Basavarajeswari Group of Institutions

BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

(Recognized by Govt. of Karnataka, Approved by AICTE, New Delhi & Affiliated to VTU, Belgaum)
"Jnana Gangotri" Campus, No.873/2, Bellary-Hospet Road, Allipur,
Bellary – 583 104 (Karnataka)
Ph: 08392-210566 / 588 / 599, Fax: 242900, e-mail: bitmbly@gmail.com, Web: www.bitm.edu.in

Proceedings of the National Conference
On

"Recent Trends in Electronics & Communication Engineering" (NCEC-2010)

NCEC-2010

24th & 25th September, 2010

Organized by:

Dept. of Electronics & Communication Engg.

NCEC-2010

MESSAGE

It is matter of pride to note that the Department of Electronics & Communication Engineering of Ballari Institute of Technology & Management, Bellary is organizing a *National Conference on "Recent Trends in Electronics & Communication Engineering*" (NCEC-2010) on 24th & 25th of September 2010.

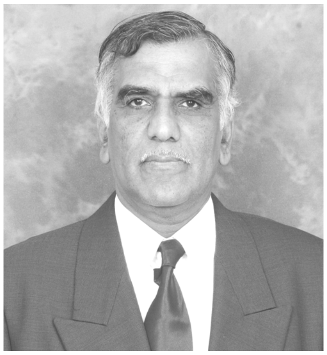
Advancement in technologies and resource keep changing the world day by day. They can be shared and made useful only when they are exchanged. This will also applicable to all engineering graduates and researchers. The conference of this type will bring academicians and researchers together and exchange the thoughts. The department brings out the proceedings and hope this will be useful to public at the large.

I take this opportunity to congratulate the organizers and wish the conference a grand success.

Dr. S. J. V Mahipal, MBBS, MD, MS
Chairman
T.E.H.R.D Trust, Bellary.

NCEC-2010

MESSAGE



I am very glad that NCEC-2010, the National Conference on Recent Trends in Electronics & Communication Engineering organized by the Department of Electronics & Communication Engineering is the path breaking initiative and will pave a way for future technological development in the field of ECE.

The conference NCEC-2010 aims at the new and recent developments in the various fields related to Electronics & Communication and will have tremendous impact on the society in future.

I am very happy to learn that NCEC-2010 has received excellent response from the academia and industry community all over India.

I hope this conference will help our faculty and students to acquire good knowledge in the latest technologies emerging in this field.

My best wishes to the organizers for the success of the conference.

Dr. Yashvanth Bhupal,
Chairman & Director /
Managing Trustee
Ballari Institute of Technology & Management,
Bellary

NCEC-2010

MESSAGE

It gives me great pleasure that our college is organizing National Conference on “Recent Trends in Electronics and Communication Engineering” by the Department of Electronics & Communication Engineering on 24th & 25th of September, 2010.

The faculty of Department of Electronics and Communication Engineering of this institute definitely deserve appreciation for their efforts in translating the mission of the institute into the reality.

The confluence of the best brains from academic institutions, scientific organization, industry and R&D organization will go on a long way in strengthening the various fields of Information Technology.

Hence, I am keenly looking forward to the outcome of the conference. I extend my sincere best wishes to the organizers for the success of the conference.

Prof. Prithviraj.Y.J.
Deputy Director & Trustee
Ballari Institute of Technology & Management,
Bellary

FOREWORD ..

It is with great expectations and prides that, we at Ballari Institute of Technology & Management look forward to NCEC-2010, the National Conference on “Recent Trends in Electronics and Communication Engineering” at this college on September 24th & 25th 2010.

We look at NCEC-2010, the National Conference organized by the Department of Electronics & Communication Engineering as a platform for the expressions of latest innovations, sharing of experience and learning new developments in the area of Electronics & Communication. The interaction between industry and academia under the aegis of this conference is expected to develop the understanding which will be beneficial to both the sections.

I look forward to this conference with anticipation of better results and increased interaction between the minds that have focused on similar fields related to information technology.

I congratulate and thank the faculty members and students for having taken untiring efforts to bring the conference a remarkable one. With the dedication and enthusiasm of my faculty, I am sure that this conference will be a great success.

Dr. U. Eranna
Principal
Ballari Institute of Technology & Management,
Bellary

NCEC-2010

PREFACE

The “National Conference on Recent Trends in Electronics & Communication Engineering” (NCEC-2010) organized by the Department of Electronics & Communication Engineering aims at bringing together researchers, academicians and industry under common platform to exchange their research findings, exploring the technological advances and highlighting current issues and future directions in the field of Electronics & Communication Engineering.

The NCEC–2010 has received good attention from all over India. It attracted around 250 papers and out of which experts accepted 100 papers for publication into the Proceedings of the Conference after thorough review. The responsibility of correctness of peer reviewed research papers, which are included in these proceedings, rests on authors not on organizers of NCEC-2010.

We take this opportunity to thank Dr. Giri NK Rangan, Managing Director, Intersil Corporation, Bengalooru and Dr. K. V. A. Balaji, Registrar, VTU, Belgaum for their great support towards the conference.

We deeply acknowledge the encouragement and support provided by Dr. Yashvanth Bhupal, Chairman & Director / Managing Trustee, BITM, Prof. Prithviraj.Y.J. Deputy Director / Trustee, BITM and Dr. U. Eranna, Principal, BITM, Bellary.

We heartily thank all the Advisory Committee Members, Technical Committee Members and Reviewers for their kind co-operation in conducting the Conference successfully.

Organizing Committee

NCEC-2010

ORGANIZING COMMITTEE

Chief Patron:

Dr. Yashvanth Bhupal
Chairman & Director / Managing Trustee

Patrons:

Prof. Prithviraj.Y.J.
Deputy Director / Trustee

Dr.U. Eranna
Principal

Conveners:

Prof. V.C.Patil
Prof. & HOD
Dept. of Electronics & Communication Engg.,

Prof. K. M.Sadyojatha
Dept. of Electronics & Communication Engg.,

Co Conveners:

Prof. Shesikala, Prof. Dept. of ECE
Prof. U.M. Rohita, Associate Prof., Dept. of ECE
Prof. Premchand, Associate Prof., Dept. of ECE
Prof. Nagabhusan Katte, Associate Prof., Dept. of ECE
Prof. Manjula R, Asst. Prof., Dept. of ECE

ADVISORY COMMITTEE

Dr. B. S. Anami
Principal, KLEIT, Hubli

Dr. Subhash S Kulkarni
Principal, JPNCE, Mahabubnagar (AP)

Dr. R. R. Mudholkar
Shivaji University, Kolhapur (Maharashtra)

Dr. Vinayadatt V Kohir
HOD E&CE, PDACE, Gulbarga

Dr. Ramamurthy
S K University, Anantapur (AP)

Dr. K Bhanuprasad
Principal SRPEC, Nalgonda (AP)

Dr. Siddarama Patil
PDACE, Gulbarga

Prof. Sandeep V.M
JPNCE, Mahabubnagar (AP)

Sri. K. V. Manjunath
Director, Global Network Services, Bengalooru

Dr. S. Ganga Shetty
IIIT, Hyderabad (AP)

Sri. B. Nagaraj Reddy
CG-Corel, Bengalooru

NCEC-2010

KEYNOTE SPEAKERS

Dr. Giri NK Rangan
Managing Director, Intersil Corporation, Bengalooru

INVITED TALK

Dr. Vinayadatt V Kohir
HOD E&CE, PDACE, Gulbarga

Dr. R. R. Mudholkar
Shivaji University, Kolhapur (Maharashtra)

Dr. Siddarama Patil
PDACE, Gulbarga

Sri. K. V. Manjunath
Director, Global Network Services, Bengalooru

Sri. B. Nagaraj Reddy
CG-Corel, Bengalooru

NCEC-2010

PROGRAMME CO-ORDINATORS

Prof. V C Patil, Prof. & HOD
Prof. K M Sadyojatha, Professor
Prof. G Seshikala, Professor
Prof. U M Rohit, Associate Professor
Prof. Premchand D R, Associate Professor
Dr. Nagabhushan .B. Katte, Associate Professor
Manjula R, Asst. Professor
Ghousia Begum, Asst. Professor
K. S. Shivakumar, Asst. Professor
U. Rajashekar, Asst. Professor
Bharani Rao, Asst. Professor
Nilam Chheda, Asst. Professor
Renuka Sagar, Asst. Professor
Mallikarjuna A, Asst. Professor
Parvathi P, Asst. Professor
Naseeruddin, Asst. Professor
Prabhakar K, Asst. Professor
Prabha. K, Asst. Professor
Sowbhagya, Asst. Professor

Technical Staff

Vishnukanth Karwa, Supervisor
Manjunath Sindigi, Instructor
Hiremathada Shiva Basavaraj, Instructor
Kesarakoni Anasuya, Instructor
Dept., of Electronics & Communication Engg.
BITM, Bellary.

Designed by: B.Basavaraj, Superintendent

NCEC-2010

About Institution

Ballari Institute of Technology & Management, Bellary started under the aegis of T.E.H.R.D. Trust, Bellary in the year 1997. The T.E.H.R.D. Trust and all the institutions at various levels run by it owe their today’s excellent reputation to Late.Smt.Basavarajeswari, Former Union Minister & Founder Chairperson of the Trust. The Institution is recognized by Govt. of Karnataka, approved by AICTE, New Delhi and affiliated to Visveswaraya Technological University, Belgaum.

BITM has many benchmarking activities to its credit. It has received ISO 9001:2008 Certification from IM Moody International Agency. The college is in the process of procuring National Board of Accreditation (NBA). Advances in Technology & Communications have brought the world closer than ever before. BITM plays a vital role in producing world class engineers tuned to the demands of fast changing world that can now be aptly called the “Global Village”. The technological changes and the constant need to stay ahead, make BITM one of Karnataka’s premier engineering institutions.

About Department

Department of Electronics & Communication Engineering established in the year 1997 has qualified & committed faculty members and has steadily grown from its humble beginning to a full-fledged department. It has well state-of-the-art laboratories and infrastructure. The department facilitates faculty to work as a team and update their knowledge and develop skills to match the industrial and technological requirements.

CONTENTS

*** *** ***

Fuzzy System Development - A General Approach

R. R. Mudholkar, Sudhakar Hegde, P. A. Kadam, S. S. Nirmale, A. C. Shaikh

Department of Electronics, Shivaji University, Kolhapur-416004.INDIA.

E-mail: neuralfuzzy.lab@gmail.com, snirmale03@gmail.com,

***Abstract*— This paper describes the general design algorithm for the development of fuzzy system in general sense. However with due modification and proper selection of fuzzy system parameter one can design and develop the fuzzy logic system suit best to this application and interest.**

***Keywords*—** LM,TSK,FLC

I. INTRODUCTION

The idea of modeling the human-thinking mechanism in terms of linguistic fuzzy values rather than numbers brought the fuzziness into the system theory and development of a new class of systems called- ‘Fuzzy Systems’. These are primarily based on the concept of Fuzzy coding or partitioning of information. In general, a fuzzy system is any system whose variables (or at least few of them) range over the fuzzy sets.Fuzzy System Models fall into two categories Linguistic Model (LM)
Takagi-Sugeno-Kang (TSK) Model. The first essentially describes the system behavior qualitatively using the natural language. The knowledge is transcribed in the form of *‘If - Then’* rules which derive the decisions through fuzzy or approximate reasoning. FLC (Fuzzy Logic Control) is prototypical example of a linguistic model.
The second is based on Takagi-Sugeno-Kang method of reasoning. This is based on logical rules with fuzzy antecedent and functional consequent. This model combines the fuzzy space with mathematical model. It was originally initiated by Zadeh and developed further by Tang and, Sugeno and Yasukawa.

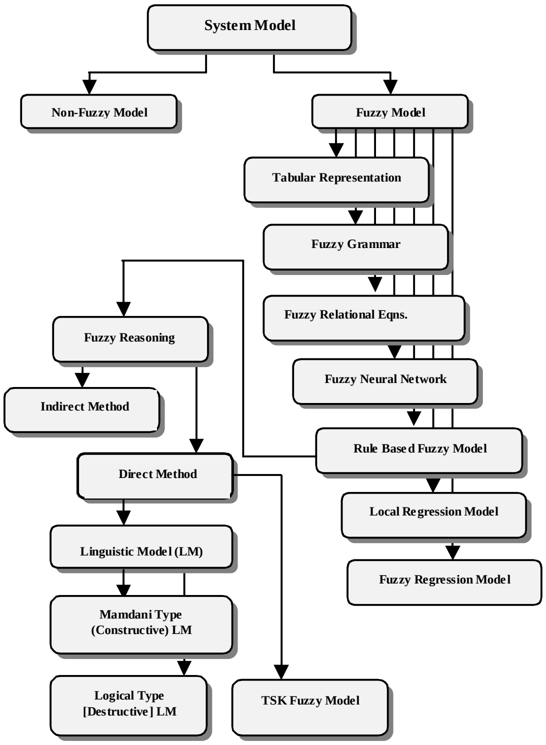


Fig.1: Classification of Fuzzy Models

Extensive literature is now available describing a systematic development of linguistic approach. The rules with vague predicates replace the arithmetical equations characterizing the behavior of the system. The decision-making ability of LM is determined by the proper interpretation of the system behavior that goes to the rule-base and reasoning mechanism. Most of LMs are founded on the approximate reasoning.
Lee (1990) has done a comprehensive classification of the methods of fuzzy reasoning and their theoretical and experimental analysis.

II. CLASSIFICATION

The categorical classification of fuzzy models depending upon the nature of structural dependencies between the system’s variables and level of knowledge captured by the specific model could be as shown in fig. 1.

III. BUILDING FUZZY SYSTEM MODEL

In this section we discuss the proto-cycling methodology of building Fuzzy System outlined in [1,4,7,8,9,12]. Initially a base system is formed that is refined, enhanced and finally extended to real-time problems. The schematic of fuzzy system development is shown in fig. 2.

Initial conceptual design is worked out on the paper. This is the critical phase of Fuzzy Model Development, where clear understanding of system behavior is made and dependencies of input-process and output-solution variables of system are worked out. To accept the model, it is verified through simulation, which can be either control surface driven or input data driven. The latter approach is widely practiced.

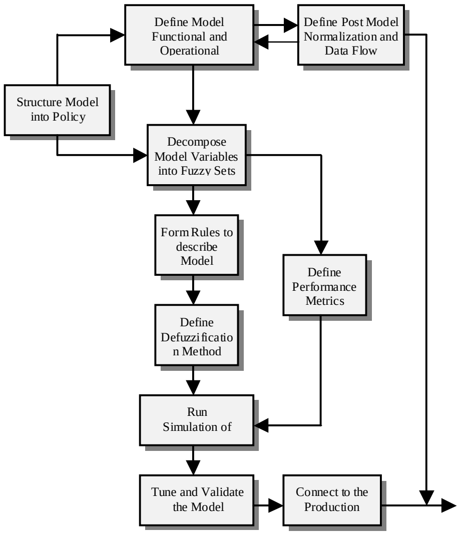

Define Model Functional and Operational
Define Post Model Normalization and Data Flow
Structure Model into Policy
Decompose Model Variables into Fuzzy Sets
Form Rules to describe Model
Define Performance Metrics
Define Defuzzification Method
Run Simulation of
Tune and Validate the Model
Connect to the Production

Fig.2: Fuzzy System Development phases

In designing a Fuzzy System many options exist than those with conventional counterpart. The designing and optimization of Fuzzy System has many degrees of freedom as illustrated in fig. 3. This lends good flexibility to the system designer in designing and optimizing his own Fuzzy System.

IV. DEVELOPMENT OF FUZZY BASED DESIGNING METHODOLOGY:

We intend to show the possibility and potentiality of Fuzzy Set Theory and Fuzzy Logic in the designing process of a transformer. To create comprehensive platform and acquaintance, first we explore the general design based methodology of fuzzy system followed by its application specific to the transformer designing.

Fuzzy Design Methodology in General:

For better establishment of architectural characteristics and easy development, the system is decomposed into sub-modules, sub-modules into policies as depicted in fig. 4.

e.g. Sub-module-2 can be fuzzification module and policy-2 can be decision regarding the better choice of the shape of membership.

Generally fuzzy modules work from Out side-In; i.e. first a good understanding of system behavior is developed and then trial simulations are carried out with respect to fuzzy surface, followed by performance optimization. Thus procedure is best applied from policies to sub-modules to overall fuzzy module.

The fundamental phases involved in the designing of fuzzy system are as follows-

Phase - 1:

Identification of input/output variables

Determination of Operating Ranges [Universe of Discourse]

Normalization of Universe of Discourse [Scaling Factor]

Phase - 2:

Expressing the variables by appropriate Fuzzy Sets [Fuzzy Quantization]

Introduction of different membership functions [Shapes]

Partition or Decomposition of Universe of Discourse

Selection of Linguistic Labels [Term Set] for each variable

Phase - 3:

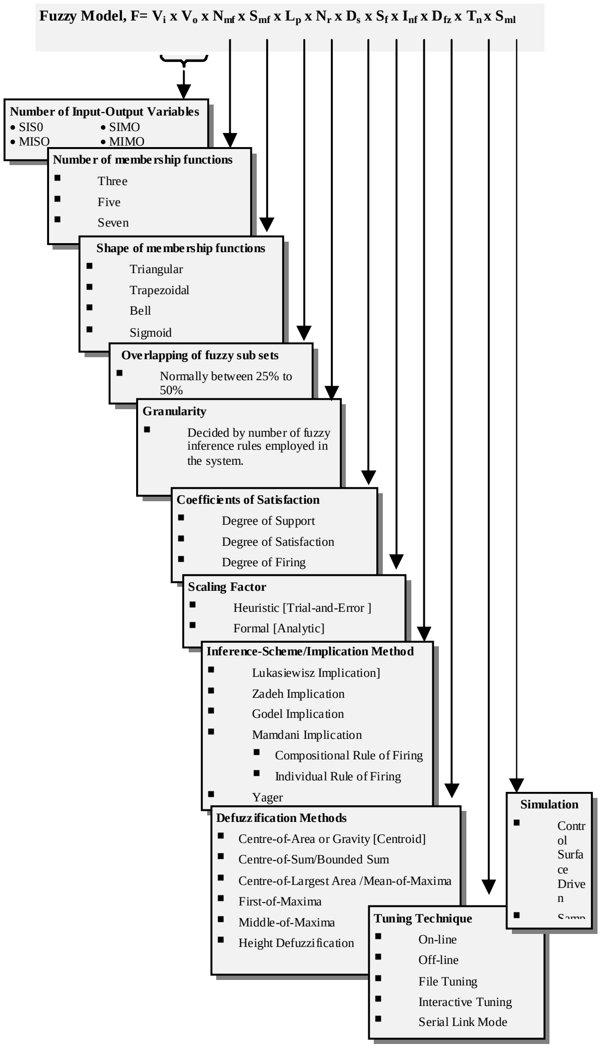

Fuzzy Model, F= V_i x V_o x N_{mf} x S_{mf} x L_p x N_r x D_s x S_f x I_{nf} x D_{fz} x T_n x S_{ml}
Number of Input-Output Variables
• SIS0
• SIMO
• MISO
• MIMO
Number of membership functions
▪ Three
▪ Five
▪ Seven
Shape of membership functions
▪ Triangular
▪ Trapezoidal
▪ Bell
▪ Sigmoid
Overlapping of fuzzy sub sets
▪ Normally between 25% to 50%
Granularity
▪ Decided by number of fuzzy inference rules employed in the system.
Coefficients of Satisfaction
▪ Degree of Support
▪ Degree of Satisfaction
▪ Degree of Firing
Scaling Factor
▪ Heuristic [Trial-and-Error]
▪ Formal [Analytic]
Inference-Scheme/Implication Method
▪ Lukasiewisz Implication]
▪ Zadeh Implication
▪ Godel Implication
▪ Mamdani Implication
▪ Compositional Rule of Firing
▪ Individual Rule of Firing
▪ Yager
Defuzzification Methods
▪ Centre-of-Area or Gravity [Centroid]
▪ Centre-of-Sum/Bounded Sum
▪ Centre-of-Largest Area /Mean-of-Maxima
▪ First-of-Maxima
▪ Middle-of-Maxima
▪ Height Defuzzification
Tuning Technique
▪ On-line
▪ Off-line
▪ File Tuning
▪ Interactive Tuning
▪ Serial Link Mode
Simulation
▪ Control Surface Driven
▪ Samp

Fig. 3: Outline of options available with Fuzzy System

Formulation of knowledge pertaining to the problem in terms of fuzzy inference rules.

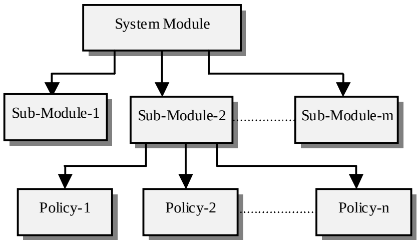


Fig. 4: Decomposition of Fuzzy System

Rules elicitation from experienced human operator and/or designer.

Rule extraction from empirical data based on their trends.

Phase – 4:

Choice of Inference Scheme/ Implication method

Phase - 5:

Defuzzification of Solution Fuzzy Space
De-normalization of Universe of Discourse [Scaling Factors]
The basic structure of fuzzy model of a system is shown in fig. 5.

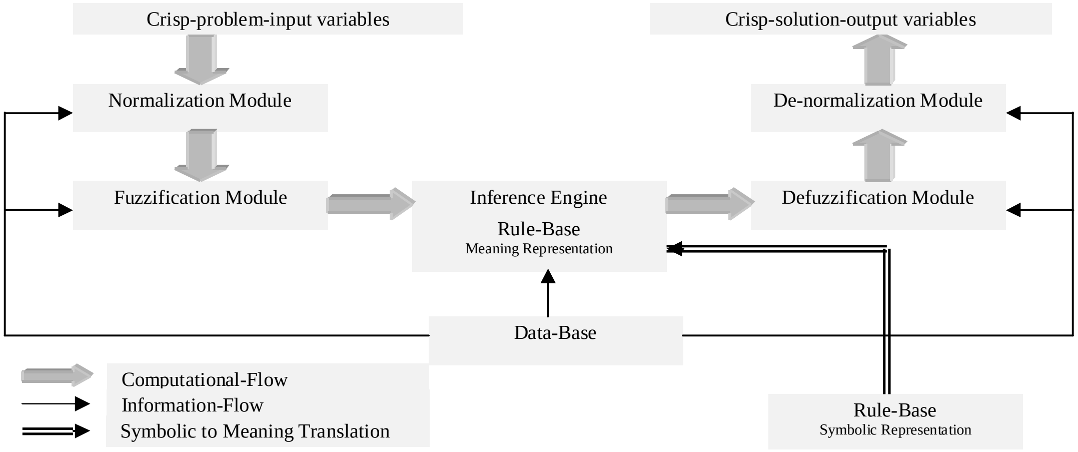


Fig. 5: Basic Structure of Fuzzy Module

Phase - 6:
Interpretation of Results

Phase-1:
Identification of Input-Output Performance Variables

This is the most difficult and tedious part of initial formulation of model. The variables that control the operation of fuzzy system and lead to desirable outputs are to be identified from the universe of continuous-discrete, qualitative-quantitative, spare-dense, certain-uncertain data after judiciously scrutinizing the problem space. The performance variables can be either endogenous or exogenous to the model.

This phase of design decides the following performance criteria-

Determination of Operating Ranges

After identifying the input-process and output-solution variables, their ranges of applicability are worked out. These ranges are called 'Operating-Domains'. The variables are described within their fuzzy spaces that are generally composed of multiple overlapping fuzzy sets.

The total fuzzy space form the smallest to largest allowable values is called 'Universe of Discourse' that constitutes the working space of a variable, where as the range of an individual fuzzy set is known as a 'Domain'.

While defining fuzzy sets, the domains are transferred directly to membership functions. For comfortable mapping between the input and output variables, the translation of variables is to be performed along similar domains. Otherwise, mis-scaling and mis-calibration of domains lead to the failure of model.

The designer of a fuzzy model for application of interest has to find the optimal and comfortable regions of operation over expected input-output data-spaces. The exact nature of input-output operational surfaces is generally explored with respect to fuzzy term set and tuned finally during simulation. At this juncture, the designer has to define limits of operational ranges for each variable.

Level of Granularity

Granularity of fuzzy model is something different than intrinsic data space granularity. Since, a fuzzy model can interpolate or interfuse input process to output solution spaces through approximation of actual space, it is possible to build fuzzy model at higher level of abstraction than the conventional mathematical model. In other words fuzzy model allows to define and modulate problem space without higher level of details and mathematical precision.

Normalization of Universe of Discourse and Scaling Factors

Through the scale transformation, domains (Universe-of-discourse) for variables are normalized. This allows the physical values of the input problem-variable map over normalized domains. This is called as 'Input-Normalization'. Likewise, output de-normalization maps the normalized value of the output solution-variables over the respective physical domain.

Scaling factors play a significant role in domain normalization and de-normalization processes. Ingenious selection of scaling factors contributes great to success of fuzzy model, and ease the model processing problem in the latter stages of simulation and validation.

There are basically two approaches of determining the scaling factors-

Heuristic approach
Formal (Analytic) approach

The first approach is based on 'trial-and-error' optimization and is popular in Fuzzy Control. In the second approach as an analytical relationship between the values of the scaling factors and overall solution-process is established. This is primarily used in the non-linear systems, whose conventional model exists.

Phase-2:
Definition and Construction of Fuzzy Sets

The fuzzy model involves the definition and construction of fuzzy sets that encode problem surface over which the process is mapped by the inference rules. Different methods of constructing fuzzy sets are employed to capture the meaning and fuzzy region, and provide anticipated mapping between domain values and their degrees of membership across the domain. The method of recognizing fuzzy attributes and drafting fuzzy sets is an important technique. The fundamental criteria involved in the construction of fuzzy sets are-

Shapes

Hedges

Reasoning-Process

The problem space of a variable is decomposed into a number overlapping fuzzy region. Each region is assigned a meaningful label-called term-name. It facilitates to refer the fuzzy sets during the inference-process and defuzzification. The collection of labels or term-names over universe of discourse is called 'Term-Set'.

Clever naming of fuzzy set has important implication for understandability, maintenance and model validation.

Incorrect mapping of related fuzzy regions conformably is a significant reason for failure of otherwise simple fuzzy system.

Decomposition and naming of fuzzy sets is shown in fig. 6 for variable- Efficiency.

Fuzzy Systems are tolerant to approximations not only in their problem space, but also in the representation of fuzzy sets. This reveals that fuzzy system perform well even the fuzzy sets do not map exactly with its model concept. This relatively insensitivity to the elastic target problem surface. Few approaches practiced towards generating fuzzy set are-

1. Probability frequency distribution
2. Neural network models
3. Mathematical surface sampling
4. Subjective approximation of problem surface
5. 'Voted for' distribution

In our model we have employed 3rd and 4th techniques to obtain the fuzzy sets. In the 3rd technique a fuzzy surface is determined by fuzzifying mathematical surface from existing mathematical model, while in the 4th technique fuzzy sets are developed from empirical data through a 'best-fit-estimation' process.

In a world of fuzzy all shapes representing fuzzy set have become the variation of the triangular shape. This shape has minimal representational requirements, yet lends maximum exploitation. Both triangular and trapezoidal sets can be represented using only four bytes- two for points in the surface (P1 and P2) and two more for slopes (S1 and S2) of fuzzy surface as shown in fig. 7.

In any case, while selecting the shape of fuzzy set, the trade-off between memory requirement and speed is made.

In spite of other shapes possible many variations in the representation of fuzzy sets makes the fuzzy model quite robust and resilient especially in the initial prototyping stage.

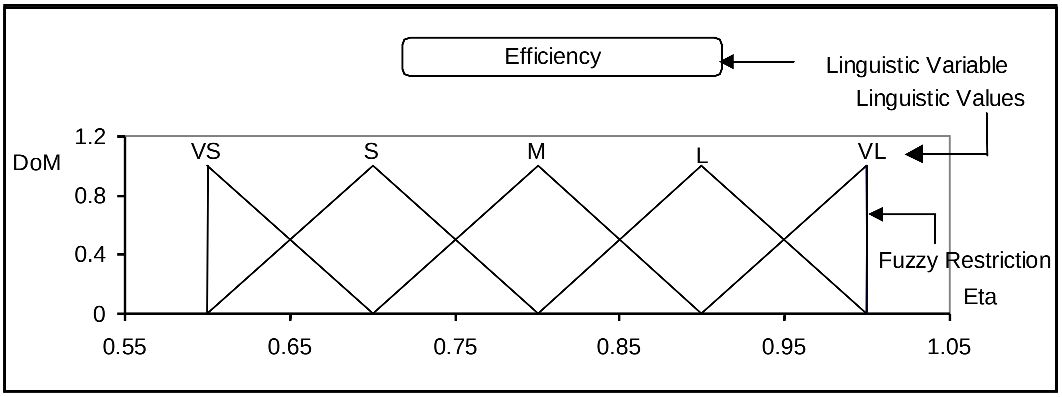


Fig. 6: Fuzzy Sets and Decomposition

Choice of shape for fuzzy set is a critical part of building a fuzzy model, since the shape of fuzzy set determines the correspondence between data and underlying concept. There are no fixed topologies to determine the shape for fuzzy set as mapping a domain and degree of membership of underlying concept can take any form, and elicitation should be elastic rather than restrictive.

