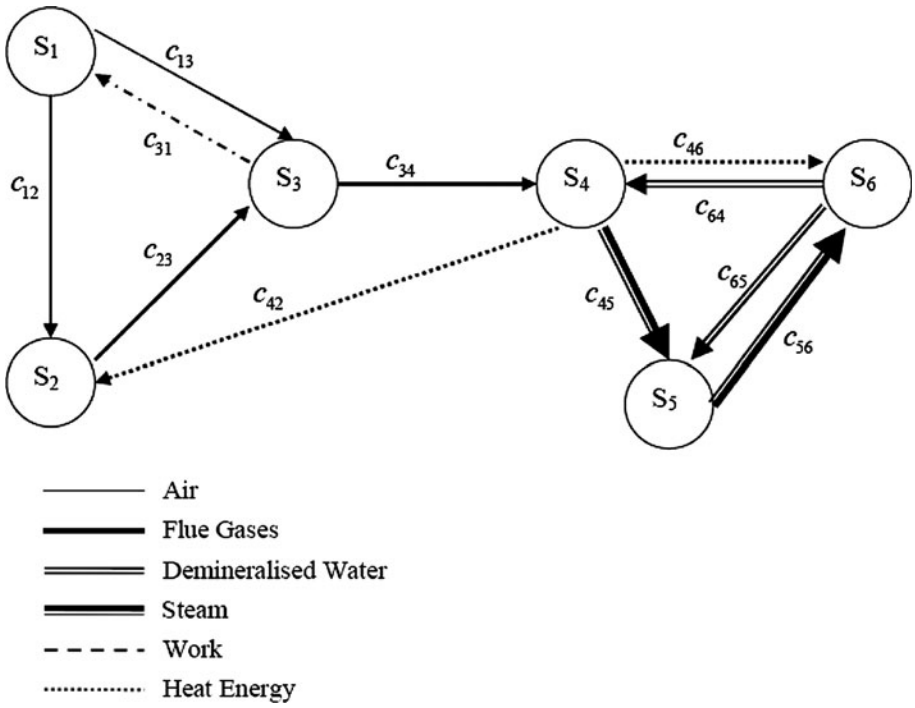


Fig. 3 System structural graph of combined cycle power plant: air compressor system (S_1), combustion chamber system (S_2), gas turbine system (S_3), heat recovery steam generator system (S_4), steam turbine system (S_5), water system (S_6)

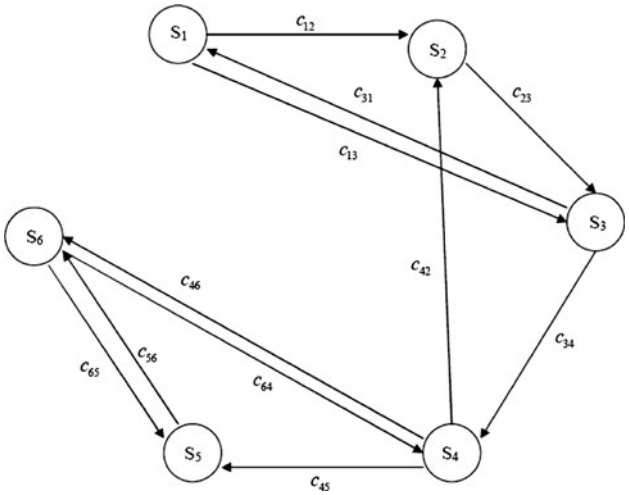


Fig. 4 Digraph showing six attributes of combined cycle power plant and their interdependencies in the system (attributes: Air compressor system (S_1), combustion chamber system (S_2), gas turbine system (S_3), heat recovery steam generator system (S_4), steam turbine system (S_5), water system (S_6))

3.2 Digraph representation of CCPP system

Digraph is used to represent the factors and their interdependencies in terms of nodes and edges. In an undirected graph, no direction is assigned to the edges in the graph. Whereas directed graphs or digraphs have directional edges. c_{ij} indicates degree of dependence of jth factor on the ith factor. In the digraph c_{ij} is represented as a directed edge from node i to node j. The digraph permits to analyse CCPP performance and gives a feeling of interactions between the systems. Diagraph for the CCPP discussed above is shown in Fig. 4.