Generally, Design Engineer and System Analyst construct fuzzy sets for the intuitive understanding of engineering applications continues to rely on triangular shape since they are -

Simple to specify

Easier to visualize

Convenient to map input-output spaces

Efficient in memory use

High in processing speed

Many applications are not overly sensitive to variations in the shape. In such cases it is convenient to use a simple shape of triangular.

Overlapping of Fuzzy Sets

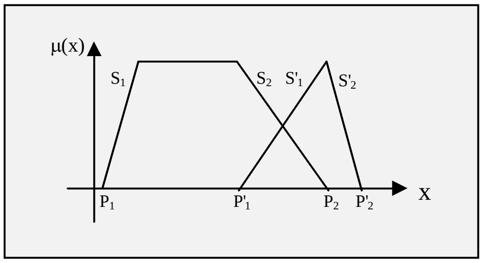


Fig. 7: Trapezoidal and Triangular Membership Functions

To convert a series of individual fuzzy regions into a single continuous and smooth surface, each fuzzy set must more or less overlap neighboring fuzzy sets. There are no definite rules and conditions as to decide the minimum and maximum degree of overlap. It is being the natural consequence of fuzziness and ambiguity of problem space, designer of fuzzy system has freedom to choose the any percentage of overlap. However a overlap of 25 to 50% is the usual in practice.

Phase-3:
Formulation of Knowledge Base:

The knowledge pertaining to the target problem is generally scattered. In fuzzy building model, the problem is thoroughly studied. Keen observations carried help accumulate and separate the knowledge and information. Thus knowledge base consists of storage of information-called database on one hand and collection of action rule-called rule-base on the other hand.

Data-Base:

Well organization of database contributes great to the success of fuzzy model. The information about the problem is kept stored in the form text, look-up tables, data-files etc. Information accumulation begins well before the actual model implementation. In the process of model development, database goes through the edition phase before its final acceptance.

The basic function of database is to keep the information stored about-

Membership function representing the meaning of problem variables

Ranges of domains [Universe of discourse]

Term sets

Scaling factors

The database supplies required information almost to all the sub-modules of fuzzy system, efficient management of data storage and retrieval of information are vital factors.

Rule-Base:

Writing rules is the final conceptual step in building a fuzzy model. A fuzzy model consists of a series of unconditional and/or conditional fuzzy propositions-called fuzzy production rules that establish the relationship between a value in the underlying domain and fuzzy space.

The basic function of rule-base is to represent policies strategically adopted for the optimization of problem solutions. The construction of rule-base is crucial and most difficult aspect of the fuzzy system design as there are no systematic tools for forming the rule- base. Yet two notable methods of rule formations are discussed in [12].

The first method is based on intuitive knowledge and experience of an personnel closely associated with system process This allows the introduction of 'rule of thumb' experience in the fuzzy system. However, the strength and quality of rule-base depends on how good the process-skills are extracted from the personnel. The process- knowledge so obtained in most of the systems can be inter-twined with a mathematical model of a process. The success of implementation of such approach in industrial applications could be found in [3,12].

The second and more formal approach is based on use of standard rule-base that utilizes the common engineering sense and on-line experience of a process. The rule-base suggested by Mac Vicar Whelan [12] is a good example of this approach. Such rule-base can be used in constructing the more specific and problem oriented knowledge base either by excluding, modifying or adding new action rules.

The fuzzy conditional or fuzzy if-then rule is symbolically expressed as-

IF <fuzzy proposition> THEN < fuzzy proposition >

or in notations as-

IF < X is A > THEN < Y is B >

Where X and Y are scalar-values from respective domains, while A and B are linguistic variables.

The ' if ' premise < X is A > is called rule-antecedent and ' then ' premise < Y is B > is called rule-consequent.

In unconditional rule the proposition is not qualified by ' if ' statement, e.g.

X is A

Such a rule is employed either to restrict the output space or to define the default solution space in the event of none-of the conditional rules is executed.

In engineering applications, the ' if-then ' rules take the following forms-

IF < condition > THEN < decision >

or

IF < condition > THEN < action>

The formation of rules is depicted by the flow chart in fig. 8.

The hints followed while formulating the fuzzy rules are -

Group together all the rules referring to the same variables

Arrange rules for easy readability

Use indentation to highlight the structure of the rule

Follow naming conventions to identify different classes of performance variables and indicate-

Source of variable

Intended use

Dimensionality

Cardinality

Make comments to describe the purpose of rule

Allow white spaces around and between rules

Indicate variable, reference fuzzy sets and hedges if any in mixed-case, while language elements in the lower case

The rules organized in well-structured manner help to improve the model maintainability, quality and expandability.

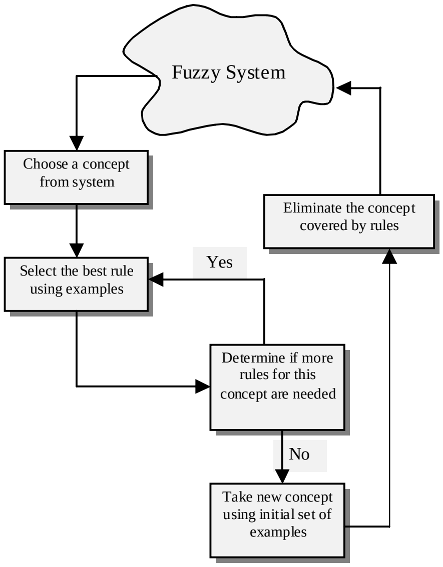

Fuzzy System
Choose a concept from system
Eliminate the concept covered by rules
Yes
Select the best rule using examples
Determine if more rules for this concept are needed
No
Take new concept using initial set of examples

Fig. 8: Formation of Inference-Rules

Phase-4:
Fuzzy Inference:

The functional tie between the degree of truth in related regions is called as method of implication, while the functional tie between fuzzy regions and expected single value from a set is called as method of defuzzification.

Both of these together constitute the backbone of Approximate Reasoning (AR) or Fuzzy Inference. In other words, AR is the way of deriving the conclusions from fuzzy hypothesis that tries to mimic human reasoning as closely as possible.

To select an appropriate fuzzy implication for AR under each particular solution is a difficult problem. Despite guidelines available, general criteria to select implication have yet to emerge. Different implication methods exist with their typicality. One of them is the Mamdani's method of implication. This is most popular of the direct methods proposed by Mamdani. It has simple structure of min- and max-operations and is therefore very popular in applications. This belongs to generalized or fuzzy modus phonon.

It consist of two types-
Compositional based inference
Individual rule based inference

In the first type, the fuzzy relations representing the meaning of each individual rule are aggregated into single relation. This relation describes the meaning of the overall set of rules with which fuzzy-inference or firing is performed through the operation composite between the crisp-input and fuzzy relation. This gives rise to a fuzzy set representing the fuzzy value of overall solution output.

In the second type, first every single rule is fired, i.e. the degree of firing or match between the crisp-problem input and fuzzy sets describing the meaning of rule-antecedent is obtained. This is followed by clipping of fuzzy sets describing the meaning of the rule-consequent to the degree of match equal to that of rule antecedent. Finally, the clipped-values for the output solution variable of each rule are aggregated to obtain the value of overall solution output.

The second type of fuzzy inference being computationally fast and memory efficient is preferred over the other in case of moderate sized rule-base.

Phase-5:
Defuzzification:

Defuzzificaction is the final phase of fuzzy reasoning. The evaluation of rules during the inference process produces the final consequent fuzzy region for each solution variables. This region is then decomposed following the extraction of single (crisp)-value for a solution variable using a technique called 'Defuzzificaction'. Practically it means to drop a plumb line across the domain axis that points to a best single value of solution variable. In general defuzzification algorithms are a compromise or grade-off between the need to find a unique valued solution and the loss of information.

Several defuzzification methods are available and the most often used methods are-
Centre-of-Area /Gravity defuzzification
Centre-of-Sums defuzzification
Centre-of –Largest Area defuzzification
First-of-Maxima defuzzification
Middle-of-Maxima defuzzification
Height-defuzzification
Height-defuzzification:
Height-defuzzification (HD) is a method that uses the individual clipped or scaled control outputs. This method takes the peak value of each fuzzy set and builds the weighted-Sum of these peak values.
Thus neither the support nor the shape of the fuzzy set play a role in the computation of crisp-value of output variable (u*). Defuzzified value of variable is given by equation (1). HD is both simple and quick method.

$$u^* = \frac{\sum_{k=1}^{m} C^{(k)}.f_k}{\sum_{k=1}^{m} f_k} \quad \text{---------} \ (1)$$

Where, $C^{(k)}$ = Peak value of fuzzy set,
f_k = Height of fuzzy set,
m = number of rules fired

It is continuous, non-ambiguous, plausible, and computationally simple and weight-counting method, hence preferred in our model.

Phase-6:
Interpretation of results:

Fuzzy model, like any conventional system, must be validated and verified through it predictive behavior against known cases or against reasonable judgments of experts or at least against common sense.

Knowledge Engineer and system analyst, both must determine and check the consistency of model with the relationships between rules and fuzzy sets, and see whether the validity of solutions obtained are as anticipated or not. If not, model is made to under go the tuning process. The positions of fuzzy sets are reshuffled and rules are scrutinized. This is carried out in iterations, still the satisfactory solutions result. Only after this it is subjected to the stress test.

CONCLUDING REMARK

Designing a fuzzy system as compare to conventional system design approaches has benefits such the system is robust and it can handle vague, imprecise data. It can tackle non-linearity very easily and require no mathematical model as it relies on linguistic description of system.

REFERENCES

[1] Klir G.J.and Yuan B. [1997], Fuzzy Sets and Fuzzy Logic. Prentice-Hall of India, New Delhi, pp.11-338.

[2] Zimmermann H.J., [1996], Fuzzy Set Theory and its Applications. Academic Publishers, Boston.

[3] Dubois D., Prade H. and Yager R.R., (eds.), [1997], Readings in Fuzzy Sets For Intelligent Systems. Morgan Kuafmann Publishers, Inc. SM, California.

[4] Tanaka K. (Translated by Tiimira T.), [1997], An Introduction to Fuzzy Logic for Practical Applications. Springer-Verlag, New York, pp.86-136.

[5] Mamdani E.H. and Gaines B.R. (eds.), [1981], Fuzzy Reasoning and Its Applications. London, Academic Press, pp. 325-334.

[6] Sugeno. M. and Yasukawa T., [1993], A fuzzy-logic based approach to qualitative modeling. IEEE Trans. on Fuzzy Systems, 1(1), pp. 7-31.

[7] Cox E., [1998], The Fuzzy Systems Handbook-A Practitioners Guide to building, using and maintaining Fuzzy Systems. AP-Professional, Boston, pp.45-469.

[8] Shaw Ian S., [1999], Fuzzy Control of Industrial Systems-Theory and Applications. Kluwer Academic Publishers, Boston, pp. 55-178.

[9] Driankov D.,.Hellendoom H and Reinfrank M., [1996], An Introduction to Fuzzy Control. Narosa Publishing House, New Delhi, pp. 1-36,37-144.

[10] Turksen I. B., [1994], Fuzzy systems modeling. Fuzzy Systems and A. I. 2(2). pp. 3-34.

[11] Lee C.C., [1990], Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part I and II. IEEE Trans. on Systems, Man, and Cybernetics, 20(2), pp. 404-418 and 20(2), pp.419-435.

[12] Yager R. R. and. Filev D. P., [1994a], Essentials of Fuzzy Modeling and Control. John Wiley, New York.

[13] Zadeh L.A., [1979], Fuzzy sets and information granularity. In Advances in Fuzzy Set Theory and Applications, M.M. Gupta, R.K. Ragade and R.R. Yager, (eds.), Amsterdam: NH, pp.3-18.

[14] Cho S. and. Ersoy O. K [1992], An algorithm to compute the degree of match in fuzzy systems. Fuzzy Sets and Systems, 49(3), pp. 285-299.

[15] Keller J. M. and Hunt D. J., [1985], Incorporating fuzzy membership functions into the perception algorithm. IEEE Trans. on Pattern Analysis and Machine intelligence. 7(6), pp. 693-699.

[16] Narazaki H. and Ralescu A. L., [1994], Iterative Induction of a Category Membership Functions. Intern. J. of Uncertainty, Fuzziness and Knowledge Based System, Vol.2, No.1, pp.91-100.

[17] Tamano K., [1991]. Optimal fuzzy inference system. SOFT- J. of Japan Society. Fuzzy Theory and Systems 3(2), pp.382-386.

[18] Zadeh L.A., [1975], The Concept of a Linguistic Variable and its Application to Approximate Reasoning. Information Science, Part I, pp.199-249, Part II, pp.301-357. Part III, pp.43-80.

[19] Saade J. J., [1996], A Unifying Approach to Defuzzification and Comparison of the Outputs of Fuzzy Controllers. IEEE Transactions on Fuzzy Systems, Vol.4, No.3, pp.227-237.

[20] Tong R. M., [1980b], The evaluation of fuzzy models derived from experimental data. Fuzzy Sets and Systems, 4(1). pp. 1-12.

Successful weaning from respiratory support system with Neural technology and Evolutionary algorithm

S. Firdosh Parveen, Prof.V C Patil
College: BITM, Bellary
firdoseks@gmail.com,patilvc@rediffmail.com

Abstract

Patients require mechanical ventilator support when the ventilator or gas exchange capabilities of their own respiratory system fail. The process of discontinuation of Mechanical ventilation from the patients is called Weaning. Unnecessary delays in discontinuation can lead to a host of complication and even death. The proposed method is to achieve successful weaning using genetic algorithm. Performance comparison of genetic algorithm with gradient descent is carried, where genetic algorithm gives much accurate results compared with gradient descent. The objective of the research is to provide very high performance in real time operation, which can improve current diagnosis very much.

KEYWORDS: Artificial Neural Networks, Genetic Algorithm, Gradient Descent, Mechanical Ventilation, Successful Weaning.

I INTRODUCTION

Mechanical ventilation is a method to mechanically assist or replace spontaneous breathing. Mechanical ventilation is typically used after an invasive intubation, a procedure wherein an endotracheal or tracheostomy tube is inserted into the airway. It is used in acute settings such as in the ICU for a short period of time during a serious illness. It may be used at home or in a nursing or rehabilitation institution if patients have chronic illnesses that require long-term ventilation assistance [1].

Mechanical ventilation is often a life-saving intervention, but carries many potential complications including pneumothorax, airway injury, alveolar damage, and ventilator-associated pneumonia. Mechanical ventilation, although the mainstay of treatment for respiratory distress syndrome, can result in physical trauma to lung tissue. Respiratory failure requiring mechanical ventilation has a large impact on hospital economics and resources.

The timing and method of discontinuation from mechanical ventilation remains an important clinical problem. Mechanical ventilation can result in life-threatening complications and therefore should be discontinued as soon as possible. However, premature attempts at weaning from respiratory support can result in failure and reinstitution of mechanical ventilation, which carries an enhanced risk of morbidity and mortality. Therefore, it is no surprise that many different strategies for successful weaning have been described in the medical literature.

All patients should be evaluated daily to determine if they are candidates for discontinuation of mechanical ventilation. To be considered a candidate, a patient should meet four criteria: (1) evidence of reversal or stability of the cause of acute respiratory failure; (2) adequate oxygenation as indicated by $Pao_2/Fio_2 > 150$ to 200, PEEP(Positive End Expiratory Pressure) in the range of ≤ 5 to 8 cm H_2O, $Fio_2 \leq 0.4$ to 0.5, and $pH > 7.25$; (3) hemodynamic stability, as defined by the absence of active myocardial ischemia and the absence of clinically significant hypotension and (4) ability to make an inspiratory effort[2].

A. Artificial Neural Networks

1. Architecture: The design and implementation of intelligent system with human capabilities is the starting point to design Artificial Neural Networks (ANN). Artificial neural networks are computational systems whose architecture and operation are inspired from the knowledge about biological neural cells (neurons) in the brain [3].

ANNs is a network of many simple processors called units, linked to certain neighbors with varying coefficients of connectivity called weights that represent the strength of these connections. The basic unit of ANNs called an artificial neuron, simulates the basic functions of natural neurons. It receives inputs process them by simple connections and threshold operations and outputs a result.

ANN's can be divided into feedforward and recurrent classes according to their connectivity. An ANN is feed forward if their exists a method which numbers all the nodes in the network such that there is no connection from a node with a large number to a node with a smaller number. All the connections are from the node with a small number to the node with a larger number. An ANN is recurrent if such a numbering method does not exists.

Figure.1 shows the schematic representation of a multilayer perceptron with eight input neurons, two hidden layers with eight hidden neurons and one output layer with single neuron. Each of the input neuron connects to each of the hidden neurons, and each of the hidden neurons connects to the output neurons

The architecture of an ANN is determined by its topological structure i.e., the overall connectivity and transfer function of each node in the network.

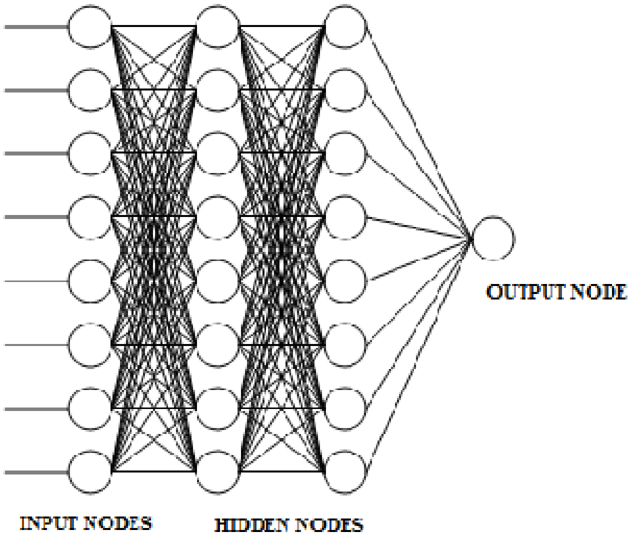


Fig.1 Schematic Representation of a Multi Layer Perceptron

2. Learning in ANN: Learning in ANN's is accomplished using examples. This is also called "training" in ANN's because the learning is achieved by adjusting the connection weight in ANN's iteratively so that trained (or learned) ANN's can perform certain tasks. Learning in ANN's can roughly be divided into supervised, unsupervised and reinforcement learning. Supervised learning is based on direct comparison between the actual output of an ANN and desired correct output also known as the target output. It is often formulated as the minimization of an error function such as the total mean square error between the actual output and the desired output summed over all available data. A gradient descent- based optimization algorithm such as back propagation (BP)[4] can then be used to adjust connection weights in the ANN iteratively in order to minimize the error. Reinforcement learning is a special case of a supervised learning where the exact desired the output is unknown. It is based only on the information of whether or not the actual output is correct. Unsupervised learning is solely base don the correlations among input data. No information on the "correct output" is available for learning.

The essence of a learning algorithm is the learning rule, i.e., a weight updating rule which determines how connection weights are changed[5].

B. Evolutionary algorithm

Evolutionary algorithms are stochastic optimization algorithms based on the mechanism of natural selection and natural genetics [6].They perform parallel search in complex search spaces. Evolutionary algorithms include genetic algorithms, evolution strategies and evolutionary programming. We deal with genetic algorithms in this paper. Genetic algorithms (GA's) were originally proposed by Holland [7].GA's have been applied to many function optimization problems and are shown to be good in finding optimal and near optimal solutions. Their robustness of search in large search spaces and their domain independent nature motivated their applications in various fields like pattern recognition, machine learning, VLSI design, etc.

C. Problem statement

Improper weaning may lead to very serious consequences, such as prolonged ventilation time, pneumonia and even death. Therefore, every effort must be extended to alleviate the problems mentioned. This paper will demonstrate which patients can be successfully weaned. The objective of the project is to

1. Develop a Dynamic Expert system for weaning the mechanical ventilation with very high accuracy and precision.
2. Evolutionary computation (Genetic algorithm) will be involved with neural technology for further performance enhancement.
3. To develop a very robust design using distributed architecture.
4. Performance comparison of genetic algorithm with Gradient descent algorithm.
5. Graphical and numerical methods for performance and comparative analysis.
6. To create low cost, high efficient health care environment in ICU.

This paper is organized as follows: section 2 discuses the related work, section 3 describes the methodology used, section 4 describes the experimental results, and section 5 concludes the paper.

II Related work

Two large (and related) evidence-based projects were conducted to review the results of published studies and evaluate the different strategies for successful weaning from mechanical ventilation. The first of these was commissioned by the Agency for Healthcare Policy and Research, who asked the McMaster University Evidence-Based Practice center to evaluate the issues surrounding ventilation, weaning and discontinuation. This group was directed to address five specific questions: (1) When should weaning be initiated? (2) What criteria should be used to determine when to begin the weaning process? (3) What are the most effective methods of weaning? (4) What are the optimal roles for nonphysician health-care providers in the weaning process? and (5) What is the value of using clinical practice algorithms or protocols in the weaning process? The McMaster project reviewed > 5,000 reports and identified 154 publications, which they used to create their comprehensive review. The second group was a task force put together by the American College of Chest Physicians, the Society for Critical Care Medicine, and the American Association for Respiratory Care. This group was charged to create clinical practice guidelines based on the McMaster report and their own literature review and expert consensus opinion [2][8].

III METHODOLOGY

The performance of the proposed method is demonstrated by employing the genetic algorithm. The performance criteria used in this research is to achieve the successful weaning of the patients. This paper compares the results of genetic algorithm and gradient descent.

The data is collected from the Royal infirmary of Edinburgh intensive care unit, Scotland. The data include the patient's respiratory parameters like Tidal Volume (VT), Respiratory Pressure (RR), Minimum Ventilation (VE) and Negative Inspiratory Pressure (NIF) are available at every instant of time and based on these values the patient shall be examined for weaning decision.

A. Gradient descent

Gradient descent is a function optimization method which uses the derivative of the function and the idea of steepest descent. The derivative of a function is simply the slope. So if we know the slope of a function, then it stands to reason that all we have to do is somehow move the function in the negative direction of the slope and that will reduce the value of the function. Gradient descent is an iterative method, so the idea is as follows:

- Compute the derivative of the function with respect to its independent variables. We can denote this derivative as dF(x), where F(x)) is the function to be minimized, and x is the vector of independent variables.

- Change the value of x as follows:
  $x_{n+1} = x_n - h\, dF(x_n)$,
  where the subscript n refers to the iteration number, and h is a step size which must be chosen so that we don't take too big or too small of a step. Too big of a step will overshoot the function minimum, and too small of a step will result in a long convergence time.
- Repeat the above two steps until we converge to a minimum of the function F(x).

Gradient descent is an attractive optimization method in that it is conceptually straightforward and often converges quickly. Its drawbacks include the fact that the derivative of the function must be available and it converges to a local minimum rather than global minimum. So in this proposed work training is given using this method, but when tested the expected output is different from the output given by the machine

B .Genetic Algorithm

GA has blossomed rapidly due to the easy availability of low cost but fast speed small computers. The complex and conflicting problems that required simultaneous solutions, which in past were considered deadlocked problems, can now be obtained with GA. However, the GA is not considered a mathematically guided algorithm. The optima obtained are evolved from generation to generation without stringent mathematical formulation such as the traditional gradient–type of optimizing procedure. Infact; GA is much different in that context. It is merely a stochastic, discrete event and a non linear process. The obtained optima are an end product containing the best elements of previous generations where the attributes of a stronger individual tend to be carried forward into the following generation. The rule of the game is “survival of the fittest will win” [9].

A simple genetic algorithm can be summed up in seven steps as follows [10]:

1. Start with a randomly generated population of n chromosomes.
2. Calculate fitness of each chromosome.
3. Select a pair of parent chromosomes from the initial population.
4. With a probability *Pcross* (the ‘crossover probability’ of the ‘crossover rate’), perform crossover to produce two offspring.
5. Mutate the two offspring with a probability *Pmut* (the mutation probability).
6. Replace the offspring in the population.
7. Check for termination or go to step 2.

Each iteration of the above steps is called a generation. The termination condition is usually a fixed number of generations typically anywhere from 50 to 500 or more. Under certain other circumstances, a check for global minimum is done after each generation and the algorithm is terminated as and when it is reached [11]. The chief aspect of this method is the representation of the parameter as strings of binary digits of 0 and 1. This composition allows simple crossover and mutation functions that can operate on the chromosomes.

C. Comparison between Genetic algorithm based training and Gradient based training

The genetic algorithm based training approach is attractive because it can handle the global search problem in a vast, complex, multimodal and non differentiable surface. It does not depend on the gradient information of the error function and thus is particularly appealing when this information is unavailable or very costly to obtain or estimate.

Genetic algorithm is less sensitive to initial condition of training. They always search for a globally optimal solution, while a gradient descent algorithm can only find a local optimum in a neighbourhood of the initial solution.

Genetic algorithm based training algorithm is significantly faster than methods that use the GDR. For the three tests in his paper [12], the Genetic algorithm based training algorithm “took a total of about 3 hours and 20 minutes, and the GDR took a total of about 23 hours and 40 minutes”.

IV EXPERIMENTAL RESULTS

A computer simulation has been developed to achieve successful weaning using genetic algorithm. The simulations have been carried out using MATLAB. Various networks were developed and tested with random initial weights. The data is collected from the intensive care unit of royal infirmary of Edinburgh, Scotland. When NN are trained using genetic algorithm and later tested gave the correct results for which patient to be weaned or not where as the gradient descent method does not give the correct result. The results of genetic algorithm and gradient descent are shown in the graph below.

Graph for Gradient descent

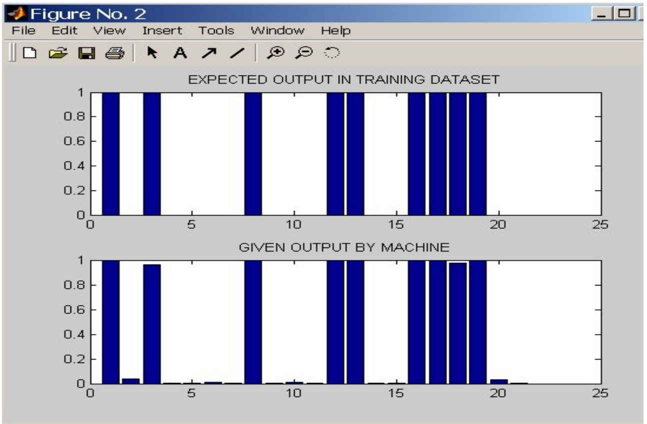


Fig 2 Snapshot showing the expected output in training dataset and output given by the machine

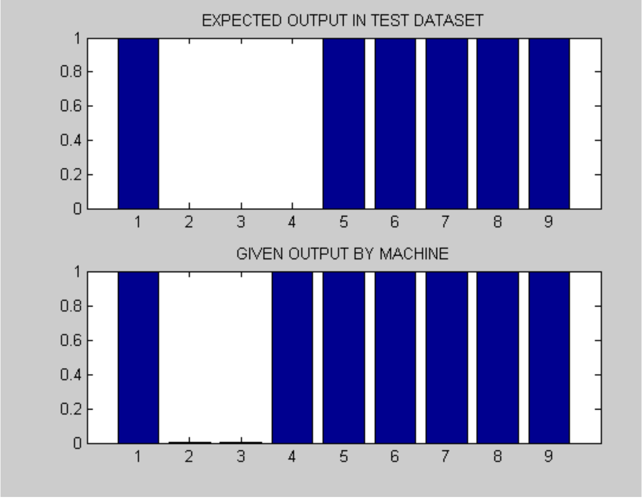


Fig 3 Snapshot showing the expected output in test dataset and output given by the machine

Graphs for Genetic algorithm

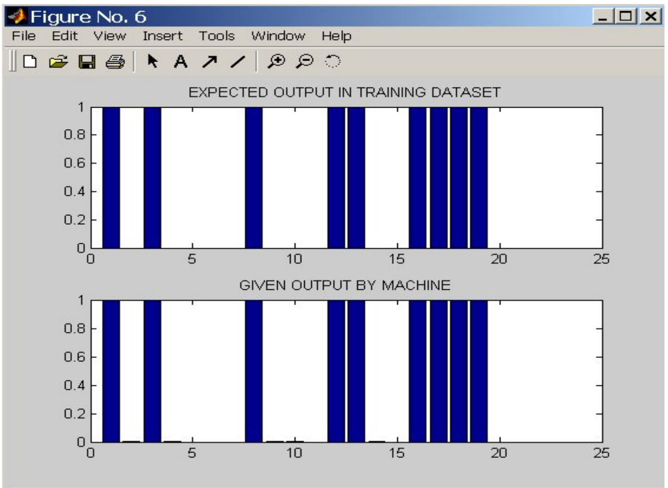


Fig 4 Snapshot showing the expected output in training dataset and output given by the machine

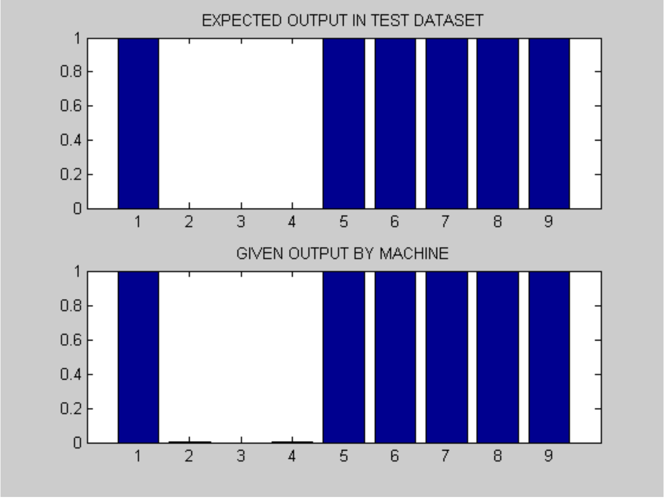


Fig 5 Snapshot showing the expected output in test dataset and output given by the machine

V CONCLUSION

Mechanical ventilation can have life treating complications it should be discontinued at the earliest possible time. the process of discontinuing mechanical ventilation termed" weaning" is one of the most challenging problems in intensive care and it account for a considerable proportion of the work load of staff in an ICU. The discontinuation of mechanical ventilation needs to be carefully timed premature discontinuation place severe stress on the respiratory and cardiovascular system which can impede the patients recovery. Unnecessary delay in discontinuation can lead to a host of complication.

Clinicians are generally inaccurate in weaning decision because there is no accurate predefined rule. Accuracy is only on average around 60% with trail and error method.

This paper demonstrates the successful weaning of the patients from mechanical ventilation with Neural Networks and evolutionary computation i.e., genetic algorithm.

VI REFERENCES

1. Collice, Gene L "Historical prespective on the development of mechanical ventilation". in Martin J Tobin. Principles and practice of Mechanical ventilation 2nd edition. Newyork: Mc graw-hill.
2. Neil R. Mac intyre MD Current Issues in Mechanical Ventilation For Respiratory Failure, 01 Nov 2005.
3. Madiha J. Jafri, Vince D. Calhoun (2006), Functional Classification of Schizophrenia Using Feed Forward Neural Networks, Proceedings of the 28th IEEE EMBS Annual International Conference, pp.6631-6634.
4. G.E.Hinton, "Connectionist learning procedures," Artificial Intel., vol. 40, no. 1–3, pp. 185–234, Sept. 1989.
5. J. Hertz, A. Krogh, and R. Palmer, Introduction to the Theory of Neural Computation. Reading, MA: Addison-Wesley, 1991.
6. D. B. Fogel, "An introduction to simulated evolutionary optimization", *IEEE Trans. Neural Networks,* vol. 5, no. 1, pp. 3–14, 1994.
7. J. H. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: Univ. of Michigan Press, 1975.
8. Chairman – Neil R. Mac Intyre M.D, F.C.C.P A collective task force facilitated by the American Collage of Chest, The American Association for respiratory care; and The American collage of critical care medicine, December, 2001.
9. D. E. Goldberg, *"Genetic Algorithms in Search, Optimization and Machine Learning"*, Reading, MA: Addison-Wesley, 1989.
10. Marco Russo, "Fu Ge Ne Sys – A Fuzzy Genetic Neural System for Fuzzy Modeling", *IEEE Transactions on Fuzzy Systems*, vol6, no,3,August 1998, pp 373 – 387.
11. Melanie Mitchell, *"An Introduction to Genetic Algorithms"*, A Bradford Book MIT Press, 1997.
12. D. L. Prados, "Training multilayered neural networks by replac-ing the least fit hidden neurons," in Proc. IEEE SOUTHEAST-CON'92, vol. 2, pp. 634–637.
13. Gilat, Amos, Matlab: An introduction with applications, 2nd edition,2004. John Wiley and sons.

COMPOSITION OF FUZZY RELATION IN VEHICAL SPEED CONTEROL

K.D. Attar, S.V. Khot, G.R. Attar, R.B. Gaikawad

Department of Electronics, Shivaji University, Kolhapur - 416004, India.
*e-mail:Attar.Khwaja@gmail.com gsmd88@gmail.com

ABSTRACT:- The measurement of speed in vehicle this Project are useful in fuzzy relation.This is the mechanism of Mamdani's direct method. In order to draw conclusions from a set of rules (rule base) one needs a mechanism that can produce an output from a collection of rules. This is done using the compositional rule of inference. It should adopt a method that better suits the objective. IF number of rule large, it is better to apply a fuzzy relations approach to computer program. Whatever the number of rules, IF-THEN rule would fit in to the compact fuzzy relation form. Such fit in operation is done offline and once we obtain the compact form, the speed of reasoning is higher than mechanism of reasoning.

Introduction:-

The Fuzzy Logic Decision System design mainly includes the process of Fuzzification, Inference composition Implication and Defuzzification. In Inference process IF-THEN fuzzy rules are employed on the fuzzified input. In process of implication these rules are applied on fuzzified input to yield a fuzzified output. The Fuzzy decision-making based on software approach is explored to operate non-fuzzy hardware. The intervention of software retards the decision-making though it renders flexibility in the general sense. An alternate approach is to incorporate the Fuzzy Logic in the hardware itself. This can be achieved by using Fuzzy Gates. Using hardwired implementation of Fuzzy Logic Gates it is possible to optimize real-time Fuzzy-Implications. The aim of the present paper is to demonstrate the realization of fuzzy implications using op-amp. It is a step close towards building fully hardwired Fuzzy Implication useful in variety of application form diverse fields.

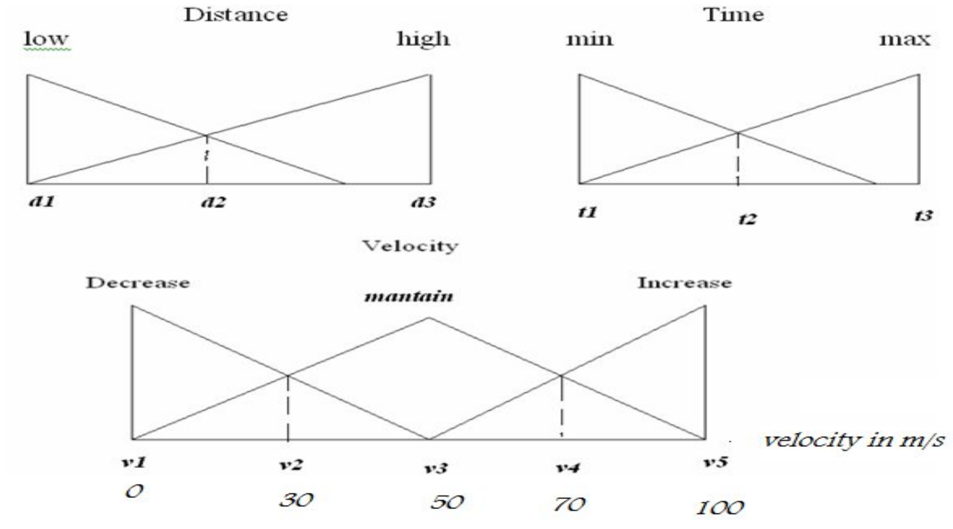


Fig.1. Fuzzy set diagram

FUZZY REASONING USING IMPLICATIONS

There are various implication methods employed in the process of fuzzy reasoning. Being more popular Mamdani's and Zadeh's implication methods are considered for op-amp implication. The various standard implication methods are as shown in Table-1.

Table-1: Various Implications methods & its Formulae

No.	Implication method	Formulae
1	Mamdani's method	$a \rightarrow b = a \cap b$
2	Algebraic product	$a \rightarrow b = a \cdot b$
3	Bounded product	$a \rightarrow b = 0 \cup (a + b - 1)$
4	Zadeh's method	$a \rightarrow b = 1 \cap (1 - a + b)$
5	Boolean logic	$a \rightarrow b = (1 - a) \cup b$
6	Goedel logic	$a \rightarrow b = \begin{cases} 1 & a \le b \\ b & a > b \end{cases}$

Description:- to control the speed of object or vehicle where as fuzzy reasoning is based on "Generalized Modus Ponens" also called as "Fuzzy Modus Ponens". Inference rules of fuzzy reasoning are expressed in IF-THEN format. In this rule the term following the IF statement iscalled the *premise*, and the term following THEN is called *consequence*.

Velocity = m/s
Distance:
Low (A1) =1/d1 + 0.4/d2 + 0/d3;
High (A2) = 0/d1 +0.4/d2 + 1/d3;

Time:
Min (B1) =1/t1 + 0.5/t2 + 0/t3;
Max (B2) = 0/t1 + 0.5/t2 + 1/t3;

Velocity (speed);
M = 0/v1 +0.6/v2 +1/v3 + 0.5/v4 + 0/v5;
I = 0/v3 + 0.5/v4 + 1/v5;
D = 1/v1 + 0.6/v2 + 0/v3;

Composition Rules: -
Fuzzy relation in composition rule are as follows;

R1 = IF *Distance* **A1** and *time* is **B1** THEN **V** is *maintain.*
R2 = IF *Distance* **A2** and *time* is **B2** THEN **V** is *maintain..*
R3 = IF *Distance* **A1** and *time* is **B2** THEN **V** is *Increase.*
R4 = IF *Distance* **A2** and *time* is **B1** THEN **V** is *decrease*

Composition in comparison with matrix multiplication / implication:-.

μd1 μd2 μd3
1.0 0.4 0.0

R1=

	V1=0.0	V2=0.6	V3=1.0	V4=0.5	V5=0.0
μt1=1.0	0.0 0.0 0.0	0.6 0.4 0.0	1.0 0.4 0.0	0.5 0.4 0.0	0.0 0.0 0.0
μt2=0.5	0.0 0.0 0.0	0.5 0.4 0.0	0.5 0.4 0.0	0.5 0.4 0.0	0.0 0.0 0.0
μt3=0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0

μd1 μd2 μd3
0.0 0.4 1.0

R2=

	V1=0.0	V2=0.6	V3=1.0	V4=0.5	V5=0.0
μt1=0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0
μt2=0.5	0.0 0.0 0.0	0.0 0.4 0.5	0.0 0.4 0.5	0.0 0.4 0.5	0.0 0.0 0.0
μt3=1.0	0.0 0.0 0.0	0.0 0.4 0.6	0.0 0.4 1.0	0.0 0.4 0.5	0.0 0.0 0.0

μd1 μd2 μd3
1.0 0.4 0.0

R3=

	V3=0.0	V4=0.5	V5=1.0
μt1= 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0
μt2= 0.5	0.0 0.0 0.0	0.5 0.4 0.0	0.5 0.4 0.0
μt3= 1.0	0.0 0.0 0.0	0.0 0.4 0.0	1.0 0.4 0.0

μd1 μd2 μd3
0.0 0.4 1.0

R4=

	V1=1.0	V2=0.6	V3=0.0
μt1= 1.0	0.0 0.4 1.0	0.0 0.4 0.6	0.0 0.0 0.0
μt2= 0.5	0.0 0.4 0.5	0.0 0.4 0.5	0.0 0.0 0.0
μt3= 0.0	0.0 0.0 0.0	0.0 0.0 0.0	0.0 0.0 0.0

Applying Mamdani's method in compilation of fuzzy relation,

R = R1 U R2 U R3 U R4

R=

	V1=0.0	V2=0.6	V3=1.0	V4=0.5	V5=0.0
	0.0 0.4 1.0	0.6 0.4 0.6	1.0 0.4 0.0	0.5 0.4 0.0	0.0 0.0 0.0
	0.0 0.4 0.5	0.5 0.4 0.5	0.5 0.4 0.5	0.5 0.4 0.5	0.5 0.4 0.0
	0.0 0.0 0.0	0.0 0.4 0.6	0.0 0.4 1.0	0.0 0.4 0.5	1.0 0.4 0.0

Reasoning Fuzzy relation:

A1'= 0.9/d1 + 0.3/d2 + 0.0/d3

B1' = 0.7/t1 + 0.4/t2 + 0.0/t3

V1 =?

Compositional rule of inference:

B1'o (A1'o R) = V'

$$A1'o\,R = (0.9\ 0.3\ 0.0) \begin{vmatrix} 0.0 & 0.4 & 1.0 \\ 0.0 & 0.4 & 0.5 \\ 0.0 & 0.0 & 0.0 \end{vmatrix}$$

={ max[min (0.9 0.0) min(0.3 0.0) min(0.0 0.0)]
max[min (0.9 0.4) min(0.3 0.4) min(0.0 0.0)]
max[min (0.9 1.0) min(0.3 0.5) min(0.0 0.0)] }

= { max[0.0 0.0 0.0] max[0.4 0.3 0.0] max[0.9 0.3 0.0] }
A1'o R = (0.0 0.4 0.9)

$$A1'o\,R = (0.9\ 0.3\ 0.0) \begin{vmatrix} 0.6 & 0.4 & 0.6 \\ 0.5 & 0.4 & 0.5 \\ 0.0 & 0.4 & 0.6 \end{vmatrix}$$

= { max(0.6 0.3 0.0) max(0.4 0.3 0.0) max(0.6 0.3 0.0) }
A1'o R = (0.6 0.4 0.6)

$$A1'o\,R = (0.9\ 0.3\ 0.0) \begin{vmatrix} 1.0 & 0.4 & 0.0 \\ 0.5 & 0.4 & 0.5 \\ 0.0 & 0.4 & 1.0 \end{vmatrix}$$

= { max(0.9 0.3 0.0) max(0.4 0.3 0.0) max(0.0 0.3 0.0) }
A1'o R = (0.9 0.4 0.3)

$$A1'o\,R = (0.9\ 0.3\ 0.0) \begin{vmatrix} 0.5 & 0.4 & 0.0 \\ 0.5 & 0.4 & 0.5 \\ 0.0 & 0.4 & 0.5 \end{vmatrix}$$

= { max(0.5 0.3 0.0) max(0.4 0.3 0.0) max(0.0 0.3 0.0) }
A1'o R = (0.5 0.4 0.3)

$$A1'o\,R = (0.9\ 0.3\ 0.0) \begin{vmatrix} 0.0 & 0.0 & 0.0 \\ 0.5 & 0.4 & 0.0 \\ 1.0 & 0.4 & 0.0 \end{vmatrix}$$

= { max(0.0 0.3 0.0) max(0.0 0.3 0.0) max(0.0 0.0 0.0) }
A1'o R = (0.3 0.3 0.0)

A1'o R = 0.0 0.4 0.9 | 0.6 0.4 0.6 | 0.9 0.4 0.3 | 0.5 0.4 0.3 | 0.3 0.3 0.0

$$(A1'o\,R) = \begin{vmatrix} 0.0 & 0.6 & 0.9 & 0.5 & 0.3 \\ 0.4 & 0.4 & o.4 & 0.4 & 0.3 \\ 0.9 & 0.6 & 0.3 & 0.3 & 0.0 \end{vmatrix}$$

$$B1'o\,(A1'o\,R) = (0.7\ 0.4\ 0.0) \begin{vmatrix} 0.0 & 0.6 & 0.9 & 0.5 & 0.3 \\ 0.4 & 0.4 & o.4 & 0.4 & 0.3 \\ 0.9 & 0.6 & 0.3 & 0.3 & 0.0 \end{vmatrix}$$

= { max(0.0 0.4 0.0) max(0.6 0.4 0.0) max(0.7 0.4 0.0)
max(0.5 0.4 0.0) max(0.3 0.3 0.0) }

Defuzzification Fuzzy set:

Defining the value (speed): v1=0 v2=30 v3=50 v4=70 v5=100 m/s

$$v' = \frac{0.4v1' + 0.6v2' + 0.7v3 + 0.5v4 + 0.3v4}{0.4 + 0.6 + 0.7 + 0.5 + 0.3}$$

$$= \frac{0.4x0 + 0.6x30 + 0.7x50 + 0.5x70 + 0.3x100}{0.4 + 0.6 + 0.7 + 0.5 + 0.3}$$

$$V' = 47.2 m/s$$

Result:

The mamdanie's method rule 1can be proved

Conclusion:

There are various methods like Fuzzification, Inference composition Implication and Defuzzification. Here we are using mamdanies method. As it is easy to handle compare to other methods used for Defuzzification. to control the speed of vehicle. Thus we can implement fuzzy logic in real system.

REFERENCES

1. Tanaka K. (Translated by Tiimira T.), [1997], "An Introduction to Fuzzy Logic for Practical Applications", Springer-Verlag, New York, pp.86-136.

2. Ibrahim A.M., [1996], "Introduction to Applied Fuzzy Electronics", Prentice-Hall, New Jersey, pp.1-96.

3. Mamdani E.H. and Gaines B.R. (eds.), [1981], "Fuzzy Reasoning and Its Applications", London, Academic Press, pp. 325-334.

4. Zadeh L.A., [1984], "Making Computers think like people", IEEE Spectrum, pp.26-32.

5. Klir G.J.and Yuan B. [1997], "Fuzzy Sets and Fuzzy Logic", Prentice-Hall of India, New Delhi, pp.11-338.

6. Kevin Self Correspondent, [1990], "Designing With Fuzzy Logic" IEEE Spectrum, pp.42-44, 105.

Fuzzy Logic Based Brightness Controller of the TV

Pooja Uchagaonkar, Priyanka Bekanalkar, Radhika Bhoite, Sanmati Patil

Department of Electronics, Shivaji University, Kolhapur. INDIA-416 004
pooja.uchagaonkar@gmail.com radhika.bhoite@gmail.com patilsanmati15@gmail.com

Abstract : One of the major cause of eye strain and other problems caused while watching television is the relative illumination between the screen and its surrounding. This can be overcome by adjusting the brightness of the screen with respect to the surrounding light. A controller based on fuzzy logic is proposed in this paper. The fuzzy controller takes in the intensity of light surrounding the screen and the present brightness of the screen as input. The output of the fuzzy controller is the grid voltage corresponding to the required brightness. This voltage is given to CRT and brightness is controller dynamically. For the given test system data, MAMDANI de-fuzzifier methods have been implemented. In order to validate the effectiveness of the proposed approach, a fuzzy controller has been designed by obtaining a test data from a real time system.

Keywords: *Outside light, Grid Voltage*

Introduction:

People look into different types of display screens, either for work or entertainment. On an average a person sits in front of them for a minimum of two hours a day. The screen used in these equipments produce images and videos by providing sufficient voltage to the cathode ray tube which is of very high intensity. People surfing through the channels get affected due to the huge contrast between the outside atmospheric light and the television brightness. When the contrast difference in contrast increases additional stress is applied causing strains, headache and many visual disorders. That is why every television or a computer monitor is provided with a brightness adjustment system. This has to be optimally adjusted so that the relative illumination is less. Manual adjustment has been practiced for quite a long period, but this has not been very effective since it requires continuous adjustment of brightness depending on the intensity of the screen and outside light. These factors vary continuously with time and they also depend upon the channel. Automatic adjustment of contrast and brightness of the screen suited to the external light can be done to reduce the viewers stress. So an automatic controller is proposed in this paper which controls the brightness depending upon the surrounding light. Fuzzy controller, one of the effective and intelligent options is employed in this paper to adjust the brightness in a professional manner enabling the viewer to reap the best results possible.

Fuzzy controller consists of a classifier, fuzzifier, rule base, interface engine. In the fuzzy rule base, various rules are formed according to the problem's requirements. The numerical input values to the fuzzier are converted into fuzzy values. The fuzzy values along with the rule base are fed into the inference engine which produces control values. As the control values are not in usable form, they have to be converted to numerical output values using the defuzzifier. The block diagram of the fuzzy controller used in this paper is shown in Fig. 1.

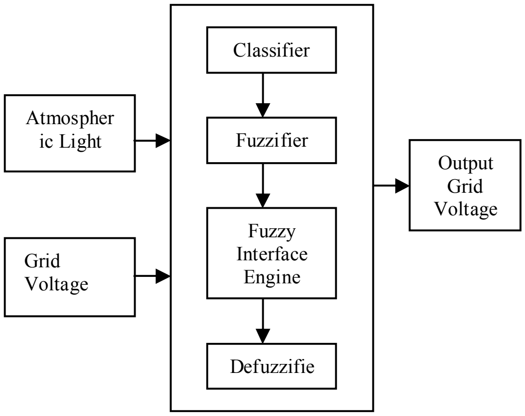


Fig. 1 Block Diagram

The first input parameter considered in this controller is the atmospheric light. This measures the intensity of light surrounding the screen. The atmospheric light is divided into three categories as dark, medium, bright. The next category is dark and this is the range of intensities during early morning or late evening with no illumination source other than the sun. The medium function corresponds to late morning or early evening when the sun's power can be felt mildly. The bright function represents that hour of the day when the sun is at its peak. The membership functions of first input parameter are shown in Fig. 2.

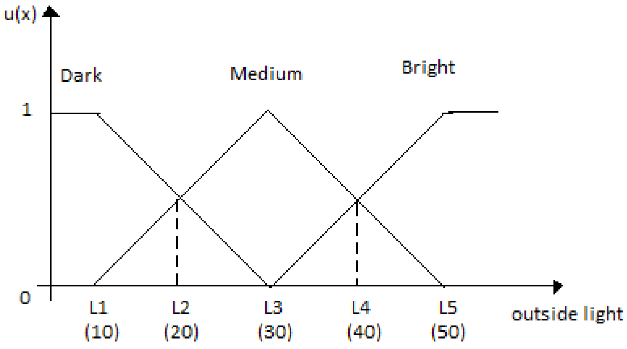


Fig 2(a): Membership function of outside light

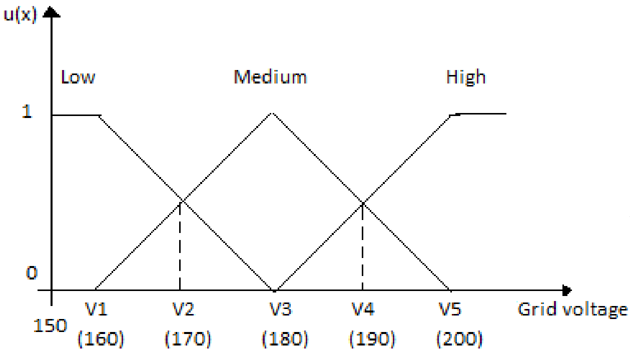


Fig 2(b): Membership function of grid voltage

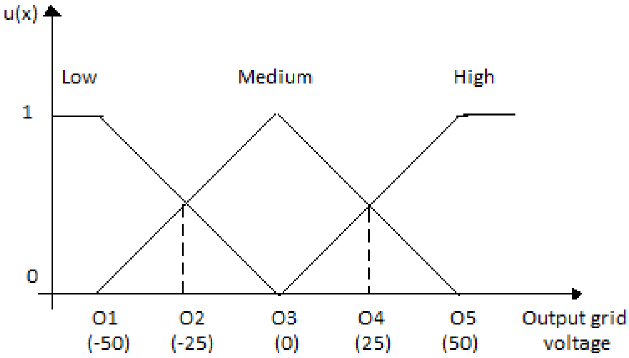


Fig 2(c): Membership function of output grid voltage

Basically the rules are if-then statements that are intuitive and easy to understand, since they are nothing but common English statements. Rules used in this paper are derived from common sense, data taken from typical home use, and experimentation in a controlled environment. The sets of rules used here to derive the output are:

1. IF outside light is Low and grid voltage is Low THEN output grid voltage is maintain;
2. IF outside light is Low and grid voltage is Medium THEN output grid voltage is Decrease;
3. IF outside light is Low and grid voltage is High THEN output grid voltage is Decrease;
4. IF outside light is Medium and grid voltage is Low THEN output grid voltage is Increase;
5. IF outside light is Medium and grid voltage is Medium THEN output grid voltage is Maintain;
6. IF outside light is Medium and grid voltage is High THEN output grid voltage is Decrease ;
7. IF outside light is High and grid voltage is Low THEN output grid voltage is Increase;
8. IF outside light is High and grid voltage is Medium THEN output grid voltage is Maintain;
9. IF outside light is High and grid voltage is High THEN output grid voltage is Maintain;

Outside light:

Low (A_1) = $1/L_1+0.5/L_2+0/L_3$
Maintain(A_2) = $0/L_1+0.5/L_2+1/L_3+0.5/L_4+0/L_5$
High (A_3) = $0/L_3+0.5/L_4+1/L_5$

Grid voltage:

Low (B_1) = $1/V_1+0.5/V_2+0/V_3$
Maintain (B_2) = $0/V_1+0.5/V_2+1/V_3+0.5/V_4+0/t\ V_5$
Greasy (B_3) = $0/V_3+0.5/V_4+1/V_5$

Output grid voltage:

Decrease (G_1) = $1/O_1+0.5/O_2+0/O_3$
Maintain (G_2) = $0/O_1+0.5/O_2+1/O_3+0.5/O_4+0/O_5$
Increase (G_3) = $0/O_3+0.5/O_4+1/O_5$

Rules:-

R_1: IF L is A_1 & V is B_1 THEN O is G_2
R_2: IF L is A_1 & V is B_2 THEN O is G_1
R_3: IF L is A_1 & V is B_3 THEN O is G_1
R_4: IF L is A_2 & V is B_1 THEN O is G_3
R_5: IF L is A_2 & V is B_2 THEN O is G_2
R_6: IF L is A_2 & V is B_3 THEN O is G_1
R_7: IF L is A_3 & V is B_1 THEN O is G_3
R_8: IF L is A_3 & V is B_2 THEN O is G_3
R_9: IF L is A_3 & V is B_3 THEN O is G_2

By using Mamdani's implication method,

R_1: A_1 & B_1→ G_2

$$R_1= \begin{array}{cc|ccccc} & & 1 & 0.5 & 0 & 0 & 0 \\ & & V_1 & V_2 & V_3 & V_4 & V_5 \\ \hline 1 & L_1 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & L_2 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & L_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_5 & 0 & 0 & 0 & 0 & 0 \end{array} \quad \begin{array}{|ccccc|} \hline 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|} \hline 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

O_2=0.5 **O_3=1** **O_4=0.5**

R_2: A_1 & B_2→ G_1

$$R_2= \begin{array}{cc|ccccc} & & 0 & 0.5 & 1 & 0.5 & 0 \\ & & V_1 & V_2 & V_3 & V_4 & V_5 \\ \hline 1 & L_1 & 0 & 0.5 & 1 & 0.5 & 0 \\ 0.5 & L_2 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & L_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_5 & 0 & 0 & 0 & 0 & 0 \end{array} \quad \begin{array}{|ccccc|} \hline 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

O_1=1 **O_2=0.5**

R_3: A_1 & B_3→ G_1

$$R_3 = \begin{array}{cc|ccccc} & & 0 & 0 & 0 & 0.5 & 1 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 1 & L_1 & 0 & 0 & 0 & 0.5 & 1 \\ 0.5 & L_2 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & L_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_5 & 0 & 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

O_1=1

O_2=0.5

R_4: A_2 & B_1→ G_3

$$R_4 = \begin{array}{cc|ccccc} & & 1 & 0.5 & 0 & 0 & 0 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 0 & L_1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & L_2 & 0.5 & 0.5 & 0 & 0 & 0 \\ 1 & L_3 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & L_4 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & L_5 & 0 & 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

O_4=0.5

O_5=1

R_5: A_2 & B_2→ G_2

$$R_5 = \begin{array}{cc|ccccc} & & 0 & 0.5 & 1 & 0.5 & 0 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 0 & L_1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & L_2 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 1 & L_3 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0.5 & L_4 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & L_5 & 0 & 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

O_2=0.5

O_3=1

O_4=0.5

R_6: A_2 & B_3→ G_1

$$R_6 = \begin{array}{cc|ccccc} & & 0 & 0 & 0 & 0.5 & 1 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 0 & L_1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & L_2 & 0 & 0 & 0 & 0.5 & 0.5 \\ 1 & L_3 & 0 & 0 & 0 & 0.5 & 1 \\ 0.5 & L_4 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & L_5 & 0 & 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

O_1=1

O_2=0.5

R_7: A_3 & B_1→ G_3

$$R_7 = \begin{array}{cc|ccccc} & & 1 & 0.5 & 0 & 0 & 0 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 0 & L_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_3 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & L_4 & 0.5 & 0.5 & 0 & 0 & 0 \\ 1 & L_5 & 0.5 & 0.5 & 0 & 0 & 0 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 1 & 0.5 & 0 & 0 & 0 \\ \hline \end{array}$$

O_4=0.5

O_5=1

R_8: A_3 & B_2→ G_3

$$R_8 = \begin{array}{cc|ccccc} & & 0 & 0.5 & 1 & 0.5 & 0 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 0 & L_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_3 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & L_4 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 1 & L_5 & 0 & 0.5 & 0.5 & 0.5 & 0 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0 \\ \hline \end{array}$$

O_4=0.5

O_5=1

R_9: A_3 & B_3→ G_2

$$R_9 = \begin{array}{cc|ccccc} & & 0 & 0 & 0 & 0.5 & 1 \\ & & v_1 & v_2 & v_3 & v_4 & v_5 \\ \hline 0 & L_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_3 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & L_4 & 0 & 0 & 0 & 0.5 & 0.5 \\ 1 & L_5 & 0 & 0 & 0 & 0.5 & 0.5 \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \\ \hline \end{array} \qquad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ \hline \end{array}$$

O_2=0.5

O_3=1

O_4=0.5

RULE COMPILATION:-

R = R1 V R2 V R3 V R4 V R5 V R6 V R7 V R8 V R9
Hence

$$R= \begin{array}{|ccccc|} v1 & v2 & v3 & v4 & v5 \\ \hline 0 & 0.5 & 1 & 0.5 & 1 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \\ \hline \end{array} \quad \begin{array}{|ccccc|} \hline 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ \hline \end{array} \quad \begin{array}{|ccccc|} \hline 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \\ \hline \end{array} \quad \begin{array}{|ccccc|} \hline 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ \hline \end{array} \quad \begin{array}{|ccccc|} \hline 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 1 & 0.5 & 1 & 0.5 & 0 \\ \hline \end{array}$$

O_1 O_2 O_3 O_4 O_5

If present conditions are
L'=30 & V'=190 then Output grid voltage = ?
Let

$L'=\{0/L_1+0/L_2+1/L_3+0/L_4+0/L_5\}$
$\quad=[0\ \ 0\ \ 1\ \ 0\ \ 0]$

$V'=\{0/V_1+0/V_2+0/V_3+0.5/V_4+0/V_5\}$
$\quad=[0\ \ 0\ \ 0\ \ 0.5\ \ 0]$

Output grid viltage = O'= V'∘ (L'∘ R)

Now

(L'∘ R) = [0 0 1 0 0] ∘ R

…(since R has obtained)

$$L'\circ R) = \begin{vmatrix} 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 1 & 0.5 & 0 & 0 & 0 \end{vmatrix}$$

Output grid voltage (G) = V' ∘ (L' ∘ R)

$$= [\,0\ 0\ 0\ 0.5\ 0\,]\circ \begin{vmatrix} 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 1 & 0.5 & 0 & 0 & 0 \end{vmatrix}$$

$$= \begin{array}{ccccc} [\,0.5 & 0.5 & 0.5 & 0.5 & 0\,] \\ O_1 & O_2 & O_3 & O_4 & O_5 \end{array}$$

Deffuzification:

Here output grid voltage value is in the form of fuzzy set. We defuzzify G by taking its center of gravity with the weighted value, we get a definite value of output grid voltage as follows.