3.3 Matrix representation of CCPP digraph

A directed graph (or a digraph) is nothing but a graph with directed edges. Firstly graphical model of the system is developed using graph theory which is named as attribute graph and after that graphical model is converted into matrix model called as attribute matrix. Finally this matrix is expressed in the form of a function called 'variable permanent function (VPF)'. The starting point in matrix representation is the adjacency matrix. Matrix representation of the CCPP attributes digraph gives one-to-one representation. Finally selected matrix should be flexible enough to incorporate the structural information of subsystem and interconnections between them.

3.3.1 System structural matrix

Let a general case of a system for example a CCPP having N systems be considered leading to symmetric adjacency matrix {0, 1} of order NXN and c_{ij} representing the connectivity between system i and j such as $c_{ij} = 1$, if system i is connected to the system j, [in the graph, this is represented by an edge (c_{ij}) between node i and j] and is equal to zero, otherwise. Thus $c_{ii} = 0$ for all i, as a system can not be connected to itself. In case it is connected to itself $c_{ii} = 1$. This implies a self loop at node i in the graph. Each row or column of the system structure matrix corresponds to a system. The off diagonal matrix elements, c_{ij} represent connection between system i and j. In this matrix $c_{ij} = c_{ji}$, as no directional consideration is taken into account and only connections between systems are considered. The possible interconnections between six systems of CCPP have been discussed in the last section. The adjacency matrix A_c for the CCPP is as following:

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

This (0, 1) adjacency matrix only shows interconnection between different systems. This matrix also does not contain any information about the six systems as diagonal terms are zeros. It simply represents connectivity amongst six systems of CCPP and does not consider directional features of connections. As this matrix considers only the connections between the systems and the characteristic features of the system do not come in the picture, a new matrix called 'Characteristic System Structure Matrix' is defined.

3.3.2 Characteristic system structure matrix

The presence of different systems of the CCPP is realized by defining a characteristic system structure matrix

$$B_c = \{SI - A_c\} \quad (4)$$

where I is the identity matrix and S is the characteristic of systems, representing its characteristic structural features. This matrix for system structural graph of CCPP is expressed as:

$$B_c = \{SI - A_c\} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{bmatrix} S & -1 & -1 & 0 & 0 & 0 \\ 0 & S & -1 & 0 & 0 & 0 \\ -1 & 0 & S & -1 & 0 & 0 \\ 0 & -1 & 0 & S & -1 & -1 \\ 0 & 0 & 0 & 0 & S & -1 \\ 0 & 0 & 0 & -1 & -1 & S \end{bmatrix} \end{array} \begin{matrix} Systems \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \quad (5)$$

This matrix is analogous to characteristic matrix in graph theory (Deo 2000; Harary 1969). The characteristics features of systems of a CCPP could be its efficiency, availability, and

cost and so on. It can be seen that this matrix contains information about the presence of system and their interconnections. It does not include information about attributes of the connections among different systems. The determinant of characteristic system structure matrix is called characteristic system structure polynomial and it is written as:

$$\det\{B_c\} = S^6 + 4S^4 - 3S^3 + 3S^2 - 4S + 1 \tag{6}$$

The characteristic system structure polynomial which is derived above is known as the invariant of the system and its value does not change by altering the labeling of systems. It is characteristic of the system structure.

However, the characteristic system structure matrix is not an invariant of the system, because a new matrix is obtained by changing the labeling of the systems. However, one matrix can be obtained from the other by proper permutation of rows and columns. From the above matrix, it is noted that a value of S is taken to be same for all the diagonal elements that is all the systems are considered to be identical. This is one of the reason that makes the characteristic polynomial non unique.

It has been reported in the literature (Harary 1969) 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 structural features. To incorporate distinct information of different systems and interconnections between them, a matrix called variable characteristic system structure matrix is proposed.

3.3.3 *Variable characteristic system structure matrix (VCSSM) of CCPP*

A variable system structure matrix T_a is defined taking into account the distinct characteristic of systems and their interconnection defined by the system structure graph. Let the off diagonal elements of a matrix, F_a, representing the connection between systems be denoted by c_{ij} instead of 1, whenever system i is connected to system j with $i, j = 1, 2, 3, 4, 5, 6$ and 0 otherwise. Let us also define a diagonal matrix D_a, with its variable diagonal elements S_i $(i = 1, 2, 3, 4, 5, 6)$ representing the characteristic structural feature of six distinct systems. For system structural graph of CCPP the VCSSM $T_a = [D_a - F_a]$ abbreviated as VCM-s is written as

$$T_a = [D_a - F_a] = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{bmatrix} S_1 & -c_{12} & -c_{13} & 0 & 0 & 0 \\ 0 & S_2 & -c_{23} & 0 & 0 & 0 \\ -c_{31} & 0 & S_3 & -c_{34} & 0 & 0 \\ 0 & -c_{42} & 0 & S_4 & -c_{45} & -c_{46} \\ 0 & 0 & 0 & 0 & S_5 & -c_{56} \\ 0 & 0 & 0 & -c_{64} & -c_{65} & S_6 \end{bmatrix} \end{array} \begin{matrix} \textit{Systems} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \tag{7}$$