O/P Grid Voltage =

$$\frac{0.5xO_1 + 0.5xO_2 + 0.5xO_3 + 0.5xO_4 + 0xO_5}{(\ 0.5 + 0.5 + 0.5 + 0.5 + 0\ )}$$

Note that ,
$O_1= -50,\ O_2 = -25,\ O_3 = 0,\ O_4 = 25,\ O_5 = 50$

We get definite value as ,

Output Grid Voltage ,

$$G = \frac{0.5x(-50) + 0.5x(-25) + 0.5x(0) + 0.5x(25) + 0x(50)}{(\ 0.5 + 0.5 + 0.5 + 0.5 + 0\ )}$$

$$= \frac{(-25) + (-12.5) + 0 + 12.5 + 0}{2}$$

$$= \ -12.5$$

This result, for the outside light L'= 30 and Grid voltage V' = 190 V, the output grid voltage will decrease by 12.5 V from the present state.

Conclusion:

There are various defuzzification methods. Here we are using Mamdani's Method. As it is easy to handle compared to other methods used for defuzzification. The inputs atmospheric light and the grid voltage cannot be classified completely in one particular range of the membership function. The fuzziness occurs in this case, which is best resolved using fuzzy logic. Thus we can implement fuzzy logic in real system.

References

1. An Introduction to Fuzzy Logic for practical application-K.Tanaka Springier
2. Introducton to applied Fuzzy electronics-A.M. Ibrahim, Prentice Hall of India,New Delhi.
3. Fuzzy sets and Fuzzy Logic-G.J. Klir and B. Yuan, Prentice Hall of India
4. Fuzzynet technical case studies http://www.aptronix.com/

5 Nikunja K. Swain, “A Survey of Application of Fuzzy Logic in Intelligent Transportation Systems (ITS) and Rural ITS” Southeast Con, Proceedings of IEEE, 2006

6. Dalal, S.; Satyanarayana, S. “Application of fuzzy logic to picturequality improvement in televisions”,

7. Weijing Zhang, Applications Engineer, Aptronics Incorporated, Copyright © 1992 by Aptronix Inc.H

CANCER DETECTION IN MEDICAL IMAGES USING FUZZY LOGIC

Mr. R.RAMESH BABU
Assistant proffessor in Electronics and Communiction Engg
Jaya Engineering College
Chennai
rammathi10@yahoo.co.in & rameshbabumtech@gmail.com

Abstract*:* The field of medical imaging gains its importance with increase in the need of accurate and efficient diagnosis over a short period of time. MR imaging has become a widely-used method of high quality medical imaging, especially in brain imaging where MR's soft tissue contrast and non-invasiveness are clear advantages. MRI segmentation is an important image processing step to identify anatomical areas of interest for diagnosis of many disorders such as brain tumor, multiple sclerosis, etc. Segmentation approaches have met with only limited success because of overlapping intensity distributions of intracranial and extra cranial tissues in order to make robust automatic brain tumor and healthy tissue segmentation. Hence the extra cranial tissues should be removed in brain MR images and making use of only intracranial tissue regions for segmentation of tumor and normal tissues for further analysis. In this paper, from the abnormal MR images, the features are extracted. Here, the feature extraction includes the first order and second order features. First order texture measures are statistics calculated from the original image values, like variance, and do not consider pixel neighbor relationships. Second order measures consider the relationship between groups of two (usually neighboring) pixels in the original image. The principal components are selected. Then, an efficient segmentation algorithm for magnetic resonance images of brain tissues using fuzzy logic is proposed. The fuzzy logic output will be compared with K- means, Neural Network for segmentation. The comparative analysis will be done in terms of performance measured parameters.

Keywords: image segmentation, clustering, fuzzy logic.

1. Introduction

Magnetic resonance imaging **(MRI)** is used as a valuable tool in the clinical and surgical environment because of its characteristics like superior soft tissue differentiation, high spatial resolution and contrast and it does not use harmful ionizing radiation to patients. The data obtained from MR images are used for detecting tissue deformities such as tumors, cancers and injuries. In order to understand MRI contrast, it is important to have some understanding of the time constant involved in relaxation processes that establish equilibrium following RF excitation. Time constants involved in MR images are T1(realign time),T2 (relaxation time) and T2f (flair). The tumors detected using MRI are the following Primary malignant tumors (glioma and meningioma) and secondary malignant tumours (metastatic brain tumor). MRI segmentation is an important image processing step to identify anatomical areas of interest for diagnosis of many disorders such as brain tumour, multiple sclerosis, etc.

1.1. Feature extraction: To classify an object in a image, we must first extract some features out of the image. Feature extraction is a special form of dimensionality reduction and features reflect properties measured at the pixel-level that can aid in discriminating between normal pixels and tumor pixels. First order texture measures are statistics calculated from the original image values, like variance, and do not consider pixel neighbour relationships. Second order measures consider the relationship between groups of two (usually neighbouring) pixels in the original image.

1.2. Cluster analysis: Cluster analysis is an interdependence technique. It is similar to multi dimensional scaling .The difference is that multi dimensional scaling identifies underlying dimensions, while cluster analysis identifies clusters. The goal of clustering is to reduce the amount of data by categorizing or grouping similar data items together. Clustering approaches have met with only limited success because of overlapping intensity distributions of intracranial and extra cranial tissues and hence the extracranial tissues are to be removed.

2. Overview of other method

Clustering can be considered the most important unsupervised learning problem, so it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. Clustering algorithms may be classified as listed below

2.1. K-Means segmentation

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main advantages of this algorithm are its simplicity and speed, which allows it to run on large datasets. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

Disadvantage of this algorithm is that the resulting clusters depend on the initial assignments. But does not ensure that the solution given is not a local minimum of variance. Several misclassified data points after segmentation of brain image. The K means algorithm is given below

Step1: Choose K Initial Centers $Z_1(1)$, $Z_2(2)$--- They are arbitrary
Step2: At the K^{th} iterative Step, distribute the sample {X} among the K Cluster domain ,using the relation $X € S_j(k)$

if $|| X- z_j(k) || < || X-z_i(k) ||$, Where $s_j(k)$-the set of samples whose cluster center is $z_j(k)$
Step3: from the result of step 2 , calculate the new clusters $z_j(k+1)$,j=1,2-----k

$z_j(k+1)=1/n_j \sum x \quad X€ S_j(k)$

where n_j number of samples in $s_j(k)$,cluster centers are sequentially updated.

Step4: if $z_j(k+1)= z_j(k)$,the algorithm has converged and procedure is terminated otherwise go to step 2.

2.2. RBF Neural Network

A self adaptive RBF neural networks method for segmenting the brain image. This method has been proposed in order to reduce computation time, ability to recognize overlapping pattern classes or degraded images. Radial Basis Function is used for approximating function and recognizing patterns .it uses Gaussian potential functions. Gaussian potential functions are also used in networks called regularization network. The task is to determine a functions in linear space such that $s(x_i)=t_{i.}$ i=1-----n. The interpolation function is a linear combination of basis functions(x)= $\sum w_i v_i(x)$,As basis function RBF of form is$V_i(x) = \varphi (|| x-x_i||)$ Where mapping R+-- R and the norm is Euclidean distance.

The architecture of RBF Network consist of 3 layers they are input, hidden and output layers as shown in Fig 2.1

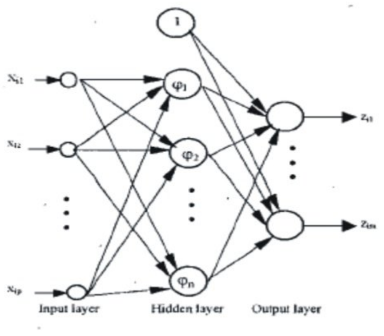


Figure 2.1 Architecture of RBF Neural Network

It is a multilayer feed forward network. There exists n number of input neurons and m number of output neurons with hidden layers existing between the input and output layers the interconnection between the input and hidden layers forms hypothetical connection and between the hidden layer and output layers forms weighted connections. The training algorithm used for update of weights in all the interconnection

3. Proposed method:

3.1. Fuzzy Logic

Fuzzy set theory provides a host of attractive aggregation connectives for integrating membership values representing uncertain information. These connectives can be categorized into the following three classes *union, intersection* and *compensation* connectives. The membership function of a fuzzy set in a functional form, typically a bell-shaped function, triangle-shaped function, trapezoid-shaped function, etc.

When fuzzy systems are applied to appropriate problems, particularly the type of problems described previously, their typical characteristics are faster and smoother response than with conventional systems. This translates to efficient and more comfortable operations for such tasks as controlling temperature, cruising speed, for example. Furthermore, this will save energy, reduce maintenance costs, and prolong machine life. In fuzzy systems, describing the control rules is usually simpler and easier, often requiring fewer

rules, and thus the systems execute faster than conventional systems. Fuzzy systems often achieve tractability, robustness, and overall low cost.

The procedure for obtaining the fuzzy output of such a knowledge base can be formulated as

1. The firing level of the *i*-th rule is determined by

$A_i(x_0) \times B_i(y_0)$.

2. The output of of the *i*-th rule is calculated by

$C\_i(w) := A_i(x_0) \times B_i(y_0) \rightarrow C_i(w)$

for all w □ W.

3. The overall system output, C, is obtained from the individual rule outputs C_i by

$C(w) = \mathbf{Agg}\{C_1, \ldots, C_n\}$for all w □ W.

The methodology proposed in this paper is as explained in Fig 2.2:

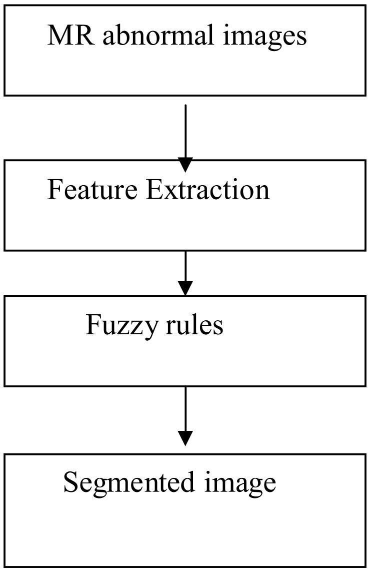


Fig 3.1 Flow Diagram for the proposed methodology

Here, in this paper, basically, the clustering of abnormal brain MR images is done using fuzzy logic which is the proposed method. From the abnormal brain MR images the first order and the second order features were extracted. First Order Features include mean, variance, skewness, kurtosis, energy and entropy while the second order features include angular second momentum, entropy, contrast, cluster shade, cluster prominence, inertia, and local homogeneity. The features were extracted and using them a feature vector was formed.Then the selected features are segmented using Fuzzy logic.Then,a comparative study is made using K-means, neural network and Fuzzy logic.

4. Result and Conclusion:

The various clustering methods are analyzed. These methods are used to perform tissue classification in MRI. The experimental results are shown in figure 4.1. The major disadvantage of K means algorithm had several misclassified data points after segmentation of brain image. The Neural network clustering also taken more time to perform the classification distinct of tissue types.

The proposed methodology is applied for T2 and T2 Flair images of slices of 150 MRI volumes. The results are compared with the radiologist labeled ground truth on pixel by pixel basis. The number of true positives (ground truth tumor pixels found algorithmically), false positives (pixels isolated as tumor though not within ground truth boundaries) and false negatives (ground truth tumor pixels not found algorithmically) for each slice of a patient-1, patient-2, patient-3 and patient-4 on a pixel level are calculated

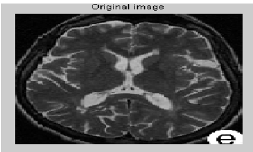


Figure 4.1 input image

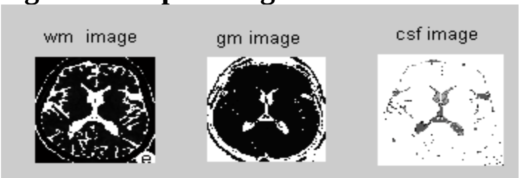


Figure 4.2 Clustering Using Fuzzy logic (No of clusters=3)

The following table shows the misclassification rate performed by various methods

Method	MR White	MR Gray	MR Csf
K-Means	30	35	99
Neural Network	10	35	38
Fuzzy Logic	0	31.46	35.8

MR – misclassification rate

References

[1]J.G.Webster, Ed. , *"Medical Instrumentation: Application and Design"*.New York:John Wiley & Sons, Inc. , 1998, pp.551-555.

[2]Haack.E et al. , 1999," *Magnetic Resonance Imaging, Physical Principles and Sequence Design"*. Wieley-Liss, New York

[3]Rui Xu, and Donald Wunsch II,"*Survey of Clustering Algorithms"*,IEEE Transaction on Neural Networks, VOL.16, NO. 3, MAY 2005

[4] Boris Cigale and Damjan Zazula, *"Segmentation of ovarian ultrasound images using cellular neural networks",* International Journal of Pattern Recognition and Artificial Intelligence Vol.18

[5] A. Slavova and V. Mladenov, *"Cellular Neural Networks: Theory and Applications"*, Nova Science Publishers, Inc., USA, 2004

[6]X.Zhang,X.L.Xiao,J.W.Tian,J.Liu and G.Y.Xu, *"Application of support vector machines in classification of Magnetic Resonance Images"*,International journal of computers and applications,vol 28,No.2.2006.

[7]Mark Schmidt, *"Automatic Brain Tumor Segmentation"*, Master Thesis,University of Alberta,CANADA,2005

[8] Boris Cigale and Damjan Zazula, *"Segmentation of ovarian ultrasound images using cellular neural networks",* International Journal of Pattern Recognition and Artificial Intelligence Vol.18

[9] M. Hanggi and G. S. Moschytz, Kluwer,*"Cellular Neural Networks",* 2000, International Journal of Electrical Engineering Education

Principle Component Analysis Based Face Detection Algorithm

J.Bhaskara Rao
Asst. Professor
janabhaskar@gmail.com
Dept of ECE, Pydah college of Engg and tech
Visakhapatnam

Jayasuryadutt Meralla
Student, ME (EI),
jayasuryadutt@gmail.com
ECE Department , College of Engineering (A),
Andhra University, Visakhaptnam-530 003

S. Santa Kumari
Associate Professor
santakseetala@yahoo.com
ECE Department , College of Engineering (A),
Andhra University, Visakhaptnam-530 003

ABSTARCT

Face recognition is a fast growing field, with many different applications to its use in society. In this paper I proposed a method to recognize a person's facial expressions by the eigenfaces using Principal Components Analysis (PCA). The Principal Components Analysis (PCA) is one of the most successful techniques that have been used to recognize faces in images. This technique consists of extracting the eigenvectors and eigenvalues of an image from a covariance matrix, which is constructed from an image database. These eigenvectors and eigenvalues are used for image classification, obtaining nice results as far as face recognition is concerned.

Keywords: Principal Components Analysis, eigenfaces, face recognition.

1.INTRODUCTION

This paper utilizes Eigenface as a method of classifying facial expression. Firstly, the train images are utilized to create a low dimensional face space. This is done by performing Principal Component Analysis (PCA) in the training image set and taking the principal components (i.e. Eigen vectors with greater Eigen values). In this process, projected versions of all the train images are also created. Secondly, the test images also are projected on the face space – as a result, all the test images are represented in terms of the selected principal components. Thirdly, the Euclidian distance of a projected test image from all the projected train images are calculated and the minimum value is chosen in order to find out the train image which is most similar to the test image. The test image is assumed to fall in the same class that the closest train image belongs to. Fourthly, in order to determine the intensity of a particular expression, its Euclidian distance from the mean of the projected neutral images is calculated. The more the distance - according to the assumption - the far it is from the neutral expression. As a result, it can be recognized as a stronger the expression.

2.Eigenfaces for Face Detection/ Recognition

Computation of the eigenfaces

Step 1: obtain face images $I_1, I_2, \ldots\ldots.., I_M$

Step 2: represent every image I_i as a vector Γ_i

Step 3: compute the average face vector Ψ:

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i$$

(1)

Step 4: subtract the mean face:

$$\Phi_i = \Gamma_i - \Psi \qquad (2)$$

Step 5: compute the covariance matrix C:

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T \quad (N^2\text{x}N^2 \text{ matrix})$$ **(3)**

Where **A = [Φ_1 Φ_2 . . . Φ_M] (N^2x*M* matrix)**

Step 6: compute the eigenvectors u_i of AA^T

The matrix AA^T is very large --> not practical !!

Step 6.1: consider the matrix $A^T A$ (*M*x*M* matrix)

Step 6.2: compute the eigenvectors v_i of $A^T A$

$\boldsymbol{A^T Av_i = \mu_i v_i}$ ***(4)***

What is the relationship between us_i and v_i?

$\boldsymbol{A^T Av_i = \mu_i v_i \Rightarrow AA^T Av_i = \mu_i Av_i \Rightarrow}$

$\boldsymbol{CAv_i = \mu_i Av_i}$ **or** $\boldsymbol{Cu_i = \mu_i u_i}$ **where** $\boldsymbol{u_i = Av_i}$ **3.6)**

Thus, AA^T and A^TA have the same eigenvalues and their eigenvectors are related as follows:

$\boldsymbol{u_i = Av_i}$ **!!** **(5)**

Step 6.3: compute the *M* best eigenvectors of AA^T : $u_i = Av_i$

Step 7: keep only *K* eigenvectors)

Representing faces onto this basis

Each face (minus the mean) Φi in the training set can be represented as a linear combination of the best *K* eigenvectors:

$$\hat{\Phi}_i - mean = \sum_{j=1}^{K} w_j u_j, \ (w_j = u_j^T \Phi_i)$$ **(6)**

Each normalized training face Φ_i is represented in this basis by a vector:

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \cdots \\ w_K^i \end{bmatrix}, \quad i = 1, 2, \ldots, M$$ **(7)**

Face Recognition Using Eigenfaces:

Given an unknown face image Γ (centered and of the same size like the training faces) follow these steps:

Step 1: normalize **Γ: $\Phi = \Gamma - \Psi$**

Step 2: project on the eigen space

$$\hat{\Phi} = \sum_{i=1}^{K} w_i u_i \quad (w_i = u_i^T \Phi)$$ **(8)**

Step 3: represent Φ as:

$$\Omega = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_K \end{bmatrix}$$ **(9)**

Step 4: find $\boldsymbol{e_r = \min_l \|\Omega - \Omega^l\|}$

Step 5: if $e_r < T_r$, then Γ is recognized as face *l* from the training set.

The distance e_r is called distance within the face space (difs)

Comment: we can use the common Euclidean distance to compute e_r, however, it has been reported that the *Mahalanobis*

$$\|\Omega - \Omega^k\| = \sum_{i=1}^{K} \frac{1}{\lambda_i}(w_i - w_i^k)^2$$ **(10)**

(Variations along all axes are treated as equally significant)

Face Detection Using Eigenfaces

Given an unknown image Γ

Step 1: compute $\boldsymbol{\Phi = \Gamma - \Psi}$

Step 2: compute

$$\hat{\Phi} = \sum_{i=1}^{K} w_i u_i \quad (w_i = u_i^T \Phi)$$ **(11)**

Step 3: compute $\boldsymbol{e_d = \|\Phi - \hat{}\Phi\|}$

Step 4: if $e_d < T_d$, then Γ is a face.

The distance e_d is called distance from face space (dffs)

Euclidean Distance

Let an arbitrary instance x be described by the feature vector

x = [a_1(x), a_2(x), . . . , a_n(x)]

where a_r(x) denotes the value of the rth attribute of instance x. Then the distance between two instances x_i and x_j is defined to be d(x_i, x_j):

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2}$$

3.Imlpementation of Face Recognition Using Eigenfaces

The entire sequence of training and testing is sequential and can be broadly classified as consisting of following two steps:

1. Database Preparation
2. Training
3. Testing

The steps are shown below.

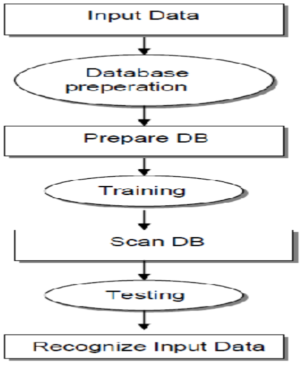


Fig 4.1: Flowchart indicating the sequence of implementation

3.1. TRAINING

1. Select any one (.bmp) file from train database using open file dialog box.
2. By using that read all the faces of each person in train folder.
3. Normalize all the faces.
4. Find significant Eigenvectors of Reduced Covariance Matrix.
5. Hence calculate the Eigenvectors of Covariance Matrix.
6. Calculate Recognizing Pattern Vectors for each image and average RPV for each person
7. For each person calculate the maximum out of the distances of all his imageRPVs from average RPV of that person.

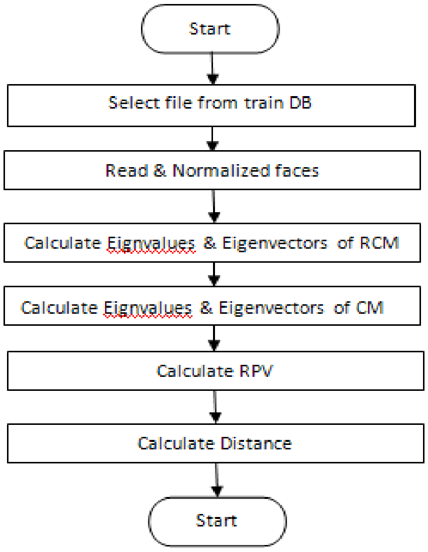

Start
Select file from train DB
Read & Normalized faces
Calculate Eignvalues & Eigenvectors of RCM
Calculate Eignvalues & Eigenvectors of CM
Calculate RPV
Calculate Distance
Start

Fig 4.3: Flowchart for training

3.2. TESTING

Testing is carried out by following steps:

1. Select an image which is to be tested using open file dialog box.
2. Image is Read and normalize.
3. Calculate the RPV of image using Eigenvector of Covariance Matrix.
4. Find the distance of this input image RPV from average RPVs of all the persons.
5. Find the person from which the distance is minimum.
6. If this minimum distance is less than the maximum distance of that person calculated during training than the person is identified as this person.

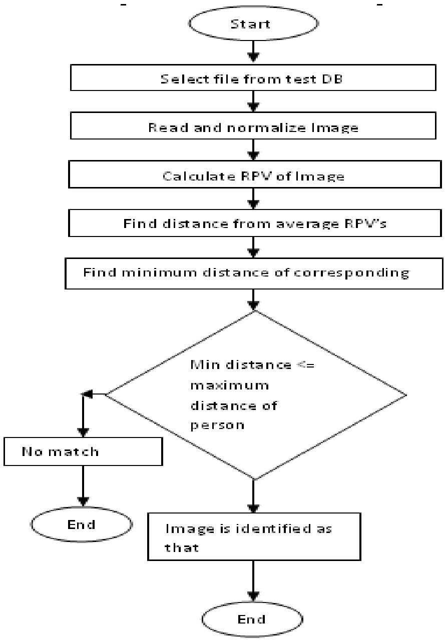
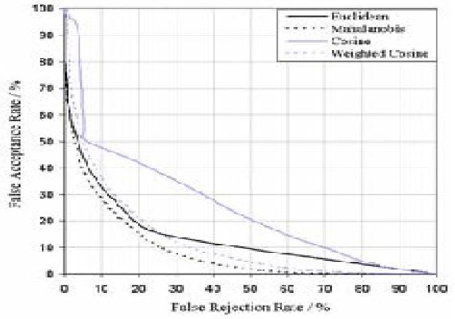

Start
Select file from test DB
Read and normalize Image
Calculate RPV of Image
Find distance from average RPV's
Find minimum distance of corresponding
Min distance <= maximum distance of person
No match
End
Image is identified as that
End

Fig 4.4: Flowchart for testing

RESULTS:



APPLICATION AREAS

1. Face Identification
2. Access Control
3. Security
4. Surveillance
5. Smart Cards
6. Law Enforcement

CONCLUSION

Face recognition is a fast growing field, with many different applications to its use in society. In this paper I proposed a face recognition method using eigen faces. This face recognition method is used to recognize the faces of various persons with different expressions like happy, anger, disgust, sad with Principle component analysis (PCA). The Euclidean Distance from the neutral image is also calculated. This face recognition method is the accurate method to recognize the persons.

REFERENCES

1. ”Eigenfaces for Recognition” (M. Turk, A. Pentland)
2. M. Turk, A. Pentland. ”Eigenfaces for Recognition”. Journal of Cognitive Neuroscience. Vol 3, No. 1. 71-86, 1991.
3. M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. of Computer Vision and Pattern Recognition*, pages 586–591. IEEE, June 1991b. URL http: //www.cs.wisc.edu/~dyer/cs540/handouts/mturk-CVPR91.pdf. (URL accessed on November 27, 2002).
4. AIDC, .Eigenface Recognition., Online Internet, 12 February 2005, Available: http://et.wcu.edu/aidc/BioWebPages/eigenfaces.htm.
5. Zhujie, Y.L.Y., 1994. Face recognition with eigenfaces. Proc. IEEE Intl. Conf. Industrial Technol.,
6. Face Recognition Edited by Miloš Oravec
7. BIOMETRICS *Person a1 Identification in Networked Society edited by* Anil K. Jain *and* Ruud Bolle and Sharath Pankanti

Enhancing Security Services in Internet based Communication Using Efficient Pseudo Embedding and Novel Compression Techniques

Mr.Gandhimathinathan [1], MrAjithanyaKumar.M.K [2]

MR.GANDHIMATHINATHAN [1]
Lecturer, ECE Dept
St.Joseph Engineering College
Mangalore-57502
agandhimathinathanbe@gmail.com

MR.AJITHANJAYAKUMAR.M.K [2]
Asst.Professor, E &E Dept
St.Joseph Engineering College
Mangalore-57502
ajithanjaya@yahoo.co.in

Abstract

The objective of this paper is two folds. First we propose a novel secured compression model wherein highly confidential text documents (business letters, banking transaction papers, military information) and error concealment information are embedded using polynomial representation into an image to the maximum safe limit, efficiently using the Pseudo Embedding algorithm. This pseudo embedded image is binarized. The binarized image is fed into the digital image processor. Here, this binarized image stream is compressed using proposed dynamic tree construction generated by computing relative frequency of distinct character set (obtained using the histogram evaluation) and iterative tree construction procedure. Secondly by applying tree coded binarized pseudo embedded image to hex-based coding system, extremely high compression ratios higher than those provided by the existing schemes were observed. The information bits are further subjected to Turbo encoding schemes for error detection and correction. Efficient authentication cryptographic procedures such as the RSA have been employed for transmission of code dictionaries.

Keywords: Pseudo-Embedded, Turbo – Encoding and Compression Techniques.

I. INTRODUCTION

Recent trends demand data to be transmitted over networks in a compressed and secured form. Military systems, online banking, online business transactions highly rely on data security.

A new possibility of digital imaging and data hiding opens wide prospects in content management and secure communications. The message hidden in a cover medium may be a plain text, cipher text, or anything that can be represented as bit stream. When embedding data, it is important to take into contemplation the following issues:

- The cover data should not be significantly degraded by the embedded data, and the embedded data should as imperceptible as possible.
- The embedded data should directly be encoded into a header or wrapper to maintain data consistency formats.
- The embedded data should be as immune as possible to modifications from intelligent attacks or anticipated manipulations like filtering and re-sampling.

Recent data hiding techniques indicate that the bit replacement or the bit substitution is inherently insecure with safe capacities, far smaller than previously thought. Further, these techniques demand lesser information to be embedded into the cover data so that the probability of introducing detectable artifacts by the embedding process is less (i.e., the size of the data to be embedded is strictly restricted to a fraction of the cover file). The proposed technique is designed to meet the above requirements and to overcome the disadvantages faced by other embedding techniques. Once the data is embedded in the image, the latter is to be stored in a compressed format, so that it occupies less number of bits. Similarly, transmitting a compressed file significantly reduces the transmission bandwidth. Since the data embedded in the image are secure text files, we don’t opt for lossy image compression schemes since the loss of even a single bit may be a disaster for highly secure information.

The currently available lossless compression schemes take into account, the diversified file characteristics like frequency etc. But, the proposed dynamic tree compression technique uses a generic algorithm that exploits the important characteristic called relative frequency. So, it always provides the best compression unlike the others.

Further compression may also be achieved using the proposed low hierarchy compression scheme that uses hex codes instead of ASCII codes to achieve much better compression levels. The compressed image carrying the secure data (one of the outputs from the compressor) should then be transmitted from the transmitting end to the receiving end. So, some error detection and correction codes are added to regain the data if some bits are lost during transmission. The other output of the data compressor is the dictionary, which requires more care during transmission. So, to transmit the dictionary in this scheme, RSA algorithm that belongs to the cryptographic family is used so that even if the compressed image is hacked by some eavesdropper, the dictionary won’t be available to decompress the image. Thus our proposed scheme introduces data security with multiple levels of data hiding so that it caters the needs of almost all the areas.

II. THE PROPOSED TECHNIQUES

Our proposed effective internet communication technique includes some novel techniques at the transmitting end and the receiving end

A. Transmitting end

1. Pseudo embedding
2. Histogram evaluation
3. Dynamic tree compression
4. Low hierarchy compression
5. Turbo encoding
6. RSA authentication scheme

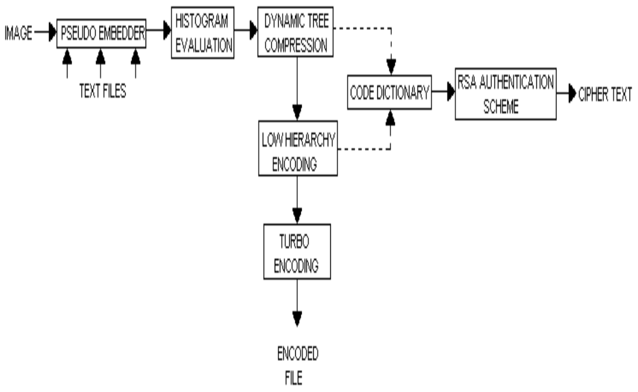


Fig 1: Transmitter Block

B. Communication channel:

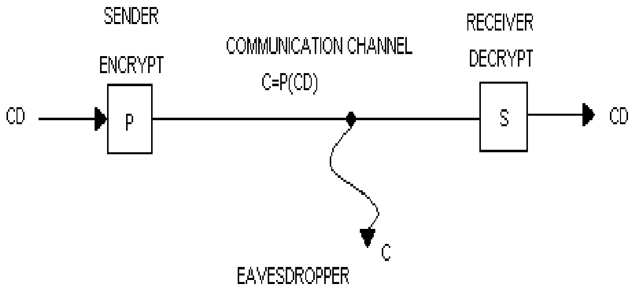


Fig 2: Channel Block

C. Receiving end

1. RSA decryption
2. Turbo decoding
3. Low hierarchy decompression
4. Dynamic tree decompression
5. Pseudo extraction

Error! Objects cannot be created from editing field codes. Fig 3: Receiver Block

III. PSEUDO EMBEDDING

The data file to be hidden onto the image and the image (host) file are read in the binary mode (as bits). A table is constructed for all the bits in the data file and the bit positions are correspondingly tabulated (taken as x-values).

For each bit in the text file, a traversal is made on the image file and when a bit match is found, an entry of the corresponding bit position of the matched image bit is made as $f(x_i)$ against the text bit position x_i. Having got a match, the search for subsequent bits of the text file is performed such that the $f(x_i)$ values for i=1 to m, yield a forward difference table in which the P^{th} differences of $f(x)$ are constant such that P<M where M is the number of bits of the text file, thus generating a polynomial of degree P.

In other words, the polynomial generated is going to perfectly correspond to the (x_i, $f(x_i)$) values for all i=1 to M so that this provides lossless encoding. The forward difference table is constructed using the following formula. The j^{th} order forward difference of *f* evaluated at x_k is

$\Delta^j f(x_k) = \sum_{i=0-j} (-1)^j\, jC_i\, y_{k+j-i}$ &

$f(x) = (1+xC_1\Delta+xC_2\Delta^2+\ldots..\Delta^p)\, f(x_0)$.

This polynomial that is obtained is partitioned into bits representing the highest degree and all the coefficients of the polynomial. This bit pattern can be considered as another data file, and then by applying the procedure once again, a still reduced bit pattern can be obtained. The data flow diagram of Pseudo embedding algorithm is given below. The various blocks in the diagram are intended to represent the logical steps involved in the algorithm. The image file (cover) and the data file to be hidden are represented in binary format and sent to the Optimal Bit Mapped Block, which sequentially maps bits in the data file to those in the image file.

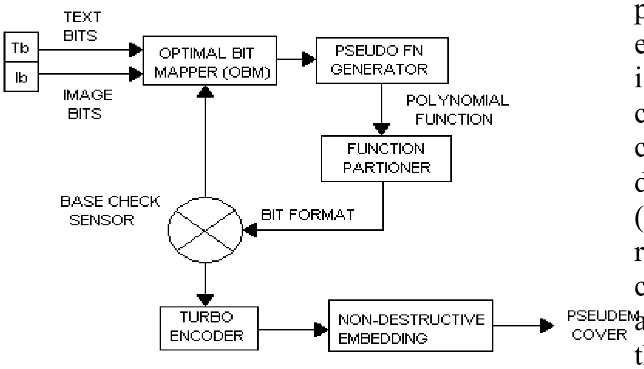


Fig 4: Principles of Pseudo Embedding

It is done such that, for each bit in the former, there is a positive integer value based on the bit position in the latter. The OBM chooses the matches in such a way that the higher order differences in the difference table vanishes thus ensuring convergence and a polynomial function of a smaller degree. The OBM uses the back propagation algorithm of neural networks for the intelligence. These set of positive integer values form an increasing trend so that the set of pairs of bit positions in the text *x* and the position indexed values in the cover *f*(*x*) generated can be applied to the working function generator block which uses Newton's forward difference to construct the appropriate polynomial to fit the data set.

The position indexed values $f(x_i)$ can exceed the maximum number of bits *N*, in the image file wherein bit mapping

Actually made indicates that bit in the position in the hypothesis: $f(x) = F_c(x) + N * R$, where F_c(x) indicates the actual position of the bit in the cover and *R*, the numbers of recursions on the image file. The working function i.e., the polynomial *f*(*x*) generated is then partitioned as bit pattern in the Function partitioned form.

Table 1: Work function the given Input level

INTENSITY LEVEL	FREQUENCY
121	837
156	987
83	244

This bit pattern is checked in the Base Check Sensor for basis function (the polynomial of the desired degree) requirements. It is fed into the OBM where it is considered as another text data file and the above procedure is recursively performed until the basis function is obtained as indicated by the Base Check Sensor.

Once the basis function is obtained, it is partitioned into bit format storing the degree and coefficients of the polynomial. This bit pattern is appended with the range of *x*, the number of recursions(*r*) of the procedure to give the Master Bit Pattern or the Basic Builder Set. If this MBP is brought down to '*m*' bytes, we embed just the '*m* x 8' bits instead of M bits of the text file (*M*>>*m*).

The MBP thus obtained has to be safely guarded since any bit change in it would lead to distorted retrieval of the hidden bit. Hence, the obtained MBP is encoded with error correction bits using turbo encoder.

Newton's Forward Difference Formula is: **F(x)** = $y_0 + {}_nC_1\ y_0 + {}_nC_2\ {}^2y_0 + \ldots$ substituting, we get **F(x) = x^2 + 2x + 2.**

MBP:

X	F(X)
1	5
2	10
3	17
4	26
5	37
6	50
7	65
8	82
9	101
10	122
11	145
12	170
13	197
14	226
15	254
16	290

Table 2: Difference Table

0	1	1	0	1	0

A. HISTOGRAM EVALUATION:

Histogram evaluation is a process of evaluating an image by identifying the distinct intensity levels and the frequency of occurrence of each intensity level in the image. In our data hiding scheme, the embedded image (output of the pseudo embedded algorithm) is fed as the input to the histogram evaluator, which tests each pixel of the image and identifies the distinct intensity levels in the image. The intensity levels (0-255) now become the ASCII character set. All the distinct intensity levels & thenumber of pixels corresponding are tabulated similar to the example shown below.

IV.DYNAMIC TREE COMPRESSION TECHNIQUE

Dynamic tree codes

In this technique the static trees are dynamically created and tested for the compression of the given ASCII file. The trees are formed by the branch-or-block trade off, that is, by varying the number and kind of nodes that are blocked and branch at every level thus giving every set of diversified tree. The trees thus obtained are tested as in the previous technique to compute the compression ratio and the tree that has the least ratio is finally used to compress the source file.

Algorithms:

A. Finding the number of distinct characters and its frequency

- Initialize the variable count=0.
- Read and store the character temporarily in variable temp
    - If count=0 then store the character read in temp to the structure containing the variable ch. Increment count.
    - If count! =0 then check the variable temp with the structure character ch.
- Repeat step 4 until count is satisfied or any match is found.

F(x)	
5	^{2}F(x)
7	2
9	2
11	2
13	2
15	2
17	2
19	2
21	2
23	2
25	2
27	2
29	2
31	2
33	2

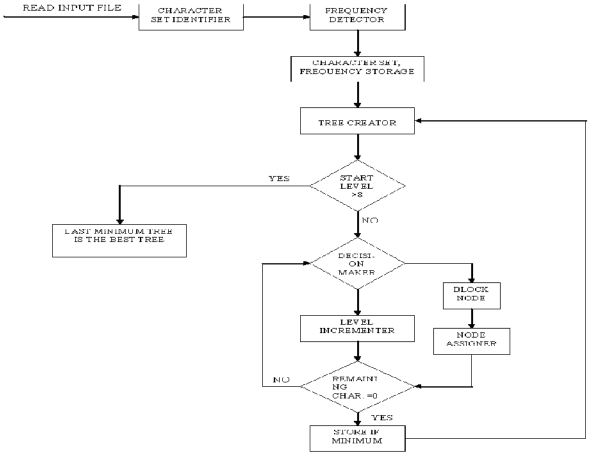
- If any match is found then increment the structure variable frequency=frequency+1 for the corresponding character.
- If no match is found until count is satisfied then ch=temp and initialize its frequency=1 and increment the variable count=count+1.
- Repeat steps 2 to 7 until the End Of File pointer is reached.
- The variable count will give the number of distinct characters in the file.

This algorithm determines the number of distinct characters and each characters frequency by sequentially reading the given input file and storing the characters and their frequency for finding the best tree and for decompression.

B. Finding the equivalent binary and decimal codes

- Initialize the two dimensional array binary, variable i=1 & variable j=i+1.
- If frequency[j]>frequency[i] then exchange frequency[i] and frequency [j].
- Repeat step 4 until j=count & repeat steps 3 to 5 until i =count.
- Traverse the static tree selected & store the binary equivalent of each character in structure two dimensional arrays binary.
- Binary values are assigned directly as the characters were already sorted based on their frequency.
- The structure variable digit =length of the binary code.
- If digit<=8 then ascii1=ASCII value of the code.
- If digit>8 then ascii1=ASCII value of the first 8 bits and ascii2=ASCII value of the remaining bits.

This algorithm assigns binary codes for each character by traversing through the binary static tree.



Flowchart 1: Algorithm for Binary code Generation

C. Finding the best static tree

Form tree (int i, long int siz, int level, int node)
The above function form tree is used to determine the best static tree. This function is called as form tree (0, 0, 1, 2) i.e. 0 characters (i) are assigned nodes and so their size is 0(siz). The node allocation starts from level 1(level) with 2 nodes (node).

1. if ((file count-i)<=node) then while(i<=file count) {dist[i++].level=level;}
2. else check if(level<8) then
3. for a=node/2 to 0 do for j=0 to a do dist[i+j].level =level;
4. call form tree(i+a,siz,level+1,2*(node-a));
5. If level>8 then only half of the nodes at that level can be taken.
   For j=0 to node/2
   Dist [i+j].level=level;
6. Then number of characters assigned increases as x=i+(node/2)
7. call form tree (x,siz,level+1,2*(node/2));
8. When all the characters are assigned then the compressed file size can be calculated by simply multiplying each characters frequency with their length of the binary code and finally adding them all.
   For i=0 to file count do siz=siz+ (dist[i].count*dist[i].level)
9. To find the best static tree this siz is temporarily stored in min size and the levels are stored in an array best [].

D. Writing the final compressed file

1. The binary codes are written into a file named *temp.txt*.
2. The corresponding length of the binary codes is written into a file named digit.txt.
3. Temp.txt is read and the byte read is stored in a variable ch1.
4. Digit.txt is also read and the length of the bit code is obtained.
5. Initialize the variable k=0.
6. The variable is ch1 then shifted right ch1<< equal to 8-length of the binary code and stored in a variable j. Then variable k=j | k.
7. If 8-length of the binary code <0 then ch1 is shifted left ch1>> equal to the difference and stored in j=ch1>>difference.

8. If difference=1 then j=j&127. If difference=2 then j=j&63 and so on.
9. Then k=k | j. The value of k is written onto to the output file.
10. k=0 and j=0. j=ch1<< (8-difference) and k=k | j.
11. If difference=0 then k=k / ch1 and k is written on to the output file.
12. Repeat steps 3 to 13 until the EOF pointer is reached.
13. For decompressing the file the distinct characters are written to the output file in the decreasing order of frequency followed by special characters.

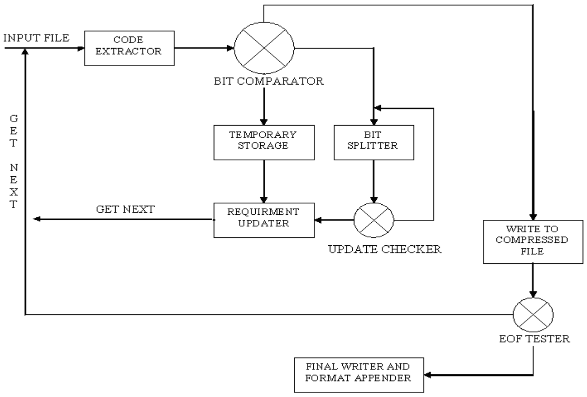


Fig 5: Principles of code extractor

E. Low Hierarchy Encoding

This new approach is designed to further compress a compressed file so that maximum compression can be achieved. After several stages of recompression using dynamic tree compression scheme, the compressed bit stream that cannot be further compressed using ASCII coding, is output. This bit stream is input to the hex code compressor that divides the bit stream into four bit patterns. Each four-bit pattern is substituted with the corresponding hex code alphabet (0-F). This gives a reduced hexadecimal character set (16) compared to ASCII character set (256).

The dynamic tree approach is again applied to compress this reduced character set file. This time, the tree constructed has lesser number of levels due to the application of hex code technique. This technique is applied recursively until maximum compression is achieved.

V. TURBO ENCODING

The turbo encoding adds parity bits to the compressed file. The code bits are generated from two encoders.

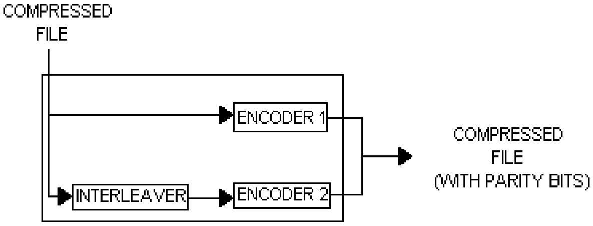


Fig 6: Principle of Turbo Encoding

A. Low Hierarchy Decomposition:

Based on the hex code dictionary received from RSA decryption the bits for higher level dynamic tree decompression are obtained.

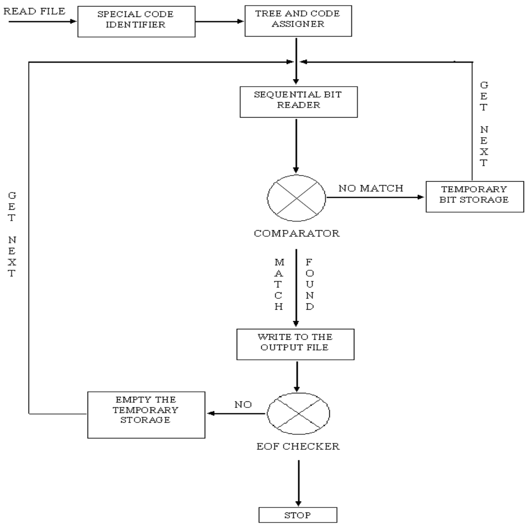
B. Dynamic Tree Decompression

This algorithm decompresses the given compressed file. For decompression first the special characters are searched and distinct characters are determined. Then the type of tree used is also determined. By finding this information the tree can be again formed and the decompression can be performed.

1. The special characters are searched and the characters are read in the structure variable ch. The type of static tree used is also found out.
2. With the help of the static tree and the characters obtained the same structure can be formed. The output file is opened and the bytes are read in the variable ch1.
3. Each bit of ch1 is examined. Compare with the available binary codes.
4. If a match is obtained then the equivalent character is written to the decompressed file.
5. If the length of binary code=8 then the corresponding character is written to the file.
6. If the length >8 then the next byte is read and (length-8) bits of LSB are read

from the next byte and the equivalent character is written to the file.

7. If the length<8 then the first character is read using step 7 and the remaining bits are taken and compared for a match and again the above steps are repeated.
8. Repeat steps 4 to 10 until the EOF pointer is reached.



Flowchart 2: Code Testing and code Correction

VI. PSEUDO EMBEDDED TEXT EXTRACTION

The pseudo cover file received at the receiving device is passed into the MBP extractor that retrieves the Basic Builder Set. This set enters the bits to the Bits to Function Converter, which converts the bits into a well-defined polynomial function in the range of *x* the function should take. t. The position is

$$F_c(x) = f(x) \; modN.$$

The obtained higher level bit pattern is sent as input to Bits to Function Converter and after similar procedures we get the next higher level bit pattern & the recursion continues until the final text data is obtained.

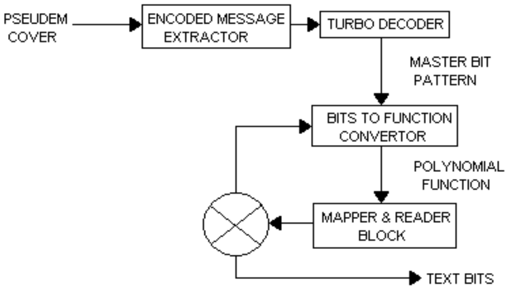


Fig 7: Pseudo Embedded Extraction

VII. CONCLUSION

The proposed model provides extremely high security to the transmitted code dictionary by use of strong cryptographic procedures. The approach of providing pseudo embedding scheme to hide the confidential text in an image provides immense security to hidden information. Most importantly compression provided by the iterative application of dynamic tree construction approach followed by hex code generation is extremely efficient compared to existing lossless compression schemes.

REFERENCES

[1] Y.Wang and P.Moulin, “Steganalysis of Block Structured Stegotext.” Proc. SPIE Conf., Vol. 5306, San Jose, CA, Jan. 2004.

[2] J.L. Cannons and P. Moulin, Design and Statistical Analysis of a Hash-Aided Image Watermarking System, to appear in IEEE Trans.On Image Processing, 2004.

[3] P.Moulin and J.A. O’Sullivan, Information-Theoretic Analysis of Information Hiding, Sep.1999 [Sep.99 postscript]; revised, Sep.2002. IEEE Trans. Information theory, Vol. 49, No. 3, pp.563-593, March 2003.

[4] P.Moulin, A Mathematical Approach to Watermarking and Data Hiding, ICASSP Tutorial, Orlando, FL, March 13, 2002

[5] Thoms H. Cormen, Charles E. Leiberson, Ronald L. Rivest & Clifford Stein, Introduction to Algorithms, 2nd edition(2004), Prentice Hall of India Pvt. Ltd. New Delhi.

Extraction of Text Regions in Natural Images

K.SANGEETHA
REG.NO.945102
BRANCH: COMMUNICATIONS AND SIGNAL PROCESSING
MASTERS OF TECHNOLOGY
G.PULLA REDDY ENGINEERING COLLEGE
KURNOOL
PH.NO:+91-9441611537
EMAIL ID.SANGEETHAKAMARTHI3@GMAIL.COM

Abstract— **The detection and extraction of text regions in an image is a well known problem in the computer vision research area. The goal of this project is to compare two basic approaches to text extraction in natural (non-document) images: edge-based and connected-component based. The algorithms are implemented and evaluated using a set of images of natural scenes that vary along the dimensions of lighting, scale and orientation. Accuracy, precision and recall rates for each approach are analyzed to determine the success and limitations of each approach. Recommendations for improvements are given based on the results.**

I. INTRODUCTION

Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos. This content can be in the form of objects, color, texture, shape as well as the relationships between them. The semantic information provided by an image can be useful for content based image retrieval, as well as for indexing and classification purposes [4,10]. As stated by Jung, Kim and Jain in [4], text data is particularly interesting, because text can be used to easily and clearly describe the contents of an image. Since the text data can be embedded in an image or video in different font styles, sizes, orientations, colors, and against a complex background, the problem of extracting the candidate text region becomes a challenging one [4]. Also, current Optical Character Recognition (OCR) techniques can only handle text against a plain monochrome background and cannot extract text from a complex or textured background [7].

Different approaches for the extraction of text regions from images have been proposed based on basic properties of text. As stated in [7], text has some common distinctive characteristics in terms of frequency and orientation information, and also spatial cohesion. *Spatial cohesion* refers to the fact that text characters of the same string appear close to each other and are of similar height, orientation and spacing [7]. Two of the main methods commonly used to determine spatial cohesion are based on edge [1,2] and connected component [3] features of text characters.

The fact that an image can be divided into categories depending on whether or not it contains any text data can also be used to classify candidate text regions. Thus other methods for text region detection, as described in more detail in the following section, utilize classification techniques such as support vector machines [9,11], k-means clustering [7] and neural network based classifiers [10]. The algorithm proposed in [8] uses the focus of attention mechanism from visual perception to detect text regions.

II. RELATED WORK

The purpose of this project is to implement, compare, and contrast the edge-based and the connected component methods. The other methods mentioned here are examples of text extraction techniques that can be used for future projects. Various methods have been proposed in the past for detection and localization of text in images and videos. These approaches take into consideration different properties related

to text in an image such as color, intensity, connected-components, edges etc. These properties are used to distinguish text regions from their background and/or other regions within the image. The algorithm proposed by Wang and Kangas in [5] is based on color clustering. The input image is first pre-processed to remove any noise if present. Then the image is grouped into different color layers and a gray component. This approach utilizes the fact that usually the color data in text characters is different from the color data in the background. The potential text regions are localized using connected component based heuristics from these layers. Also an aligning and merging analysis (AMA) method is used in which each row and column value is analyzed [5]. The experiments conducted show that the algorithm is robust in locating mostly Chinese and English characters in images; some false alarms occurred due to uneven lighting or reflection conditions in the test images.

The text detection algorithm in [6] is also based on color continuity. In addition it also uses multi-resolution wavelet transforms and combines low as well as high level image features for text region extraction. The text finder algorithm proposed in [7] is based on the frequency, orientation and spacing of text within an image. Texture based segmentation is used to distinguish text from its background. Further a bottom-up ‘chip generation’ process is carried out which uses the spatial cohesion property of text characters. The chips are collections of pixels in the image consisting of potential text strokes and edges. The results show that the algorithm is of bust in most cases, except for very small text characters that are not properly

detected. Also in the case of low contrast in the image, misclassifications occur in the texture segmentation.
A focus of attention based system for text region localization has been proposed by Liu and Samarabandu in [8]. The intensity profiles and spatial variance is used to detect text regions in images. A Gaussian pyramid is created with the original image at different resolutions or scales. The text regions are detected in the highest resolution image and then in each successive lower resolution image in the pyramid.

The approach used in [9, 11] utilizes a support vector machine (SVM) classifier to segment text from non-text in an image or video frame. Initially text is detected in multi scale images using edge based techniques, morphological operations and projection profiles of the image [11]. These detected text regions are then verified using wavelet features and SVM. The algorithm is robust with respect to variance in color and size of font as well as language

III. APPROACH

The goal of the project is to implement, test, and compare and contrast two approaches for text region extraction in natural images, and to discover how the algorithms perform under variations of lighting, orientation, and scale transformations of the text. The algorithms are from Liu and Samarabandu in [1,2] and Gllavata, Ewerth and Freisleben in [3]. The comparison is based on the accuracy of the results obtained, and precision and recall rates. The technique used in [1,2] is an edge-based text extraction approach, and the technique used in [3] is a connected-component based approach. In order to test the robustness and performance of the approaches used, each algorithm was first implemented in the original proposed format. The algorithms were tested on the image data set provided by Xiaoqing Liu (xliu65@uwo.ca) and Jagath Samarabandu (jagath@uwo.ca), as well as another data set which consists of a combination of indoor and outdoor images taken from a digital camera. The results obtained were recorded based on criteria such as invariance with respect to lighting conditions, color, rotation, and distance from the camera (scale) as well as horizontal and/or vertical alignment of text in an image. The experiments have also been conducted for images containing different font styles and text characters belonging to language types other than English. Also, the precision and recall rates (Equations (1) and (2)), have been computed based on the number of correctly detected words in an image in order to further evaluate the efficiency and robustness of each algorithm.

The Precision rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false positives. *False positives* are those regions in the image which are actually not characters of a text, but have been detected by the algorithm as text regions.

$$\text{Precision Rate} = \frac{\text{Correctly detected words}}{\text{Correctly detected words} + \text{False positives}} \times 100\% \qquad (1$$

The Recall rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false negatives. *False Negatives* are those regions in the image which are actually text characters, but have not been detected by the algorithm.

$$\text{Recall Rate} = \frac{\text{Correctly detected words}}{\text{Correctly detected words} + \text{False Negatives}} \times 100\% \qquad (2)$$

1. ALGORITHM FOR EDGE BASED DETECTION[1,2]

The basic steps of the edge-based text extraction algorithm are given below, and diagrammed in Figure 1. The details are explained in the following sections.

i. Create a Gaussian pyramid by convolving the input image with a Gaussian kernel and successively down-sample each direction by half. (Levels: 4)
ii. Create directional kernels to detect edges at 0, 45, 90 and 135 orientations.
iii. Convolve each image in the Gaussian pyramid with each orientation filter.
iv. Combine the results of step 3 to create the Feature Map.
v. Dilate the resultant image using a sufficiently large structuring element (7x7 [1]) to cluster candidate text regions together.
vi. Create final output image with text in white pixels against a plain black background.

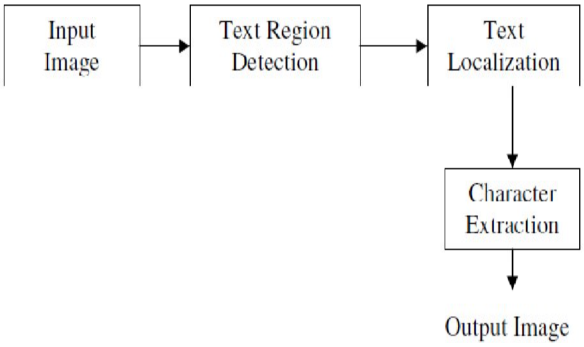
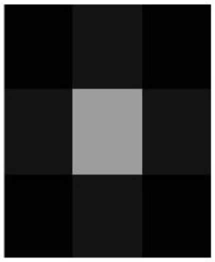


FIGURE 1 . BASIC BLOCK DIAGRAM FOR EDGE BASED TEXT EXTRACTION

As given in [1][2], the procedure for extracting a text region from an image can be broadly classified into three basic steps: (1)detection of the text region in the image, (2)localization of the region, and (3) creating the extracted output character image

1.1 DETECTION

Given an input image, the region with a possibility of text in the image is detected [1,2]. A Gaussian pyramid is created by successively filtering the input image with a Gaussian kernel of size 3x3 and down sampling the image in each direction by half. Down sampling refers to the process whereby an image is resized to a lower resolution from its original resolution. A Gaussian filter of size 3x3 will be used as shown in Figure 2. Each level in the pyramid corresponds to the input image at a different resolution. A sample Gaussian pyramid with 4 levels of resolution is shown in Figure 3. These images are next convolved with directional filters at different orientation kernels for edge detection in the horizontal (0°), vertical (90°) and diagonal (45°, 135°) directions. The kernels used are shown in Figure5.