The determinant of VCSSM is called variable characteristic system structural function (VCF-s) is written as

$$\begin{aligned} Per[T_a] = {} & [S_1S_2S_3S_4S_5S_6 + (c_{13}c_{31})(S_2S_4S_5S_6) \\ & + (c_{34}c_{43})(S_1S_2S_5S_6) + (c_{46}c_{64})(S_1S_2S_3S_5) \\ & + (c_{56}c_{65})(S_1S_2S_3S_4) - S_1S_2S_3(c_{45}c_{56}c_{64}) \\ & - S_1S_5S_6(c_{23}c_{34}c_{42}) - \text{x}\ S_4S_5S_6(c_{12}c_{23}c_{31}) \\ & + S_1S_2(c_{34}c_{43})(c_{56}c_{65}) + S_2S_4(c_{13}c_{31})(c_{56}c_{65}) \\ & + S_2S_5(c_{13}c_{31})(c_{46}c_{64}) - S_1(c_{56}c_{65})(c_{23}c_{34}c_{42}) \\ & - S_2(c_{13}c_{31})(c_{45}c_{56}c_{64}) - S_4(c_{56}c_{65})(c_{12}c_{23}c_{31}) \\ & - S_5(c_{46}c_{64})(c_{12}c_{23}c_{31}) + (c_{12}c_{23}c_{31})(c_{45}c_{56}c_{64})] \end{aligned} \tag{8}$$

Every term in the Per $[T_a]$ is representing the part of the system and their interlink. For example the $S_1S_2S_3S_4S_5S_6$ shows that all six systems are linked to each other. Any of the system cannot be omitted.

The multinomial (8), consist of variable structural components such as Si, c_{ij}, $c_{ij}c_{jk}$, $c_{ij}c_{jk}c_{kl}$ and so on. Here Si is the structural characteristic feature (SCF) of the system that is S_1 represents the SCF of the compressor system, S_2 represents the SCF of the combustion system, S_3 represents the SCF of gas turbine system and so on. By associating proper physical meaning to these structural components appropriate information about the system and their interlink is obtained.

The first term of the multinomial (8) is $S_1S_2S_3S_4S_5S_6$ and in it all the six systems are present. It may be considered as a set of six systems of CCPP. $c_{ij}c_{jk}c_{kl}$ is a 3-system structural loop. Only three such loops are visible in the SSG of CCPP. These are between the system (1, 2, 3), (2, 3, 4) and (4, 5, 6). Each term of the multinomial is considered as a set of different structural components. To read the expression (8) in a systematic manner, the terms are arranged in $N + 1$ grouping (with $N = 6$ in the example). The first grouping contains only one term and is $S_1S_2S_3S_4S_5S_6$ signifying that for the CCPP system to work or to exist all its systems must be in place. The second grouping is absent in the absence of self loop in the graph. This means, this grouping in the expression will appear if a system is connected to itself. The third grouping consists of number of terms and each is collection of two- system structural loop and four system characteristics structural features. The absence of the term containing loops such as c_{24}, c_{42} signifies absence of direct interaction of combustion chamber system and HRSG system.

The fourth grouping consists of a term which is a set of three system structural loop and three system characteristic structural features. The absence of term c_{14}, c_{15}, c_{16}, c_{24}, c_{26} etc. shows that there is no direct interaction between compressor system, combustion chamber system and HRSG system, steam turbine system and water system.

Each term of the fifth grouping is a collection of two two-system structural loop and two-system characteristic structural features. The terms of sixth grouping are a set of three-system structural loop, two system structural dyad and system characteristic structural features. The term of seventh grouping is a set of three two-system structural loops.