0.0113	0.0838	0.0113
0.0838	0.6193	0.0838
0.0113	0.0838	0.0113

FIGURE 2.DEFAULT FILTER RETURNED BY THE FSPECIAL GAUSSIAN FUNCTION IN MATLAB.SIZE [3 3], SIGMA 0.5

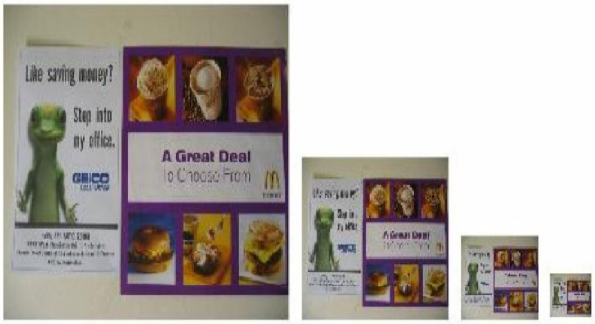


FIGURE 3. SAMPLE GAUSSIAN PYRAMID WITH 4 LEVELS

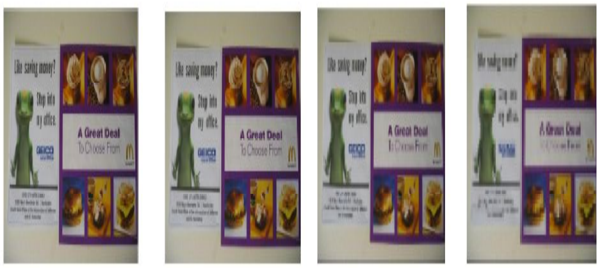


FIGURE 4. EACH RESOLUTION IMAGE RESIZED TO ORIGINAL IMAGE SIZE

After convolving the image with the orientation kernels, a feature map is created. A weighting factor is associated with each pixel to classify it as a candidate or non candidate for text region. A pixel is a candidate for text if it is highlighted in all of the edge maps created by the directional filters. Thus, the feature map is a combination of all edge maps at different scales and orientations with the highest weighted pixels present in the resultant map.

-1	-1	-1
2	2	2
-1	-1	-1

0° kernel

-1	-1	2
-1	2	-1
2	-1	-1

45° kernel

-1	2	-1
-1	2	-1
-1	2	-1

90° kernel

2	-1	-1
-1	2	-1
-1	-1	2

135° kernel

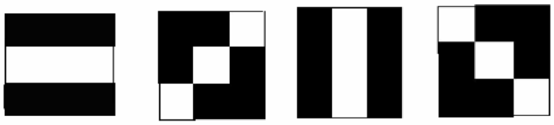


FIGURE 5. THE DIRECTIONAL KERNELS [1]

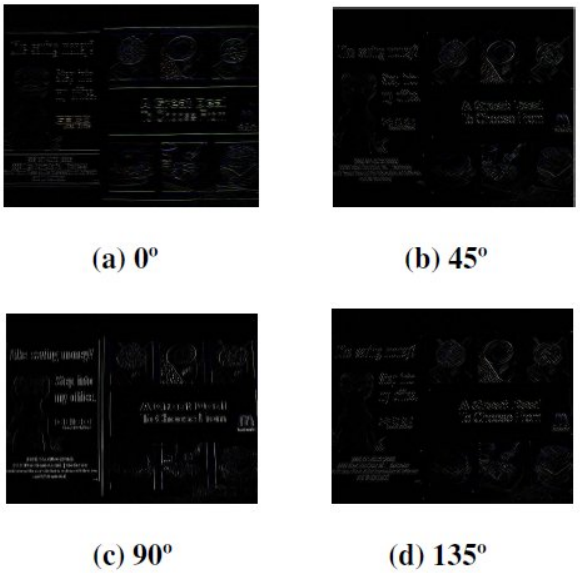


FIGURE 6. SAMPLE IMAGE FROM FIGURE 3 AFTER CONVOLUTION WITH EACH DIRECTIONAL KERNEL NOTE HOW THE EDGE INFORMATION IN EACH DIRECTION IS HIGHLIGHTED.

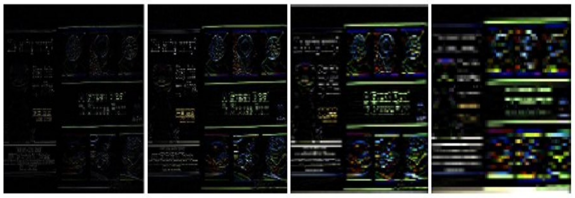


FIGURE 7. SAMPLE RESIZED IMAGE OF THE PYRAMID AFTER CONVOLUTION WITH 0° KERNEL

1.2. LOCALIZATION

The process of localization involves further enhancing the text regions by eliminating non-text regions [1,2]. One of the properties of text is that usually all characters appear close to each other in the image, thus forming a cluster. By using a morphological dilation operation, these possible text pixels can be clustered together, eliminating pixels that are far from the candidate text regions. *Dilation* is an operation which expands or enhances the region of interest, using a structural element of the required shape and/or size. The process of dilation is carried out using a very large structuring element in order to enhance the regions which lie close to each other. In this algorithm, a structuring element of size [7x7] has been used [1]. Figure 8 below shows the result before and after dilation.

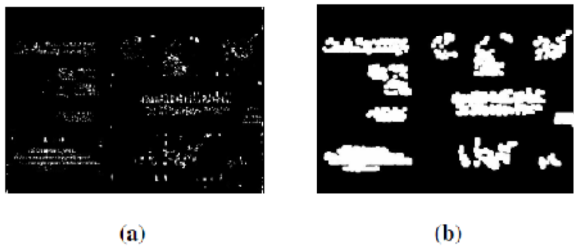


FIGURE 8 .(A) BEFORE DILATION (B) AFTER DILATION

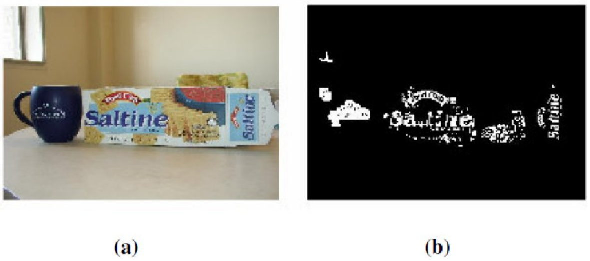


FIGURE 9. (A) ORIGINAL IMAGE [1,2] (B) RESULT

1.3. ALGORITHM FOR CONNECTED COMPONENT BASED TEXT REGION EXTRACTION [3]

The basic steps of the connected-component text extraction algorithm are given below, and diagrammed in Figure 10. The details are discussed in the following sections.

1. Convert the input image to YUV color space. The luminance(Y) value is used for further processing. The output is a gray image.
2. Convert the gray image to an edge image.
3. Compute the horizontal and vertical projection profiles of candidate text regions using a histogram with an appropriate threshold value.
4. Use geometric properties of text such as width to height ratio of characters to eliminate possible non-text regions.
5. Binarize the edge image enhancing only the text regions against a plain black background.
6. Create the Gap Image (as explained in the next section) using the gap-filling process and use this as a reference to further eliminate non-text regions from the output.

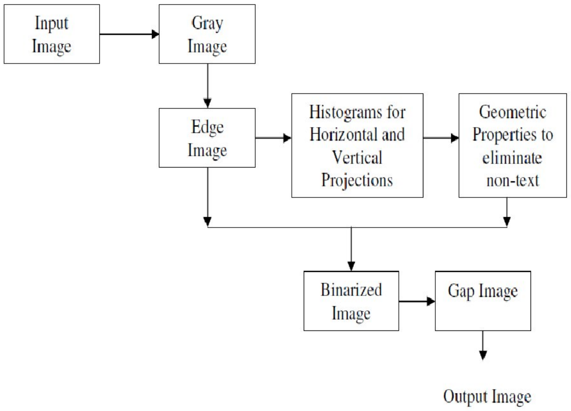

Input Image
Gray Image
Edge Image
Histograms for Horizontal and Vertical Projections
Geometric Properties to eliminate non-text
Binarized Image
Gap Image
Output Image

FIGURE 10. BASIC BLOCK DIAGRAM FOR CONNECTED COMPONENT BASED TEXT EXTRACTION.

1.4. DETECTION OF EDGES

This section corresponds to Step 2 of 3.2. In this process, the connected-component based approach is used to make possible text regions stand out as compared to non-text regions. Every pixel in the edge image is assigned a weight with respect to its neighbors in each direction. As depicted in Figure 12, this weight value is the maximum value between the pixel and its neighbors in the left (L), upper (U) and upper-right (UR) directions [3]. The algorithm proposed in [3] uses these three neighbor values to detect edges in horizontal, vertical and diagonal directions. The resultant edge image obtained is sharpened in order to increase contrast between the detected edges and its background, making it easier to extract text regions. Figure 13 below shows the sharpened edge image for the Y Channel gray image G from Figure 11, obtained by the algorithm proposed in [3].

The algorithm for computing the edge image E, as proposed in [3] is as follows:

```
1. Assign left, upper, upperRight to 0.
2. For all the pixels in the gray image G(x,y) do
a. left = (G(x,y) – G(x-1,y))
b. upper = (G(x,y) – G(x,y-1))
c. upperRight = (G(x,y)-G(x+1,y-1))
d. E(x,y) = max( left, upper, upperRight )
3. Sharpen the image E by convolving it with a sharpening filter.
```

$$W(x,y) = \max(L,U,UR)$$

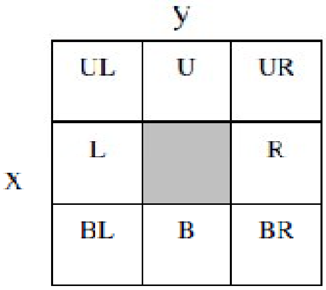


FIGURE 12. WEIGHT FOR PIXEL (X,Y)

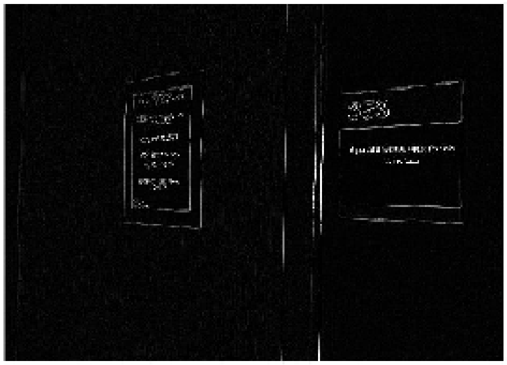


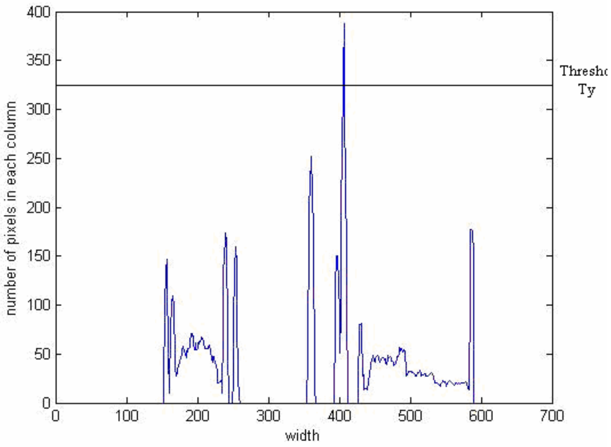
FIGURE 13. SHARPENED EDGE IMAGE

1.5. LOCALIZATION

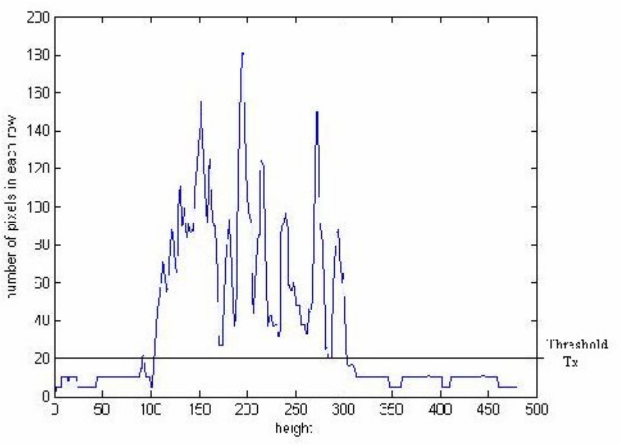
In this step, the horizontal and vertical projection profiles for the candidate text regions are analyzed. The sharpened edge image is considered as the input intensity image for computing the projection profiles, with white candidate text regions against a black background. The vertical projection profile shows the sum of pixels present in each column of the intensity or the sharpened image. Similarly, the horizontal projection profile shows the sum of pixels present in each row of the intensity image. These projection profiles are essentially histograms where each bin is a count of the total number of pixels present in each row or column. The vertical and horizontal projection profiles for the sharpened edge image from Figure 13, are shown in Figure14 (a) and (b) respectively.

Candidate text regions are segmented based on adaptive threshold values, Ty and Tx, calculated for the vertical and horizontal projections respectively. Only regions that fall within the threshold limits are considered as candidates for text. The value of threshold Ty is selected to eliminate possible non text regions such as doors, window edges etc. that have a strong vertical orientation. Similarly, the value of threshold Tx is selected to eliminates regions which might

benon text or long edges in the horizontal orientation.



(A)



(B)

FIGURE14. (A) VERTICAL PROJECTION PROFILE (B) HORIZONTAL PROJECTION PROFILE FOR SHARPENED IMAGE IN FIGURE 13.

$$Tx = \frac{\text{Mean(Horizontal projection profile)}}{20} \qquad (3)$$

$$Ty = \text{Mean(Vertical projection profile)} - \frac{\text{Max (Vertical projection profile)}}{10} \qquad (4)$$

1.6. ENHANCEMENT AND GAP FILLING

The geometric ratio between the width and the height of the text characters is considered to eliminate possible non-text regions. This ratio value will be defined after experimenting on different kinds of images to get an average value. In this project, regions with minor to major axis ratio less than 10 are considered as candidate text regions for further processing. Next a gap image will be created which will be used as a reference to refine the localization of the detected text regions [3]. If a pixel in the binary edge image created is surrounded by black (background) pixels in the vertical, horizontal and diagonal directions, this pixel is also substituted with the background value. This process is known as *gap filling.* An example of extracted text using this technique is shown in Figure 15.

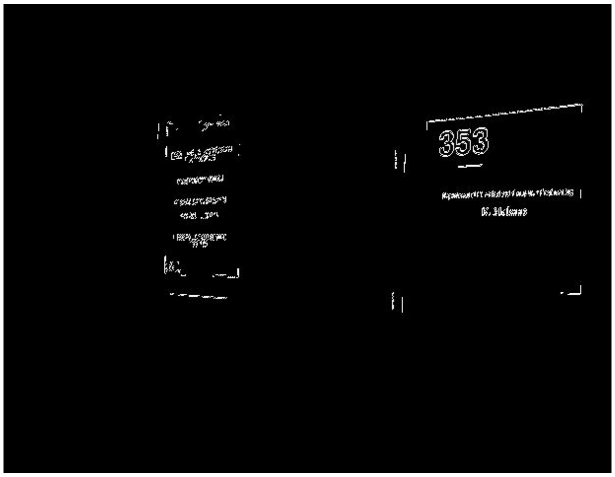


Figure 15. Result Obtained By Connected Component Based Text Detection Algorithm for test image in Figure 11 (1)

IV. Conclusion and Recommendations

The results obtained by each algorithm on a varied set of images were compared with respect to precision and recall rates. In terms of scale variance, the connected component algorithm is more robust as compared to the edge based algorithm for text region extraction. In terms of lighting variance also, the connected component based algorithm is more robust than the edge based algorithm. In terms of rotation or orientation variance, the precision rate obtained by the connected component based algorithm is higher than the edge based, and the recall rate obtained by the edge based is higher than the connected component based The average precision rates obtained by each algorithm for the remaining test images are similar, whereas the average recall rate obtained by the connected component algorithm is a little lower than the edge based algorithm. Thus, the results from the experiments indicate that in most of the cases, the connected component based algorithm is more robust and invariant to scale, lighting and orientation as compared to the edge based algorithm for text region extraction.

V. References

[1] Xiaoqing Liu and Jagath Samarabandu, *An Edge-based text region extraction algorithm for Indoor mobile robot navigation*, Proceedings of the IEEE, July 2005.

[2] Xiaoqing Liu and Jagath Samarabandu, *Multiscale edge-based Text extraction from Complex images*, IEEE, 2006.

[3] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, *A Robust algorithm for Text detection in images*, Proceedings of the 3rd international symposium on Image and Signal Processing and Analysis, 2003.

[4] Keechul Jung, Kwang In Kim and Anil K. Jain, *Text information extraction in images*
and video: a survey, The journal of the Pattern Recognition society, 2004.

[5] Kongqiao Wang and Jari A. Kangas, *Character location in scene images from digital*
camera, The journal of the Pattern Recognition society, March 2003.

[6] K.C. Kim, H.R. Byun, Y.J. Song, Y.W. Choi, S.Y. Chi, K.K. Kim and Y.K Chung,*Scene Text Extraction in Natural Scene Images using Hierarchical Feature Combining and verification*, Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04), IEEE.

[7] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, *TextFinder: An Automatic System to Detect and Recognize Text in Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, November 1999.

[8] Xiaoqing Liu and Jagath Samarabandu, *A Simple and Fast Text Localization Algorithm for Indoor Mobile Robot Navigation*, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5672, 2005.

[9] Qixiang Ye, Qingming Huang, Wen Gao and Debin Zhao, *Fast and Robust text detection in images and video frames*, Image and Vision Computing 23, 2005.

[10] Rainer Lienhart and Axel Wernicke, *Localizing and Segmenting Text in Images and Videos*, IEEE Transactions on Circuits and Systems for Video Technology, Vol.12, No.4, April 2002.

[11] Qixiang Ye, Wen Gao, Weiqiang Wang and Wei Zeng, *A Robust Text Detection Algorithm in Images and Video Frames*, IEEE, 2003.

[12]
http://softpixel.com/~cwright/programming/colorspace/yuv/

[13] http://www.captcha.net

[14] http://images.google.com

STATISTICAL ANALYSIS OF EMOTION DETECTION USING FUNDAMENTAL FREQUENCY

A.Vasavi[1] *Dr.D.Satyanaraya*[2] *R. Hanuma naik*[3]

[1]*PG Student RGMCET, Nandyal* vasaviakula@gmail.com

[2]*Professor Department of ECE, RGMCET, Nandyal* dsn2003@rediffmail.com

[3]*Assistant Professor Dept of EIE, RGMCET, Nandyal* rhnaik.1717@gmail.com

Abstract—During expressive speech, the voice is enriched to convey not only the intended semantic message but also the emotional state of the speaker. The pitch contour is one of the important properties of speech that is affected by this emotional modulation.. This paper presents an analysis of the statistics derived from the pitch contour. First, pitch features derived from emotional speech samples are compared with the ones derived from neutral speech, by using symmetric Kullback-Leibler distance. Then, the emotionally discriminative power of the pitch features is quantified by comparing nested logistic regression models. The results indicate that gross pitch contour statistics such as mean, maximum, minimum, and range are more emotionally prominent than features describing the pitch shape. Also, analyzing the pitch statistics at the utterance level is found to be more accurate and robust than analyzing the pitch statistics for shorter speech regions (e.g., voiced segments). Finally, the best features are selected to build a binary emotion detection system for distinguishing between emotional versus neutral speech. A new two-step approach is proposed. In the first step, reference models for the pitch features are trained with neutral speech, and the input features are contrasted with the neutral model. In the second step, a fitness measure is used to assess whether the input speech is similar to, in the case of neutral speech, or different from, in the case of emotional speech, the reference models. The proposed approach is tested with four acted emotional databases spanning different emotional categories, recording settings, speakers and languages.

Index Terms—Emotional speech analysis, emotional speech recognition, expressive speech, intonation, pitch contour analysis.

I. INTRODUCTION

EMOTION plays a crucial role in day-to-day interpersonal human interactions. Recent findings have suggested that emotion is integral to our rational and intelligent decisions. It helps us to relate with each other by expressing our feelings and providing feedback. This important aspect of human interaction needs to be considered in the design of human-machine interfaces (HMIs).

Speech prosody is one of the important communicative channels that is influenced by and enriched with emotional modulation.

The goal of this paper is two fold. The first is to study which aspects of the pitch contour are manipulated during expressive speech (e.g., curvature, contour, shape, dynamics). For this purpose, we present a novel framework based on Kullback-Leibler divergence (KLD) and logistic regression models to identify, quantify, and rank the most emotionally salient aspects of the FO contour. First, the symmetric Kullback-Leibler distance is used to compare the distributions of different pitch statistics (e.g., mean, maximum) between emotional speech and reference neutral speech. Then, a logistic regression analysis is implemented to discriminate emotional speech from neutral speech using the pitch statistics as input. These experiments provide insights about the aspects of pitch that are modulated to convey emotional goals. The second goal is to use these emotionally salient features to build robust prosody speech models to detect emotional speech Gaussian mixture models (GMMs) are trained using the most discriminative aspects of the pitch contour, following the analysis results presented in this paper.

II. METHODOLOGY

A. **Overview**

The fundamental frequency or FO contour (pitch), which is a prosodic feature, provides the tonal and rhythmic properties of the speech.

The fundamental frequency is also a supra-segmental speech feature, where information is conveyed over longer time scales than other segmental speech correlates such as spectral envelope features. Therefore, rather than using the pitch value itself, it is commonly accepted to estimate global statistics of the pitch contour over an entire utterance or sentence (sentence-level) such as the mean, maximum, and standard deviation.

B. Databases

In this paper, five databases are considered: one non-emotional corpus used as a neutral speech reference, and four acted emotional databases with different properties.

For the analysis and the training of the models (Sections IV-VI), three emotional corpora were considered. These emotional databases were chosen to span different emotional categories, speakers, genders, and even languages, with the purpose to include, to some extent, the variability found in the pitch. The first database was collected at the University of Southern California (USC) using an electromagnetic artic-ulography (EMA) system. In this database, which will be referred to here on as EMA, one male and two female subjects (two of them with formal theatrical vocal training) read ten sentences five times portraying the emotions sadness, anger, and happiness, in addition to neutral state. Although this database contains articulatory information, only the acoustic signals are analyzed in this study

The second emotional corpus corresponds to the Emotional Prosody Speech and Transcripts database (EPSAT). This database was collected at the University of Pennsylvania and is comprised of recordings from eight professional actors (five female and three male) who were asked to read short semantically neutral utterances corresponding to dates and numbers, expressing 14 emotional categories in addition to the neutral state

The third emotional corpus is the Database of German Emotional Speech (GES) which was collected at the Technical University of Berlin . This database was recorded from ten participants, five female, and five male, who were selected based on the naturalness and the emotional quality of the participant's performance in audition sessions. The emotional categories considered in the database are anger, happiness, sadness, boredom, disgust, and fear, in addition to neutral state.

C. Speaker Dependent Normalization

Normalization is a critical step in emotion recognition. The goal is to eliminate speaker and recording variability while keeping the emotional discrimination. For this analysis, a two-step approach is proposed: 1) energy normalization and 2) pitch normalization.

In the first step, the speech files are scaled such that the average RMS energy of the neutral reference database (£$_r$ef) ^{an}d the neutral subset in the emotional databases (£^$_{eu}$) are the same for each speaker s. This normalization is separately applied for each subject in each database. The goal of this normalization is to compensate for different recording settings among the databases.

$$S^{s}_{Energy} = \sqrt{\frac{E_{ref}}{E^{s}_{neu}}} \quad ....(1)$$

In the second step, the pitch contour is normalized for each subject (speaker-dependent normalization). The average pitch across speakers in the neutral reference database is estimated F0$_{re}$f. Then, the average pitch value for the neutral set of the emotional databases is estimated for each speaker F0$_{neu}$. Finally, a scaling factor (Sp$_0$) is estimated by taking the ratio between –F0$_{ref}$ and F0$_{neu}$.Therefore, the neutral samples of each speaker in the databases will have a similar FO mean value.

$$s_{F0} = F0_{ref} \,/\, F0_{neutral} \ ...(2)$$

One assumption made in this two-step approach is that neutral speech will be available for each speaker. For real-life applications, this assumption is reasonable when either the speakers are known or a few seconds of their neutral speech can be prerecorded.

D. Pitch Features

The pitch contour was extracted with the Praat speech processing software, using an autocorrelation method. The analysis window was set to 40 ms with an overlap of 30 ms, producing 100 frames per second. The pitch was smoothed to remove any spurious spikes by using the corresponding option provided by the Praat software.

Describing the pitch shape for emotional modulation analysis is a challenging problem, and different approaches have been proposed. The Tones and Break Indices System (ToBI) is a well-known technique to transcribe prosody (or intonation). Although progress has been made toward automatic ToBI transcription [30], an accurate and more complete prosodic transcription requires hand labeling. Furthermore, linguistic models of intonation may not be the most appropriate labels to describe the emotions . Taylor has proposed an alternative pitch contour parameterization called Tilt Intonation Mode! . In this approach, the pitch contour needs to be pre-segmented into intonation events. However, there is no straightforward or readily available system to estimate these segments. Given these limitations, we follow a similar approach presented by Grabe et a!. . The voiced regions, which are automatically segmented from the pitch values, are parameterized using polynomials. This parameterization captures the local shape of the FO contour with few parameters, which provides clear physical interpretation of the curves. Here, the slope (tti), curvature, and inflexion (c3) are estimated to capture the local shape of the pitch contour by fitting a first-, second-, and third-order polynomial to each voiced region segment

$$y = a1.x + a0 \text{..........................(3)}$$

$$y = b2.x^2 + b1.x + b0 \text{..............(4)}$$

$$y = c3.x^3 + c2.x^2 + c1.x + c0 \text{...(5)}$$

These statistics provide insights about the local dynamics of the pitch contour. For example, while the pitch range at the sentence-level (Srange) gives the extreme value distance of the pitch contour over the entire sentence, SVmeanRange, the mean of the range of the voiced regions, will indicate whether the voiced regions have flat or inflected shape.

III. EXPERIMENT I: COMPARISONS USING SYMMETRIC KULLBACK-LEIBLER DISTANCE

This section presents our approach to identifying and quantifying the pitch features with higher levels of emotional modulation. Instead of comparing just the mean, the distributions of the pitch features extracted from the emotional databases are compared with the distributions of the pitch features extracted from the neutral reference corpus using KLD . KLD provides a measure of the distance between two distributions. It is an appealing approach to robustly estimate the differences between the distributions of two random variables.

Since the KLD is not a symmetric metric, we propose the use of the symmetric Kullback-Leibler distance or ^-divergence, which is defined as

$$J_{(q,p)} = D(q//p) + D(p//q)/2 \text{ .... (6)}$$

Where D(p //q) is the conventional KLD

$$D(q//p) = \sum_{x \in X} q(X)\log(q(x)/p(x)) \text{ .. (7)}$$

The first step is to estimate the distribution of the pitch features for each database, including the neutral reference corpus. For this purpose, we proposed the use of the K-means clustering algorithm to estimate the bins. This nonparametric approach was preferred since the KLD is sensitive to the bins' estimation. To compare the symmetric KLD in terms of features and emotional categories k the number of bins, was set constant for each distribution (k = 40 empirically chosen). Notice that these feature-dependent nonuniform bins were estimated considering all the databases to include the entire range spanned by the features. After the bins were calculated, the distribution ($p_f^{(d,e)}$)of each pitch feature (f) was estimated for each database (d), and for each emotional category (e). Therefore, the true feature distribution for each subset is approximated by counting the number of samples assigned to each bin. The same procedure was used to estimate the distribution of the pitch features in the reference neutral corpus, q_f^{ref}.

The next step is to compute the symmetric KLD between the distribution of the emotional databases and the distribution estimated from the reference database $J_f^{(d,e)}$ ($p_f^{(d,e)}$, $q_f^{(ref)}$) . This procedure is repeated for each database and for each emotional category.

A good pitch feature for emotion discrimination ideally would have $J_f^{(d,neytral)}$ close to zero (neutral speech of the database d is similar to the reference corpus) and a high value for $J_f^{(d,e)}$,where e is any emotional category except the neutral state. Notice that if $J_f^{(d,neytral)}$ and $J_f^{(d,e)}$, have high values, this test would indicate that the speech from the emotional database is different from the reference database (how neutral is the neutral speech?). Likewise, if both values were similar, this feature would not be relevant for emotion discrimination. Therefore, instead of directly comparing the symmetric KLD, we propose to estimate the ratio between . $J_f^{(d,e)}$,and $J_f^{(d,neytral)}$ .That is, after matching the feature distributions with the reference feature distributions, the emotional speech is directly compared with the neutral set of the same emotional database by taking the ratio. High values of this ratio will indicate that the pitch features for emotional speech are different from their neutral counterparts, and therefore are relevant to discriminate emotional speech from neutral speech.

$$r_f^{(d,\bar{e})} = J_f^{(d,\bar{e})} / J_f^{(d,neutral)} \quad \ldots\ldots..(8)$$

The pitch features with higher values are SVmeanMin, SVmeanMax, Sdiqr, and Smean for the sentence-level features and Vrange, Vstd, Vdrange, and Vdiqr for the voiced-level features.

IV. EXPERIMENT 2: LOGISTIC REGRESSION ANALYSIS

Logistic regression is a well-known technique to model binary or dichotomous variables. In this technique, the conditional expectation of the variable given the input variables is modeled with the specific form described in (9). After applying the logit transformation (10), the regression problem becomes linear in its parameters (β)A nice property of this technique is that the significance of the coefficients can be measured using the log-likelihood ratio test between two nested models (the input variables of one model are included in the other model). This procedure provides estimates about the discriminative power of each input feature

$$E(Y / f1, f2.....fn) = \pi(x) = \frac{e^{\beta 0+\beta 1x1+....+\beta nxn}}{1+e^{\beta 0+\beta 1x1+....+\beta nxn}} \ldots..(9)$$

$$g(x) = \ln[\pi(x)/1-\pi(x)]$$

$$=^{\beta 0+\beta 1x1+....+\beta nxn} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(10)$$

Logistic regression analysis is used witb/orward feature selection (FFS) to discriminate between each emotional category and neutral state (i.e., neutral-anger).