From the above, it is obvious that, in general, the *i*th grouping contains $(N + 1 - i)$ S*i*s and the remaining structural components (for example, c_{ij}, $c_{ij}c_{jk}$, $c_{ij}c_{jk}c_{kl}$, etc.). The terms of the expression are arranged in decreasing number of S*i*s for better interpretation of the system. The distinct structural components appear in any term and the distinct terms appear in any grouping of multinomial. Because of the arrangement of the terms of system structure in the same grouping, and distinct structures of different groupings and physical meaning behind each of these, the complete multinomial can be written by visual inspection of the graph. For finding the value of the diagonal attributes, that is, $S_1,\ldots,S_6$, and so on (for each of the systems of the CCPP), an SSG for these systems can be developed and the determinant value of the corresponding VCM-s can be calculated.

3.4 Variable permanent system structure matrix (VPSSM) for a CCPP

In order to obtain proper characterization of CCPP systems as derived from combinatorial considerations, a permanent matrix, also known as the permanent function, is proposed. Use of this concept in system structure modelling helps in retaining the structural information of the system.

The negative signs in Eq. (8) indicate subtraction of information about loops of systems, or system attributes such as reliability, efficiency, and so on, which will not project a true picture of the CCPP under analysis. For realistic understanding and characterization, a permanent function is proposed, as no negative sign will appear in the expression. Application of the permanent concept will thus lead to a better appreciation of the complete structure, in general. In view of the above, a variable permanent system structure matrix (VPSSM) T_c abbreviated as VPM-s for the combined cycle power plant is written as:

$$T_c = [D_c + F_c] = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & \textit{Systems} \end{matrix} \\ \begin{bmatrix} S_1 & c_{12} & c_{13} & 0 & 0 & 0 & 1 \\ 0 & S_2 & c_{23} & 0 & 0 & 0 & 2 \\ c_{31} & 0 & S_3 & c_{34} & 0 & 0 & 3 \\ 0 & c_{42} & c_{43} & S_4 & c_{45} & c_{46} & 4 \\ 0 & 0 & 0 & 0 & S_5 & c_{56} & 5 \\ 0 & 0 & 0 & c_{64} & c_{65} & S_6 & 6 \end{bmatrix} \end{array} \tag{9}$$

where the S_1s, c_{ij}s, $\mathbf{D_c}$, and $\mathbf{F_c}$ have the same meaning as in the matrix of expression (7). The permanent of VPSSM is called the variable permanent system structure function and is abbreviated as VPF-s. The only difference between matrices (7) and (9) is in the signs of the off-diagonal elements. In the VCSSM, expression (7), the off-diagonal elements c_{ij} have negative signs, while these are positive in the VPSSM of expression (9). VPF-s for matrix (9) is written as:

$$\begin{aligned} Per[T_c] = {} & [S_1S_2S_3S_4S_5S_6 + (c_{13}c_{31})(S_2S_4S_5S_6) \\ & + (c_{34}c_{43})(S_1S_2S_5S_6) + (c_{46}c_{64})(S_1S_2S_3S_5) \\ & + (c_{56}c_{65})(S_1S_2S_3S_4) + S_1S_2S_3(c_{45}c_{56}c_{64}) \\ & + S_1S_5S_6(c_{23}c_{34}c_{42}) + S_4S_5S_6(c_{12}c_{23}c_{31}) \\ & + S_1S_2(c_{34}c_{43})(c_{56}c_{65}) + S_2S_4(c_{13}c_{31})(c_{56}c_{65}) \\ & + S_2S_5(c_{13}c_{31})(c_{46}c_{64}) + S_1(c_{56}c_{65})(c_{23}c_{34}c_{42}) \\ & + S_2(c_{13}c_{31})(c_{45}c_{56}c_{64}) + S_4(c_{56}c_{65})(c_{12}c_{23}c_{31}) \\ & + S_5(c_{46}c_{64})(c_{12}c_{23}c_{31}) + (c_{12}c_{23}c_{31})(c_{45}c_{56}c_{64})] \end{aligned} \tag{10}$$

Comparing expressions (8) and (10), that is, the VCF-s and VPF-s, respectively, for the CCPP systems of Fig. 2, it is noted that all the terms are exactly the same in both expressions. However, they differ in their signs. A close look at the multinomial reveals that its various characteristic coefficients carry both positive and negative signs.

The variable characteristic function may not be able to provide the total objective value, when the numerical values for S_i and a_{ij} are substituted in the multinomial, because some of the information is lost by subtraction and addition operations in the determinant function. Considering these factors, the 'variable permanent function (VPF)' is defined. In VPF-s expression (10), all the terms carry positive signs, while in the VCM-s of expression (8) both positive and negative signs appear in the multinomial. Equations (8) and (10) uniquely represent the CCPP systems of Fig. 2.

Every term of these equations represents a physical subset of the system. It is possible to write these equations simply by visual inspection of the SSG of Fig. 3. To achieve this objective, the permanent function of Eq. (10) is written in standard form in $(N + 1)$ groups. The multinomial, that is, the permanent function when written down in $N + 1$ groups, present an exhaustive way of structural analysis of an CCPP at different levels. It helps in identifying critical components and links to improve reliability, efficiency, and cost of the system.

4 Example

For demonstration the proposed methodology, a CCPP is taken as an example. It is proposed to find the value of CCPP performance index. For determining the index we require numerical values of inheritance of all systems and

their interdependencies, i.e. in VPM-s for CCPP (Eq. (9)). Step by step methodology discussed in the previous section is used to evaluate CCPP performance index in this example.

1. Step 1. The various systems affecting the CCPP performance are identified in Fig. 3.
2. Step 2. The dependencies of systems are visualized through digraph shown in Fig. 4.
3. Tables 1 and 2 are used to determine numerical values for inheritance of systems and their interactions. It is suggested to find hypothetical best and hypothetical worst value of CCPP performance index. CCPP performance index is at its best when the inheritance of all its systems is at its best. Maximum value of Per [T_c] is obtained when inheritance of all systems is maximum, i.e. value taken from Table 1 is 9. The values of off diagonal elements are taken from Table 2 as explained below:

   - Performance of air compressor has a direct effect on the performance of combustion chamber and that is why c_{12} is assigned a value of 5.
   - Compressed air from the compressor is sent to the gas turbine for the cooling of gas turbine blades. This cooling air decreases the temperature of the combustion gases in the gas turbine. Their effect is not much pronounced and hence c_{13} is assigned with the value of 3.
   - Gas turbine and air compressor is attached with other and power to drive the compressor comes from the turbine. But there is no power transfer from combustion chamber to the air compressor. So c_{21} is assigned value 0 and c_{31} as 3. After combustion in combustion chamber flue gases enter the gas turbine. As the performance of gas turbine is dependent upon the condition of flue gases coming out of combustion chamber and gas turbine efficiency, a value of 5 is given to c_{23}.
   - As HRSG is at the outlet of gas turbine its performance is more effected by the gas turbine than combustion chamber. Due to this reason c_{34} is given a value of 4 and c_{42} is with 2.
   - Due to very high effect of steam condition on steam turbine c_{45} is assigned a value of 5.
   - There are water losses in the steam cycle. The extra amount of water comes from the water system. It (c_{64}) is taken as 3.
   - The temperature of steam entering the steam turbine is controlled by spraying extra amount of water. Steam turbine system and water system affects each other equally. So c_{65} and c_{56} are given a equal value of 3.
   - Some amount of steam is taken from the HRSG which goes to water system and helps in water purification. As this amount of steam is not very high, so c_{46} is assigned value of 2.

Thus, Eq. (9) may be written for the maximum value of Per T_c as

$$T_1 = \begin{bmatrix} 9 & 5 & 3 & 0 & 0 & 0 \\ 0 & 9 & 5 & 0 & 0 & 0 \\ 3 & 0 & 9 & 4 & 0 & 0 \\ 0 & 2 & 0 & 9 & 5 & 2 \\ 0 & 0 & 0 & 0 & 9 & 3 \\ 0 & 0 & 0 & 3 & 3 & 9 \end{bmatrix} \tag{11}$$

The value of permanent of the above function is 836865, i.e. maximum Per $T_c = 836865$.

Similarly CCPP performance index is at its worst when the inheritance of all its systems is at its worst i.e. value taken from Table 1 is 1.

Thus, Eq. (9) may be rewritten for the minimum value of Per T_c as

$$T_1 = \begin{bmatrix} 1 & 5 & 3 & 0 & 0 & 0 \\ 0 & 1 & 5 & 0 & 0 & 0 \\ 3 & 0 & 1 & 4 & 0 & 0 \\ 0 & 2 & 0 & 1 & 5 & 2 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 3 & 3 & 1 \end{bmatrix} \tag{12}$$

The value of permanent of the above function is 5585, i.e. minimum Per $T_c = 5585$.

The value of Per T_c indicates the value of CCPP performance index. Thus, the maximum and minimum value of performance index indicates the range with in which it can vary. Experts can use this range to decide a threshold value for a given set of similar power plant.

Monitoring at regular interval may be carried out by third party to assess power plant performance. Moreover, the values may be carried out at regular interval.