VI. EMOTIONAL DISCRIMINATION RESULTS USING NEUTRAL MODELS

to recognize expressive speech using the acoustic likelihood scores obtained from hidden Markov models (HMMs) [6]. The models were trained with neutral (non-emotional) speech using spectral features. In this section, the ideas are extended to build neutral models for the selected sentence-and voiced-level pitch features .

A. **Motivation and Proponed Approach**

Automatic emotion recognition in real-life applications is a nontrivial problem due to the inherent inter-speaker variability of expressive speech. Furthermore, the emotional descriptors are not clearly established. The feature selection and the models are trained for specific databases with the risk of sparseness in the feature space and over-fitting. It is also fairly difficult, if not infeasible, to collect enough emotional speech data so that one can train robust and universal acoustic models of individual emotions. Therefore, it is not surprising that the models built with these individual databases (usually offline) do not easily generalize to different databases or online recognition tasks in which blending of emotions is observed .

In the first step, neutral models are built to measure the degree of similarity between the input speech and the reference neutral speech. The output of this block is & fitness measure of the input speech. In the second step, these measures are used as features to infer whether the input speech is emotional or neutral. If the features from the expressive speech differ in any aspect from their neutral counterparts, the fitness measure will decrease. Therefore, we hypothesize that setting thresholds over these fitness measures is easier and more robust than setting thresholds over the features themselves.

FO contour is assumed to be largely independent of the specific lexical content, in contrast to spectral speech features. Therefore, a single lexical-independent model is adequate to model the selected pitch features. For this task, we propose the use of univariate GMM for each pitch feature.

The maximum likelihood estimates of the parameters in the GMM 8 are computed using the expectation-maximization (EM) algorithm. For a given input speech, the likelihoods of the models, Ff(Xt — x/|©), are used as fitness measures. In the second step, a Linear Discriminant Classifier (LDC) was implemented to discriminate between neutral and expressive speech. For a given input speech, the likelihoods of the models, Ff(Xt — x/|©), are used as fitness measures. In the second step, a Linear Discriminant Classifier (LDC) was implemented to discriminate between neutral and expressive speech.

$$F_f(X_f = x_f \,|\Theta) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sigma_j \sqrt{2\pi}} \exp(\frac{(X_f - \mu_j)^2}{-2\sigma_j^2})\ldots(12)$$

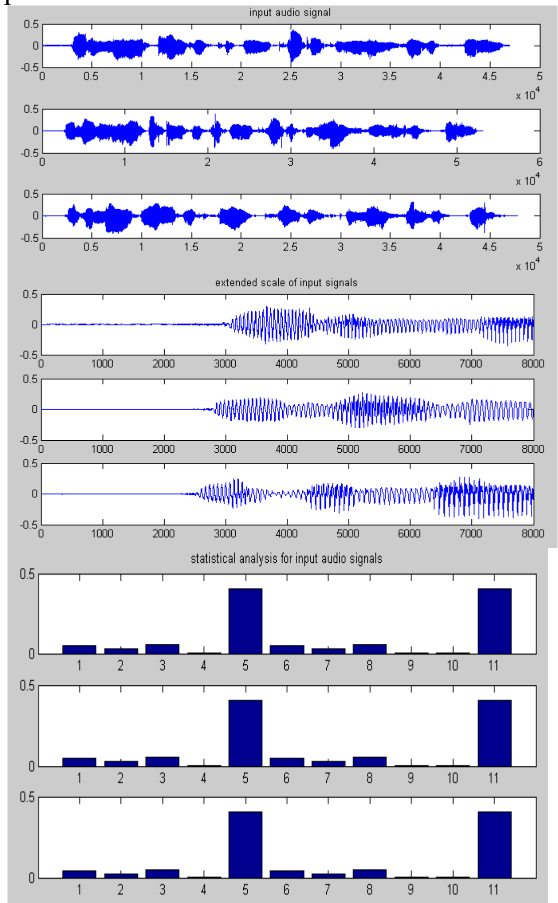
with

$$\Theta = \left\{\alpha_j, \mu_j, \sigma_j\right\}_{j=1}^{K}, \alpha_j > 0\, j = 1,.....K, \sum_{j=1}^{K} \alpha_j = 1$$

B. Results

The recognition results presented in this section are the average values over 400 realizations. Since the emotional categories are grouped together, the number of emotional samples is higher than the neutral samples.

An important parameter of the GMM is the number of mixtures, the performance of the GMM-based pitch neutrai models for different numbers of mixtures. The figure shows that the proposed approach is not sensitive to this parameter.



C.Conclusion

This paper presented an analysis of different expressive pitch contour statistics with the goal of finding the emotionally salient aspects of the FO contour (pitch). For this purpose, two experiments were proposed.

In the first experiment, the distribution of different pitch features was compared with the distribution of the features derived from neutral speech using the symmetric KLD. In the second experiment, the emotional discriminative power of the pitch features was quantified within a logistic regression framework. Both experiments indicate that dynamic statistics such as mean, maximum, minimum, and range of the pitch are the most salient aspects of expressive pitch contour. The statistics were computed at sentence and voiced region levels. The results indicate that the system based on sentence-level features outperforms the one with voiced-level statistics both in accuracy and robustness, which facilitates a turn-by-turn processing in emotion detection.

REFERENCES

(I J R. W. Picard, "Affective Computing," MIT Media Laboratory Perceptual Computing Section, Cambridge, MA, USA, Tech. Rep. 321, Nov. 1995.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias. W. Fellcnz, and J. Taylor, "Emoiion recognition in human-computer interaction." IEEE Signal Process. Mag., vol. 18, no. 1- pp. 32-80, Jan. 2001.

|3| A. Alvarez, I. Cearreta, J. Lopez, A. Arruti, E. Lazkano, B. Sierra, and N. Garay, "Feature subset selection based on evolutionary algorithm.[1], for automatic emotion recognition in spoken Spanish and standard basque language." in Proc. 9th Int. Conf. Text, Speech and Dialogue (TSD 2006), Brno, Czech Republic, Sep. 2006. pp. 565-572.

[4] D. Vervcridis and C. Kotropoulos, "Fast sequential floating forward selection applied lo emotional .speech features estimated on DES and S US AS data collect ions," in Prot.. XIV Ear. Signal Process. Conf. (EU-SiPCO'Of,), Florence, Italy, Sep. 2006, pp. 929-932.

[5] M. Sedaaghi, C. Kotropoulos, and D. Ververidis, "Using adaptive genetic algorithms to improve speech emotion recognition." in Proc. Int. Workshop Multimedia Signal Process. (MMSP '07). Chania, Crete. Greece, Oct. 2007, pp. 461^164.

[6] C. Busso, S. Lee, and S. Narayanan, "Using neutral speech models for emotional speech analysis," in Proc. Interspeech'07—Eurospeech. Antwerp, Belgium, Aug. 2007,

Estimation of Motion Vector Parameters Using Modified Diamond Search

T.Swati (M.Tech), Sri M.V. R Vittal, Senior Assistant Professor, M.Tech., (Ph.d)
Electronics and Communication Department,
GPREC (Sri Krishna
Devaraya University)
Kurnool, Andhra Pradesh.
Swati.thimmapuram@gmail.com

Abstract

A fast block motion estimation algorithm is proposed using modified diamond search patterns. This algorithm utilizes the directions and magnitudes of motion vectors between interblocks and uses a smaller
number of search points than conventional diamond search patterns. Simulation results show that the proposed method significantly improves computational speed over other fast motion estimation algorithms without degradation of distortion.

I. INTRODUCTION

Many video coding standards use block matching motion estimation (BMME) to reduce temporal redundancy between successive frames. However, the full search BMME requires a heavy computational burden. To speed up the motion estimation process, many fast block matching algorithms (BMAs) have been developed using different search patterns, such as square-shape, diamond-shape, and hexagon-shape by exploiting the motion vector distribution characteristics. All fast BMAs are based on the assumption that the distortion of the block matching increases monotonically away from the global minimum distortion. The new three-step search (NTSS) employs a halfway-stop technique of the three-step search (TSS) using sparse square-shaped patterns, and results in accelerated motion estimation in stationary blocks. The diamond search (DS) algorithm and the crossdiamond
search (CDS) algorithm utilize centre-biased motion vector distribution (MVD) characteristics in real-life sequences. DS and CDS algorithms are effective in cases of small motion, since their search patterns are based on the centre-biased MVD with a radius of 2 pels and use fewer search points with similar distortion performance compared to NTSS and TSS. In this Letter, we propose a faster and less distorted BMA which utilizes modified diamond patterns based on the motion correlation between neighbour blocks. Modified diamond search patterns: Most blocks in real-life image sequences have highly correlated motion characteristics between spatial and=or temporal neighbour blocks, i.e. the current motion vector can be predicted from the neighbour blocks in a spatial and=or temporal sense. Hsieh and Lu introduced a halfway-stop technique that thresholds the motion compensation error with priority order between neighbour blocks. However, this motion estimation technique may not find the global minimum distortion or sudden changes of motion vectors along object boundaries, because the search regions are limited around the point predicted by motion vectors of neighbour blocks. In this Letter we propose the use of modified diamond search

Diamond Search (DS): This is also one of the algorithms that are very popular for motion estimation and in fact it is used for motion estimation by MPEG-Tool. The basic idea behind this algorithm is to divide the search window into a number of regions and do a full search only in one of these regions. It may be described as:

Step 1: An initial step size is picked. Thirteen search points are selected surrounding a selected midpoint, in the pattern of a diamond, with the selected step size.

Step 2: A search is performed in the selected points and the point with minimum mod value is picked. Making this point as centre another nine points are selected in the diamond pattern.

Step 3: A search is performed in the selected points and the point with minimum mod value is picked. If the point with minimum mod value is the midpoint then a diamond pattern of five search points is selected making the minimum mod value point as the centre.

Step 4: If the selected midpoint in the step 3 is not the midpoint then another diamond pattern with nine search points is selected around this point. This procedure is repeated until the minimum midpoint occurred is the midpoint.

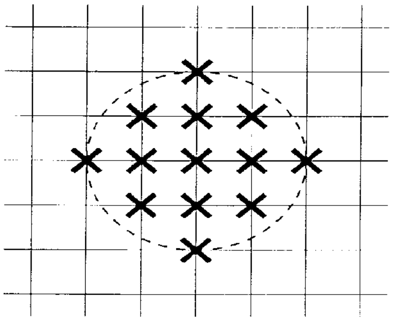


Fig. a Thirteen point pattern for diamond search

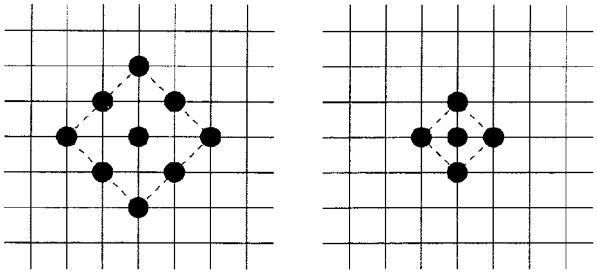


Fig. b Nine and five point pattern for diamond search

In the modified diamond search pattern with fewer search points by adapting interblock motion correlations. Three neighbour blocks, the left block and the upper block of a current block in a current frame and the block of the same pixel position in the previous frame, are used without the priority order between neighbour blocks to increase the likelihood of finding the global minimum distortion and sudden changes of motion vectors. As shown in Fig. 1, different search patterns are employed according to the motion vectors of neighbour blocks. These patterns are based on the MVD found through experimentations; most image sequences have centre-biased motion vectors located in the central 5_5 area, which account for more than 80% of all motion vectors . If all the motion vectors in the neighbour blocks have a distance of less than 3 pel (p < 3), the proposed method uses the small sparse diamond (SSD) pattern (Fig. 1a). Otherwise, an initial search pattern is determined by motion directions predicted from those of neighbour blocks. If the neighbour blocks have the same motion directions within 90_, one of the quarter sparse diamond (QSD) patterns, QSD1, QSD2, QSD3, and QSD4, is selected based on their motion directions (Fig. 1b). If they are within 180_, one of the half sparse diamond (HSD) patterns (Fig. 1c), HSD1, HSD2, HSD3, and HSD4, is selected as the initial search pattern. In all other cases, HSD1 is used. Note that these QSD or HSD patterns are used for only the initial search pattern.

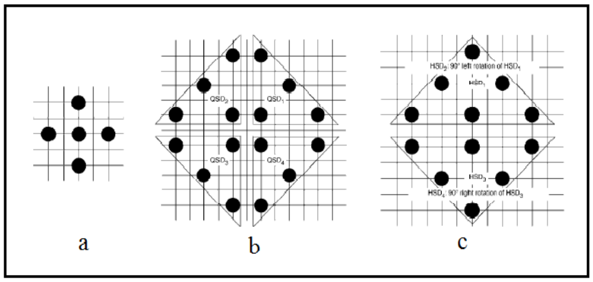

QSD₁
QSD₂
QSD₃
QSD₄
HSD₂: 90° left rotation of HSD₁
HSD₁
HSD₃
HSD₄: 90° right rotation of HSD₃
a
b
c

Fig. 1 Modified diamond search patterns exploiting motion correlation
a SSD pattern used as initial search pattern when p<3
b QSD patterns when p_3 and motion directions within 90_
c HSD patterns when p_3 and motion directions within 180_, and in all other
cases

Modified diamond search (MDS) algorithm: The MDS patterns are based on the motion correlation characteristics of real-life image sequences. The MDS algorithm can be used for a faster BMA because they use fewer search points than DS and CDS patterns. The following is a summary of the MDS algorithm.

Step 1: If at least one of the displacements of the neighbour blocks is larger than 3, select one of the QSD or HSD patterns as the initial search pattern, depending on the directions of the neighbour motion vectors. Otherwise, set the centre point of the SSD pattern as the motion vector of (0, 0) and go to step 2. The point of the minimum block
distortion measure (BDM) is found from the points in the initial search pattern, and this point is set as the centre point of the SSD pattern. Goto step 2.

Step 2: Find the minimum distortion point using the centre point of the SSD pattern determined in step 1. If it is at the centre of the SSD pattern, go to step 3. Otherwise, repeat step 2 using the SSD pattern centred at the newly found minimum distortion point within the search range.

Step 3: Find the sets of two points with the smallest BDM from the outside points of the SSD pattern (denoted as or in Figs. 2a and b). Those two points and the centre point of the SSD pattern define the boundaries of the candidates of final motion vectors. Identify a new BDM point from among these candidates, which is the final motion vector.

Fig. 2 shows two examples of the motion estimation steps when the initial search patterns are different and the final motion vectors are the same as (0, 0). The circle, represents the minimum BDM point found in step 3. Note that the number of search points (NSP) in Figs. 2a and b is 8 and 13, respectively.

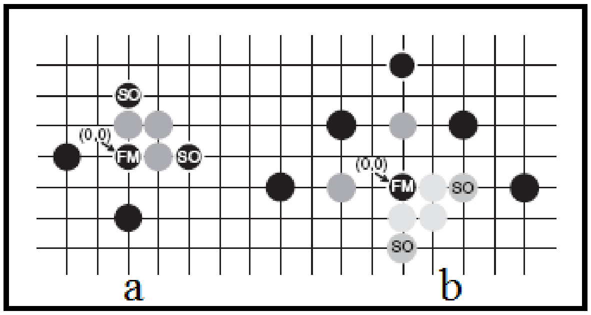
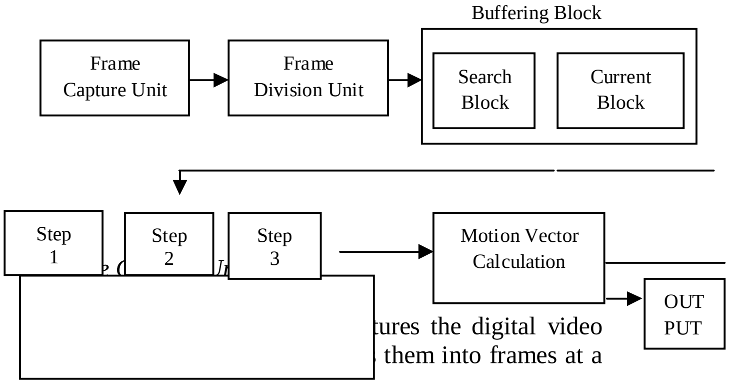


Fig. 2 Examples of proposed motion estimation algorithm
a When p < 3 in neighbour blocks
b When p < 3 in neighbour blocks and motion directions within 180_

BLOCK DIAGRAM



tures the digital video
them into frames at a

frequency of 30frames/second. The frame capture unit supplies the captured frames to the frame division unit.

Frame Division Unit

Frame division unit collects the frame from the frame capture unit and divides them into blocks. Block motion estimation technique picks up a block from the current frame and compares it with the block in the previous frame. The frame division unit divides the current frame i.e. the frame to be estimated, into 16x16 blocks. The previous frame, which is taken as the reference, is divided into 32x32 blocks.

Buffering Block

The block from the current frame which is to be compared is placed into the current block and the block with which it is to be compared is placed into the search block. The address positions of the current and search blocks are always updated. When the first comparison is completed the position of the current block is to be moved by a length of 8pixels in all the possible ways. The position of the search block remains constant for the first complete comparison. A complete comparison is completed when the 16x16 current block is compared with the 32x32 search frame in all the possible ways as shown in the following figure 3.2.

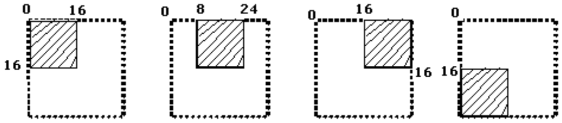


Fig. 3. Current block on search block in all possible positions

Algorithmic Block

Modified diamond search block consists of three independent blocks each indicating one individual step in the diamond search algorithm. The first step deals with the thirteen point diamond search and calculation of the block difference at that particular position. The second step is the nine point diamond search which is done making the minimum mod value in the thirteen point search as the mid point and forming the nine point diamond around it. In the case of second search if the minimum mod value point is not the centre, then the nine point search is repeated, with the obtained minimum mod value point as the centre, until the point is obtained at the centre. The third step is performed with five search points and the centre being the centre of the nine point search.

Motion Vector Block

The 64x2 motion vectors are calculated from the obtained block difference and these are used to estimate the motion between the frames and thus to estimate the second frame using the first frame and the difference.

Experimental results: The proposed MDS algorithm is simulated and the results were compared to those of FS, TSS, DS, CDS algorithms in terms of NSP, mean absolute difference (MAD) as the BDM, and the product of NSP and MAD (Table 1). All simulations were carried out with QCIF or CIF MPEG image sequences, block sizes of 16, and search window sizes of _8. When there were small motions in the image sequences, such as those in the QCIF 'Miss America', MDS was able to achieve a similar performance as the next best algorithm, CDS, but about 15% faster. For slightly higher degrees of motion in QCIF sequences, such as 'Table Tennis', MDS was able to increase the speed by about 19%. For CIF sequences with larger motions, MDS achieved 45% ('Akiyo with Crowd') and 75% ('Flower Garden') speed improvement with even better distortion performance. A smaller product of NSP and MAD can be considered as the figure-of-merit (FOM) for motion estimation algorithms. The MDS algorithm always resulted in a better FOM for all of the sequences, with up to 80% improvement ('Flower Garden') over CDS.

Table 1: Performance comparisons with other BMAs

BMA	NSP	MAD	NSP *MAD	NSP	MAD	NSP * MAD
	Miss America(QCIF)			Table Tennis(QCIF)		
FS	289.0	1.775	339.586	289.0	1.84	531.625
TSS	25.00	1.177	29.425	24.96	2.032	50.734
DS	13.33	1.176	15.678	13.71	1.84	25.221
CDS	9.48	1.176	11.149	10.11	1.840	18.998
MDS	8.23	1.178	9.693	8.45	1.889	15.963

CONCLUSION

A fast BMA using MDS patterns is proposed by exploiting motion correlations in real-life image sequences. The experimental results show that this MDS algorithm can achieve faster motion estimation speeds with similar or even better distortions. The MDS algorithm outperforms other fast BMAs, and hence, it can be applied to various video applications for high speed motion estimation.

REFERENCES

1 Li, R., Zeng, B., and Liou, M.L.: 'A new three-step search algorithm for block motion estimation', IEEE Trans. Circuits Syst. Video Technol., 1994, 4, pp. 438–442

2 Shan, Z., and Ma, K.: 'A new diamond search algorithm for fast Blockmatching motion estimation', IEEE Trans. Image Process., 2000, 6, pp. 313–317

3 Cheung, C., and Po, L.: 'A novel cross-diamond search algorithm for fast block motion estimation', IEEE Trans. Circuits Syst. Video Technol.,2003, 12, pp. 1168–1177

4 Hsieh, C.H., and Lu, P.C.: 'Motion estimation using interblock correlation'. Proc. Int. Symp. on Circuits and Systems, May 1990, pp. 995–998

EFFICIENT IMAGE SMOOTHING AND SHARPENING BY USING ADAPTIVE BILATERAL FILTER

B.NARESH REDDY,
C.S.P (M.TECH),
Email id: naresh200745@gmail.com,
Mobile: 09963978938,
GPREC, KURNOOL.

Mr. N.B.R. KUMAR, M.E,
Asst PROFESSOR,
Email id: kumarnbr@gmail.com,
Mobile: 09441312354,
GPREC, KURNOOL.

ABSTRACT

In this paper, the adaptive bilateral filter (ABF) for sharpness enhancement and noise removal is presented. The ABF sharpens an image by increasing the slope of the edges without producing overshoot or undershoot. It is an approach to sharpness enhancement that is fundamentally different from the unsharp mask. This new approach to slope restoration also differs significantly from previous slope restoration algorithms in that the ABF does not involve detection of edges or their orientation or extraction of edge profiles. In the ABF the edge slope is enhanced by transforming the histogram via a range filter with adaptive offset and width. The ABF is able to smooth the noise while enhancing edges and textures in the image. The parameters of the ABF are optimized with a training procedure. ABF restored images are significantly sharper than those restored by the bilateral filter. Compared with an USM based sharpening method ABF restored edges are as sharp as those rendered by the OUM, but without the halo artifacts that appear in the OUM restored image. In terms of noise removal, ABF also outperforms the bilateral filter. We demonstrate that ABF works well for both natural images and text images.

INTRODUCTION

Image restoration refers to the genre of techniques that aim to recover a high quality original image from a degraded version of that image given a specific model for the degradation process. This is in contrast to image enhancement techniques that seek to improve the appearance of an image without reference to a specific model for the degradation process. The restoration framework is particularly valuable because in conjunction with a training based approach, it provides a context within which the free parameters of the restoration algorithm may be optimized. Training based approaches have been used to develop imaging algorithms for a variety of applications, including image interpolation, image restoration, digital halftoning, descreening, and color correction .The ingredients that training-based approaches have in common when used for development of imaging algorithms are: 1) a set of training pairs each consisting of an input image and a desired output image, 2) an architecture for the algorithm consisting of free parameters, and 3) a cost function under which those free parameters may be optimized. In many cases, the architecture contains a classifier that allows for parameter optimization separately within different pixel classes according to the value of an appropriately chosen feature vector.

In this paper, we propose a new training based approach to image restoration. Once the restoration algorithm has been fully developed, we are, however, free to apply it to images for which the degradation process is unknown. This puts us back in the domain of enhancement. The success of this broader application of the restoration algorithm will depend on how general is the degradation model under which the algorithm was developed, as well as how robust is the overall structure of the algorithm to deviations from the assumed degradation model. The scope of this paper is to deal with images that are appropriate for dig-ital photography. We do not consider images that are severely degraded.

The two most common forms of degradation an image suffers are loss of sharpness or blur, and noise. The degradation model we use consists of a linear, shift-invariant blur followed by additive noise, described in detail in. The problem we are interested in is twofold. First we seek to develop a sharpening method that is fundamentally different from the unsharp mask filter (USM), which sharpens an image by enhancing the high-frequency components of the image.

In the spatial domain, the boosted high frequency components lead to overshoot and undershoot around edges, which causes objectionable ringing or halo artifacts. Our goal is to develop a sharpening algorithm that increases the slope of edges without producing overshoot and undershoot, which renders clean, crisp, and artifact-free edges, thereby improving the overall appearance of the image. The second aspect of the problem we wish to address is noise removal. We want to present a unified solution to both sharpness enhancement and noise removal. In most applications, the degraded image contains both noise and blur. A sharpening algorithm that works well only for noise-free images will not be applicable in these situations.

In terms of noise removal, conventional linear filters work well for removing additive Gaussian noise, but they also significantly blur the edge structures of an image. Therefore, a great deal of research has been done on edge-preserving noise reduction. One of the major endeavors in this area has been to utilize rank order information. Due to a lack of the sense of spatial ordering, rank order filters generally do not retain the frequency selective properties of the linear filters and do not suppress Gaussian noise optimally. Hybrid schemes combining both rank order filtering and linear filtering have been proposed in order to take advantage of both approaches,

These nonlinear rank order approaches in general improve the edge sharpness, but they are more complex to implement than a spatial linear filter. The idea of bilateral filtering has since found its way into many applications not only in the area of image de-noising, but also computer graphics, video processing, image interpolation, illumination estimation, as well as relighting and texture manipulation, dynamic range compression, and several others pointed out in. Several researchers have provided a theoretical analysis of the bilateral filter and connected it with the classical approaches to noise removal. Elad demonstrated that the bilateral filter emerges from the well known Bayesian approach when a novel penalty function is used. Based on this observation, he proposed methods to speed up the bilateral filtering and to implement a bilateral filter for piecewise linear signals. In Elad also pointed out that the bilateral filter is a discrete version of the short-time effective kernel of the Beltrami flow discussed in and. Barash and Comaniciu demonstrated that the nature of the bilateral filter resembles that of anisotropic diffusion and outlined a common framework for bilateral filtering, nonlinear diffusion, adaptive smoothing, and a mean shift procedure. The rest of this paper is organized as follows. In Section II, we describe the bilateral filter and how it works. In Section III, we present our proposed adaptive bilateral filter (ABF).

II. BILATERAL FILTER AND ITS PROPERTIES

The bilateral filter proposed by Tomasi and Manduchi in 1998 is a nonlinear filter that smoothes the noise while preserving edge structures. The shift-variant filtering operation of the bilateral filter is given by $[\quad,] = \sum\ \sum h[\quad,\quad;\ ],\ [\quad$,W]here $[\quad,]$ is the restored image $h[\quad,\quad;\ ]$,,is the response at *[m,n]* to an impulse at *[k,l]*, and *g[m,n]* is the degraded image. For the Bilateral filter impulse response would be given as

$$h[m_0,n_0;m,n]=\begin{cases} r_{m_0,n_0}^{-1}\exp\left(-\frac{(m-m_0)^2+(n-n_0)^2}{2\sigma_d^2}\right)\exp\left(-\frac{(g[m,n]-g[m_0,n_0])^2}{2\sigma_r^2}\right), & [m,n]\in\Omega_{m_0,n_0} \\ 0, & \text{else} \end{cases}$$

Where [,] is the center pixel of the window, σ and σ are the standard deviations of the domain and range Gaussian filters, respectively, and

$$r_{m_0,n_0}=\sum_{m=m_0-N}^{m_0+N}\sum_{n=n_0-N}^{n_0+N}\exp\left(-\frac{(m-m_0)^2+(n-n_0)^2}{2\sigma_d^2}\right)\quad\times\exp\left(-\frac{(g[m,n]-g[m_0,n_0])^2}{2\sigma_r^2}\right)$$

is a normalization factor that assures that the filter preserves average gray value in constant areas of the image. The edge-preserving denoising bilateral filter adopts a low pass Gaussian filter for both the domain filter and the range filter. The domain low pass Gaussian filter gives higher weight to pixels that are spatially close to the center pixel. The range low pass Gaussian filter gives higher weight to pixels that are similar to the center pixel in gray value. Combining the range filter and the domain filter, a bilateral filter at an edge pixel becomes an elongated Gaussian filter that is oriented along the edge. This ensures that averaging is done mostly along the edge and is greatly reduced in the gradient direction. This is the reason why the bilateral filter can smooth the noise while preserving edge structures. From a frequency domain perspective, bilateral filter is able to preserve edges while removing noise. On the other hand, the bilateral filter is essentially a smoothing filter. It does not sharpen edges. the edge rendered by the bilateral filter has the same level of blurriness as in the original degraded image, although the noise is greatly reduced. The results of the bilateral filtering are a significant improvement over a conventional linear low-pass filter. However, in order to enhance the sharpness of an image, we need to make some modifications to this filter.