5 Conclusion

In this paper a methodology for evaluation of CCPP performance index is proposed using digraph and matrix approach. The graph theoretical methodology consists of the CCPP digraph, the CCPP matrix and the CCPP permanent function. The CCPP digraph is the visual representation of the characteristics and their interdependence. The CCPP performance matrix converts digraph into mathematical form. The CCPP permanent function is a mathematical model, which helps to determine CCPP index. Thus the approach helps to express CCPP performance in quantitative terms, which has more often been

expressed in qualitative terms. The procedure also helps to compare different industries in terms of efficiency, cost and reliability and rate them for a particular period of time.

To face the challenges of World Trade Organization and ease the opportunities to remain competitive in global business free from trade barriers, following salient features of proposed methodology are identified:

1. It is a qualitative cum quantitative method for modeling CCPP performance.
2. It permits modeling of interactions/dependencies existing between factors/subsystems.
3. CCPP performance can be represented by graph theoretic, matrix and permanent function models.
4. Performance is quantified by a single numerical index representing its competitiveness and suitability.
5. These models can easily be modified to consider new factors/subsystems emerging with technological development.
6. Sensitivity analysis to identify the critical elements is easily carried out.
7. The method permits to generate alternative for CCPP performance improvement.
8. This is an effective tool for evaluation, comparison, ranking and selection of an optimum CCPP system.

Practical implementation of the proposed methodology in a systematic manner will help industry to identify, analyse and evaluate factors responsible for CCPP performance. Evaluation and comparison will also lead to identify critical areas that are roadblocks to CCPP performance. The CCPP performance index not only help an organization to achieve intangible objectives- better efficiency, reliability through continuous improvement but also have long lasting effects on tangible objective – profitability through productivity.

References

Deo N (2000) Graph theory with applications to engineering and computer science. Prentice Hall India, New Delhi

Gandhi OP, Agrawal VP (1996) Failure cause analysis—a structural approach. J Press Vessel Technol Trans ASME 118:434–440

Gandhi OP, Agrawal VP, Shishodia KS (1991) Reliability analysis and evaluation of systems. Reliab Eng Syst Safety 32:283–305

Garg RK, Agrawal VP, Gupta VK (2006) Selection of power plants by evaluation and comparison using graph theoretical methodology. Electr Power Energy Syst 28:429–435

Grover S, Agrawal VP, Khan IA (2000) A digraph approach to TQM evaluation of an industry. Int J Prod Res 42:4031–4053

Grover S, Agrawal VP, Khan IA (2006) Role of human factor in TQM: a graph theoretic approach. Benchmarking: An Int J 13(4):447–468

Harary F (1969) Graph theory. Addison-Wesley, Reading

Mohan M, Gandhi OP, Agrawal VP (2003) Systems modeling of a coal based steam power plant. Proc Inst. Mech Eng A: J Power Energy 217:259–277

Mohan M, Gandhi OP, Agrawal VP (2004) Maintenance strategy for a coal based steam power plant equipment—a graph theoretic Approach. Proc Inst Mech Eng A J Power Energy 218:619–636

Mohan M, Gandhi OP, Agrawal VP (2008) Real time reliability index of steam power plant—a systems approach. Proc IMechE A J Power Energy 220:355–369

Rao RV (2004) Digraph and matrix methods for evaluating environmentally conscious manufacturing programs. Int J Environ Conscious Des Manuf 12:23–33

Rao RV (2006) A decision making framework model for evaluating flexible manufacturing systems using digraph and matrix methods. Int J Adv Manuf Technol 30:1101–1110

Rao RV, Gandhi OP (2001) Digraph and matrix method for selection, identification and comparison of metal cutting fluids. Proc IME J Eng Tribol 212:307–318

Rao RV, Padmanabhan KK (2006) Selection, identification and comparison of industrial robots using digraph and matrix methods. Robot Comput Integr Manuf 22:373–383

Tamassia R (1999) Advances in the theory and practice of graph drawing. Theor Comput Sci 217:235–254

Venkataswamy R, Agrawal VP (1997) A digraph approach to quality evaluation of an automatic vehicle. Qual Eng 9(3):405–417

Wani MF, Gandhi OP (1999) Development of maintainability index for mechanical systems. Reliab Eng Syst Saf 65:259–270

Yadav N, Khan IA, Grover S (2011) A graph theoretic approach to quality evaluation of typical gas turbine system. Int J Multidiscip Sci Technol 2(2)