III. ADAPTIVE BILATERAL FILTER (ABF) FOR IMAGE SHARPENING AND DE-NOISING

In this section, we present a new sharpening and smoothing algorithm: the adaptive bilateral filter (ABF). The response at [,]of the proposed shift-variant ABF to an impulse at *[m,n]* is Given by

$$h[m_0,n_0;m,n]=\begin{cases} r_{m_0,n_0}^{-1}\exp\left(-\frac{(m-m_0)^2+(n-n_0)^2}{2\sigma_d^2}\right)\exp\left(-\frac{(g[m,n]-g[m_0,n_0]-\zeta[m_0,n_0])^2}{2\sigma_r^2}\right), & [m,n]\in\Omega_{m_0,n_0} \\ 0, & \text{else} \end{cases}$$

And the normalization factor is given by

$$r_{m_0,n_0}=\sum_{m=m_0-N}^{m_0+N}\sum_{n=n_0-N}^{n_0+N}\exp\left(-\frac{(m-m_0)^2+(n-n_0)^2}{2\sigma_d^2}\right)$$
$$\times\exp\left(-\frac{(g[m,n]-g[m_0,n_0]-\varsigma[m_0,n_0])^2}{2\sigma_r^2[m_0,n_0]}\right)$$

The ABF retains the general form of a bilateral filter, but contains two important modifications. First, an offset ζ is introduced to the range filter in the ABF. Second, both ζ and the width of the range filter in the ABF are locally adaptive. If ζ=0 and is fixed, the ABF will degenerate into a conventional bilateral filter. For the domain filter, a fixed low-pass Gaussian filter with is adopted in the ABF. The combination of a locally adaptive ζ and transforms the bilateral filter into a much more powerful filter that is capable of both smoothing and sharpening. Moreover, it sharpens an image by increasing the slope of the edges. To understand how the ABF works, we need to understand the role of ζ and in the Adaptive Bilateral Filter.

Role of ζ in the ABF

The range filter can be interpreted as a 1-D filter that processes the histogram of the image. We will illustrate this viewpoint for the window of data enclosed in the box in the images. We index the images in the table by their [row, column] coordinates. For the conventional bilateral filter, the range filter is located on the histogram at the gray value of the current pixel and rolls off as the pixel values fall farther away from the center pixel value. By adding an offset ζ to the range filter, we are now able to shift the range filter on the histogram. As before, let Ω , denote the set of pixels in the (2N+1) ×(2N+1) window of pixels centered at [,] Let MIN, MAX, and MEAN denote the operations of taking the minimum, maximum, and average value of the data in Ω , respectively.We will demonstrate the effect of bilateral filtering with a fixed domain

Gaussian filter ($\sigma_{d=1.0)}$ and a range filter ($\sigma_{r=}20_{)}$ shifted by the following choices for ζ:

1) No offset (conventional bilateral filter): $[\ ,\ ]=0$.
2) Shifting towards the MEAN : $[\ ,\ ] = -\Delta_{\ ,}$
3) Shifting away from the MEAN : $[\ ,\ ] = \Delta_{\ ,}$
4) Shifting away from the MEAN, to the MIN/MAX

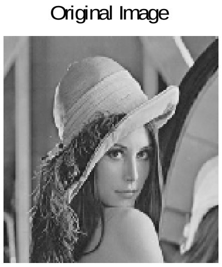
$$[\ ,\ ] =$$
$$\begin{matrix} \Omega_{\ ,} & -\ [\ ,\ ], & \Delta_{,} > 0 \\ \Omega_{\ ,} & -\ [\ ,\ ], & \Delta_{,} < 0 \\ 0 & , & \Delta_{,} = 0 \end{matrix}$$

Role of σ_r in the ABF

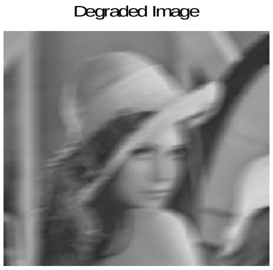
The parameter of the range filter controls the width of the range filter. It determines how selective the range filter is in choosing the pixels that are similar enough in gray value to be included in the averaging operation. If σ_r is large compared to the range of the data in the window, the range filter will assign similar weight to every pixel in the range. Then, it will not have much effect on the overall bilateral filter. On the other hand, a small σ_r will make the range filter dominate the bilateral filter. The bilateral filtered image resembles the range filtered image when, and it resembles the domain filtered image when $\sigma_{r=}5$ and it resembles the domain filtered image when σ_r $_{=}50$.

Summary of the Rationale for the ABF The pixel dependent offset ζ in the ABF is the key to slope restoration. With ζ , we are able to restore the slope by transforming the local histogram of the image, thus circumventing the cumbersome process of locating edge normal and detecting edge profiles. Since at any pixel $[m^{-}_{0},n_{0}]$ in the image, the ABF output is bounded between MIN($\Omega_{\ ,}$) and MAX($\Omega_{\ ,}$). In general the ABF does not produce overshoot and undershoot. By making ζ and σ_r adaptive and jointly optimizing both parameters, we transform the bilateral filter into a much more powerful and versatile filter. To smooth the image at a given pixel, we can shift the range filter towards MEAN ($\Omega_{\ ,}$), and/or use a large σ_r which enables the spatial Gaussian filter to take charge of the bilateral filtering. To sharpen the image at a given pixel, we can shift the range filter away from the midpoint of the edge slope which will be approximately equal to MEAN ($\Omega_{\ ,}$), towards MAX ($\Omega_{\ ,}$) or $_{\mathrm{MIN}}$ ($\Omega_{\ ,}$) depending on the position of the edge pixel on the edge slope. At the same time, we would reduce σ_r accordingly. With a small σ_r, the range filter dominates the bilateral filter and effectively pulls up or pushes down the pixels on the edge slope.

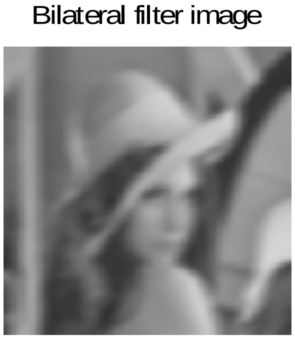
IVRESULTS



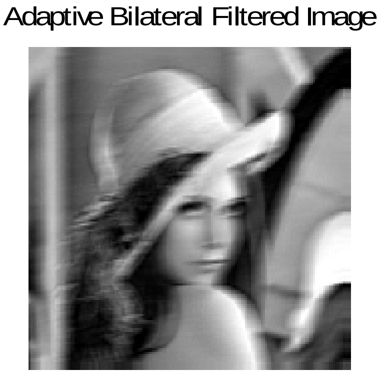
The above figure illustrates the original test image. This image is degraded and tested under two algorithms namely bilateral filter and ABF and resulting observations are observed.



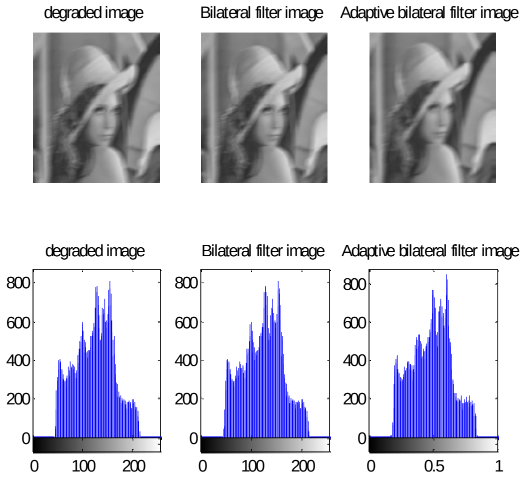
The original image is degraded by adding additive white Gaussian noise (AWGN) and also blurred by point spread function (PSF).



By applying bilateral filter to the degraded image the image is restored as above. in this it is clearly observed that only smoothing is performed but not edge sharpening. These Variations are observed in the following figure.



The above figure clearly mentions that both sharpening and smoothing are happened when compared to the previous approaches. It is concluded that the restored image by using ABF is approximately equals to original image.



The above figure gives a clear idea of comparison between existing method and proposed method. The comparison is shown in histogram.

V CONCLUSION

The adaptive bilateral filter is similar to bilateral filter but it contains two important modifications. First an offset ζ is introduced to range filter in ABF, second both ζ and width of the range filter σ_r are locally adaptive. The MSE between the original and reconstructed images is minimized for each class of pixels. The signal to noise ratio is improved compared with bilateral filter. The adaptive bilateral filter outperforms in noise removal. At the same time it renders much sharp images than the bilateral filter does. The quality of restored image is significantly improved compared with bilateral filter.

The ABF restored edges are as sharp as the OUM restored edges, but without the halo artifacts that the OUM produces. The ABF also achieves better noise suppression than the OUM. The ABF sharpens an image by increasing the slope of the edges. Previous slope restoration algorithms involve complex algorithms to determine edge orientation and edge profiles. In these approaches, the adjustment of the edge profiles tends to produce artifacts. In the ABF a new approaches to slope restoration: restoring edge slope by transforming the histogram of the edges. The ABF is efficient to implement, and provides a more reliable and more robust solution to slope restoration. The ABF works well for both natural images and text images.

REFERENCES

[1] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982, vol. 1.

[2] A. C. Bovik, Ed., "Regularization in image restoration and reconstruction," in *Handbook of Image & Video Processing*. San Diego, CA: Academic, 2000, ch. 3.6 (W. C. Karl), pp. 141–160.

[3] C. B. Atkins, C. A. Bouman, J. P. Allebach, J. S. Gondek, M. T. Schramm, and F. W. Sliz, "Computerized Method for Improving Data Resolution," U.S. Patent 058248, 2000.

[4] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Optimal image scaling using pixel classification," in *Proc. ICIP*, 2001, vol. 3, pp. 864–867.

[5] H. Hu and G. de Haan, "Classification-based hybrid filters for image processing," in *Proc. SPIE Int. Soc. Opt. Eng.*, 2006, vol. 6077, pp. 607711–607711.

[6] B. Zhang, J. Gondek, M. Schramm, and J. P. Allebach, "Improved resolution synthesis for image interpolation," in *Proc. IS&T's NIP22*, 2006, pp. 343–345.

[7] J. Tegenbosch, P. Hofman, and M. Bosma, "Improving nonlinear up-scaling by adapting to the local edge orientation," in *Proc. SPIE Int. Soc. Opt. Eng.*, 2004, vol. 5308, pp. 1181

[8] H. Kotera, Y. Yamada, and K. Shimo, "Sharpness improvement adaptive to edge strength of color image," in *Proc. IS&T/SID Eighth Color Imaging Conf.*, 2000

[9] H. Kotera and W. Hui, "Multi-scale image sharpening with noise re-duction," in *Proc. IS&T's NIP18*, 2002, pp. 590–594.

[10] S. Guillon, P. Baylou, M. Najim, and N. Keskes, "Adaptive nonlinear filters for 2D and 3D image enhancement," *Signal Process.*, vol. 67, pp. 237–254, 1998.

[11] R. Gonzalez and R. Woods, *Digital Image Processing* . Norwell, MA:Addison Wesley, 1992.

[12]A. Buades, B. Coll, and J.-M. Morel, "The staircasing effect in neigh-borhood filters and its solution," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1499–1505, Jun. 2006.

THE CONTENT-BASED IMAGE RETRIEVAL METHOD USING MULTIPLE FEATURES

k.santha (M.Tech), **Project Guide: K. Kantamma**

MITS College, Madanapalli. **Associate Professor, MITS College, Madanapalli,**

mail: kundusantha@gmail.com **Email: kantha.srinivas@yahoo.in**

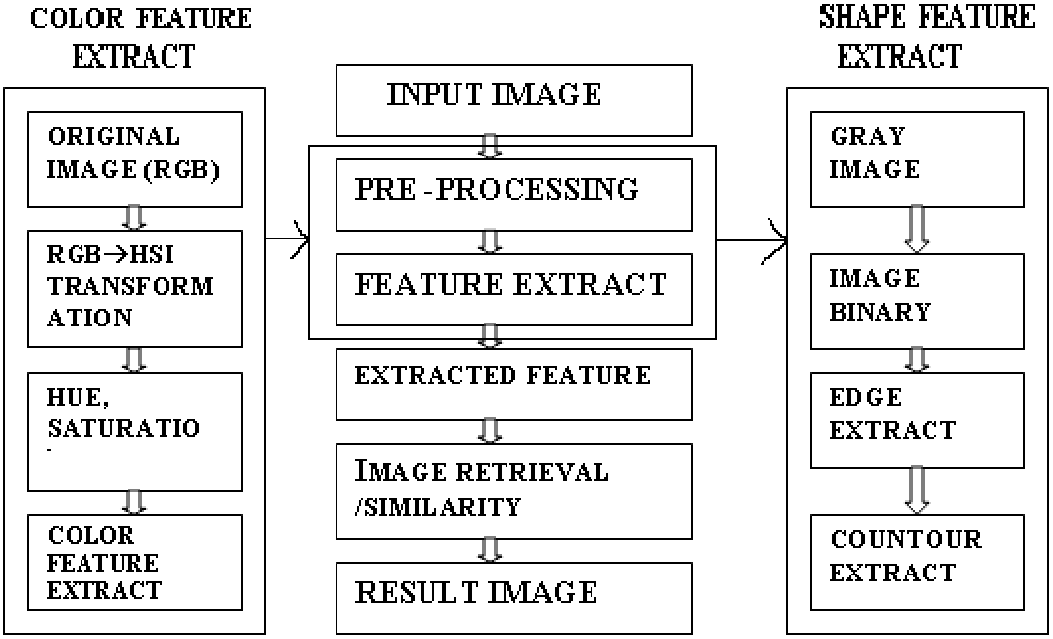
Introduction

Advanced in memory technologies and processing speed have made it feasible to store a large number of images in computers. This has given rise to the problem of organizing them for a rapid access to their content. An image database system aims to help people in this regard and enable them to find their desired images as quickly as possible. In content-based image database system, intrinsic properties of images are captured in some feature vectors which are indexed or compared to one another during query processing to find similar images from the database. Image can express a full of shape and color. So, on this paper propose both of adaptive color and shape information express mixed-features using by CSS(Curvature scale Space) and HSI Color Space that is one of model for can comparison and retrieval the image.

Scope of the Project

This paper be formed 4 steps propose, preprocessing, extract of feature, store information of Image and retrieval the Image. We used CSS(Curvature Scale Space) and HSI(Hue, Saturation, Intensity) to extract the feature points. On pre-processing, implement the Image processing for next step. Extract the RGB of pixel color information for color feature and the gray-level of pixel information for shape feature. On extract of feature, can extract feature of visual, this is retrieval. This is consisting of vector of feature that base on the retrieval similarity measure from color and shape. Extract process of color information show up the progress that transfer from original image data RGB value to HSI value. On extract of shape, one of step for can get the CSS Image, extract edge after transfer inputted color image to gray-level. Obtain the CSS image after extract contour by progress of contour tracking then, remove the noise by clustering. On storage information of image, efficiently can be storage and management the feature information of image and, store the vector and linked image file though the indexing progress on an image. Then, as last step, retrieval progress of image and measurement of similarity, extract and show up the best of quality. For example, user query by example image to here, first time extract maxima coordinates value store from
between vector of feature and image database then, compare the vector with the CSS image of query image. After output the image follow the top priority.

BLOCK DIAGRAM



Literature Survey

[1] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," IEEE Workshop on Applications of Computer Vision, pp. 96-102, 1996

In this paper the k.santha ,my project guiode:k.kantamma says,

Color histograms are widely used for content-based image retrieval. Their advantages are efficiency, and insensitivity to small changes in camera viewpoint. However, a histogram is a coarse characterization of an image, and so images with very di_erent appearances can have similar histograms. We describe a technique for comparing images called histogram re- _nement, which imposes additional constraints on histogram based matching. Histogram re_nement splits the pixels in a given bucket into several classes, based upon some local property. Within a given bucket, only pixels in the same class are compared. We describe a split histogram called a color coherence vector (CCV), which partitions each histogram bucket based on spatial coherence. CCV's can be computed at over 5 images per second on a standard workstation. A database with 15,000 images can be queried using CCV's in under 2 seconds. We demonstrate that histogram re_nement can be used to distinguish

[2] M. K. Mandal, T. Aboulnasr, and S. Panchanathan,"Image /indexing Using Moments and Wavelets," IEEE Transactions on Consumer Electronics, vol. 42, no. 3, pp. 557-565, Aug 1996.

In this paper the k.santha, project guide:K.kantamma says, Histogram comparison is a popular technique for image and video indexing. The complexity of the technique can be reduced by representing the histogram by its moments. In this paper, we propose two techniques to improve the performance of the basic histogram/moment-based technique. First, we propose to use orthogonal Legendre moments for representing histograms. Since Legendre moments are orthogonal, they provide superior indexing performance compared to regular moments at a similar complexity. Secondly, we propose to compare the histograms of wavelet coefficients at different scales. The wavelet coefficients provide important directional information, and hence improve the performance of the basic histogram-based technique. The proposed scheme can be easily extended to color images and also be integrated into a wavelet-based image coder.

[3] Jing Huang, S. Ravi Kumar, mandar Mitra, Wei-Jing Zhu, Ramin Zabih," "Image Indexing Using Color Correlogram," "International Conference on Computer Vision and Pattern Recognition, IEEE, 1997.

In this paper the k.santha,projectguide:K.kantamma says,

We define a new image feature called the color correlogram and use it for image indexing and comparison. This feature distills the spatial correlation of colors, and is both effective and inexpensive for content-based image retrieval. The correlogram robustly tolerates large changes in appearance and shape caused by changes in viewing positions, camera zooms, etc. Experimental evidence suggests that this new feature outperforms not only the traditional color histogram method but also the recently proposed histogram refinement methods for image indexing/retrieval.

Module Separation

- Shape feature extract.
- Color feature extract.

Module Description

- ***Shape feature extract.***

  An extract method of feature-vector is following below steps.

  1. Transfer RGB color information of extract pixels to gray-level information on preprocessing.

  2. Make a binarized by threshold after transfer.

  3. Extract contour of image by apply LoG (laplacian of Gaussian).

  4. Make a contour tracing from extract of contour

  5. Get the Circularity use by equation (1) after tracking the contour

  6. Following the sequential smoothing on contour, is doing until the curve be not extant.

- ***Color feature extract.***

The correct Extract of feature vector means extract visual feature information that extract pixel RGB color information and gray-level information with by preprocessing. As color feature, uses the intersection of histogram by gets the values (hue, saturation and

intensity) from translate RGB Model to HSI model. This method is that use indexing information that pixels into image color and, checked color histogram of query image and color histogram on DB (all image). Meaning is that simply compute method because, compute the frequency of color. If it similar each other histogram, can be retrieval. So, case of object shift or rotation on image, can be get the more result of retrieval. But, it have weakness that can't use spatial information if not include it. RGB set up an ideal for image. But, for the description of color restrict to use RGB. But HSI is can be get it. Method of image extract on HSI model use by hue and intensity. Object has brightness unlike pixels of foreground so, preferentially, find threshold by similarity brightness. In case of similarity brightness, it is a different that between object and side of one. For color histogram configuration, it use color factor (H) in HSI. As use by hue (H) and saturation(S), get the advantage for reduce the intensity (I) translation. Also, useful for get the memory and compute because, can use histogram on planar.

ALGORITHMS USED

- HSV
- Edge Detection
- CSS(Curvature Scale Space)

Advantages

- Accuracy is high compared with single feature use of image ,which results rotation-transition

Applications

- Image database system.

Reference and bibliography

[1] Apostol, N., Rajeev, R,. Kyuseok, S.:WALRUS: A similarity retrieval algorithm for image database. In: Proceeding of the ACM SIGMOD International Conference on Management of Data, pp. 395-406. ACM Press, New York (1999)

[2] S. Abbasi, Curvature scale space in shape similarity retrieval, Ph.D. thesis, Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 5XH, England, 1998.

[3] F. Mokhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and twodimensional shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8,no. 1, pp. 34-43,1986.

[4] Sadegh Abbasi and Farzin Mokhtarian, "Robustness of Shape Similarity Retrieval under Affine Transformation", 1999.

[5] G. Bebis, G. Papadourakis, and S. Orphanoudakis, "Curvature scale space driven object recognition with an indexing scheme based on artificial neural networks", accepted Pattern Recognition (also available from http://www.cs.unr.edu/~bebis).

[6] M. Swain and D. Ballard, "Color Indexing," Intl'l Journal of Computer Vision, Vol. 7, No. 1, pp.11-32, 1991.

[7] Q.T. Luong, "Color in Computer vision," handbook of Pattern Recognition and Computer Vision, pp.311-368, 1993.

[8] V. N. Gudivada and V. V. Raghavan, "Content-based image retrieval systems, "IEEE Computer, pp.18-22, Sept. 1995

[9] M. Stricker and A. Dimai, "Color Indexing with Weak Spatial Constraints," Storage and Retrieval for Image and Video Databases IV, SPIE Proceedings vol. 2670, 1996

[10] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," IEEE Workshop on Applications of Computer Vision, pp. 96-102, 1996 [11] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Image /indexing Using Moments and Wavelets," IEEE Transactions on Consumer Electronics, vol. 42, no. 3, pp. 557-565, Aug 1996.

[11] Jing Huang, S. Ravi Kumar, mandar Mitra, Wei-Jing Zhu, Ramin Zabih," "Image Indexing Using Color Correlogram," "International Conference on Computer Vision and Pattern Recognition, IEEE, 1997.

Fractional order singular value decomposition representation for face – recognition

J.Kamalakar,
M.Tech (DECS) ,
MITS, Madanapalli,
Chittor(dt),
AndraPradesh .
jyothikamal23@gmail.com,

K.Kantamma, Associate. Professor,
Department of ECE
MITS, Madanapalli,
Chittor (dt),
AndraPradesh.
kantha.srinivas@yahoo.in,

Abstract: Face representation (FR) plays a typically important role in face recognition and methods such as principal component analysis (PCA) and linear discriminate analysis (LDA) have been received wide attention recently. However, despite of the achieved successes, these FR methods will inevitably lead to poor classification performance in case of great facial variations such as expression, lighting, occlusion and so on, due to the fact that the image gray value matrices on which they manipulate are very sensitive to these facial variations. In this paper, we take notice of the facts that every image matrix can always have the well-known singular value decomposition (SVD) and can be regarded as a composition of a set of base images generated by SVD, and we further point out that the leading base images (those corresponding to large singular values) on one hand are sensitive to the aforementioned facial variations and on the other hand dominate the composition of the face image. Then based on these observations, we subtly deflate the weights of the facial variation sensitive base images by a parameter *з* and propose a novel fractional order singular value decomposition representation (FSVDR) to alleviate facial variations for face recognition. Finally, our experimental results show that FSVDR can: (1) effectively alleviate facial variations; and (2) form an intermediate representation for many FR methods such as PCA and LDA to significantly improve their classification performance in case of great facial variations.

***Keywords**:* Singular value decomposition (SVD); Fractional order singular value decomposition representation (FSVDR); Face representation (FR); Intermediate representation (IR); Face recognition

I. Introduction: Machine recognition of human face from still and video images has become an active research area in the communities of image processing, pattern recognition, neural networks and computer vision. The most remarkable abilities of human vision are that of face recognition. It develops over several years of development, and is important for several aspects of our social life, and together with related abilities, such as estimating the expression of face with which we process, has played an important role in the course of face recognition evolution. The problem of face recognition was considered in the early stages of computer vision and is still undergoing various evolutions from a long period. Different techniques were proposed in past for the enhancement of face recognition. This enhancement is motivated by wide applications ranging from static matching of controlled format photographs such as passports, credit cards, driving licenses, access control systems, model-based video coding, criminal identification and authentication in secure system like computer or bank teller machines etc. to real-time matching of surveillance video images presenting different constraints in terms of processing requirements.

Although many face recognition by human beings and machines were developed in past, it is still difficult to design an automatic system for the task because in real world, illumination, complex background, visual angle and facial expression for face images are highly variable. it is still difficult to design an automated system for accurate face recognition, especially in real-time identifications. The main reasons for the existing face recognition system can be described as:

Furthermore, the lighting, background, scale, and parameters of the acquisition are all variables in facial images acquired under real-world scenarios. The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to changes in the face identity. This makes face recognition a great challenging problem.

II.Face Recognition

Although human beings can easily detect and identify faces in a scene, it is very challenging for an automated system to achieve such objectives. The challenges become more profound when large variations exist in the face images. Despite of these challenges, face recognition has drawn wide attention from researchers in areas of machine learning, computer vision, pattern recognition, neural networks, and are efficiently been used in areas of access control, information security, law enforcement and surveillance, smart cards and so on.

Face recognition plays an important role in biometrics based personal identification. A biometrics verification system is designed to verify or recognize the identity of a living person on the basis of his/her physiological characters, such as face, fingerprint, and iris, or some aspects of behavior such as handwriting or keystroke pattern. The need for reliable identification of interacting users is obvious. The biometrics verification technique acts as an efficient method and has wide applications in the areas of information retrieval, automatic banking, control of access to security areas, buildings, and so on. In general, an unsupervised learning approach cannot get a high recognition rate. Under conditions where we cannot acquire a large number of face images for every person, utilizing all available samples is very important. This means that not only positive samples but also negative samples need to be learned for such face recognition systems. These systems Detroit from the accuracy level of estimation when the trained information's are limited. This limitation result in the demand for an enhanced methodology for face recognition.

A. Face Recognition Objective

In the language of information theory, the objective is to extract the relevant information in a face image, encode it as efficiently as possible, and compare one face encoding with a database of models encoded in the same way. A simple approach to extract the information contained in a face image is to somehow capture the variation in a collection of face images, independent of any judgement of features, and use this information to encode and compare individual face images.

In mathematical terms, the objective is to find the principal components of the distribution offices, or the eigenvectors of the covariance matrix of the set of face images. These eigenvectors can be thought of as a set of features which together characterize the variation between face images. Each image location contributes more or less to each eigenvector, so that we can display the eigenvector as a sort of ghostly face called an eigenface.

Each face image in the training set can be represented exactly in terms of a linear combination of the eigenfaces. The number of possible eigenfaces is equal to the number of face images in the training set. However, the faces can also be approximated using only the "best" eigenfaces those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images. The primary reason for using fewer eigenfaces is computational efficiency. The most meaningful M eigenfaces span an M-dimensional subspace "face space" of all possible images. The eigenfaces are essentially the basis vectors of the eigenface decomposition.

The idea of using eigenfaces was motivated by a technique for efficiently representing pictures of faces using principal component analysis. It is argued that a collection of face images can be approximately reconstructed by storing a small collection of weights for each face and a small set of standard pictures. Therefore, if a multitude of face images can be reconstructed by weighted sum of a small collection of characteristic images, then an efficient way to learn and recognize faces might be to build the characteristic features from known face images and to recognize particular faces by comparing the feature weights needed to (approximately) reconstruct them with the weights associated with the known individuals.

B. Existing Methodologies

As mentioned earlier Face recognition is a problem, which is one of the most researched problems in Computer Vision and Artificial Intelligence.

In this section, we review existing techniques to detect faces from a single intensity or color image. Single image detection methods can be classified into four categories

a) Knowledge-based methods: These rule-based methods encode human knowledge of what constitutes a typical face. Usually, the rules capture the relationships between facial features. These methods are designed mainly for face localization.

b) Feature invariant approaches: These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these to locate faces. These methods are designed mainly for face localization.

c) Template matching methods: Several standard patterns of a face are stored to describe the face as a whole or the facial features separately. The correlations between an Table 1.1:Categorization of methods for Face Detection input image and the stored patterns are computed for detection. These methods have been used for both face localization and detection.

d) Appearance-based methods: In contrast to template matching, the models are learned from a set of training images which should capture the representative variability of facial appearance. These learned models are then used for detection. These methods are designed mainly for face detection.

III. Singular Value decomposition (SVD)

In current face recognition system; there have been studies of employing SVD to obtain representation for face images. Hong [12] proposed to apply SVD to each Face Image to obtain singular values (SVs) to represent this face image, and then to perform classification based on these SVs. Cheng et al. [13] made use of the SVs as an IR, and then employed an optimal discriminate transformation to transform the SVs into a new space for subsequent classification. Tian et al. [14] pointed out that the SVs contained little useful information for face recognition and attributed the good performance reported in [13] to the small testing database. Though SVD based approach is found to be efficient for face recognition this technique is limited under facial expression variation and illumination effects. To elevate the problem of facial expression, occlusion, illumination a fractional order singular value decomposition representation (FSVD) is suggested in this thesis.

The singular value decomposition of a matrix A of m x n matrix is given in the form,

$$A = U\Sigma V^T$$

Where U is an m x m orthogonal matrix; V an n x n orthogonal matrix, and ∑ is an m x n matrix containing the singular values of A.

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

along its main diagonal.

A similar technique, known as the eigenvalue decomposition (EVD), diagonalizes matrix A, but with this case, A must be a square matrix. The EVD diagonalizes A as

$$A = VDV^{-1}$$

Where D is a diagonal matrix comprised of the eigenvalues, and V is a matrix whose columns contain the corresponding eigenvectors. Where Eigen value decomposition may not be possible for all facial images SVD is the result.

IV. Fractional order singular value decomposition (FSVD)

To alleviate the facial variations on face images, a novel FSVDR is suggested.

The main ideas of FSVD approach are that;

(1) The weights of the leading base images $u_i v_i^T$ should be deflated, since they are very sensitive to the great facial variations within the image matrix A itself.

(2) the weights of base images $u_i v_i^T$ corresponding to relatively small σ_i 's should be inflated, since they may be less sensitive to the facial variations within A.

The order of the weights of the base images $u_i v_i^T$ in formulating the new representation of SVD should be retained. More specifically, for each face image matrix A which has the SVD, its FSVD 'B' can be defined as,

$$\mathbf{B=U\ \Sigma^{\alpha} V^T}$$

Where U, Σ and V are the SV matrices, and in order to achieve the above underlying ideas, α is a fractional parameter that satisfies:

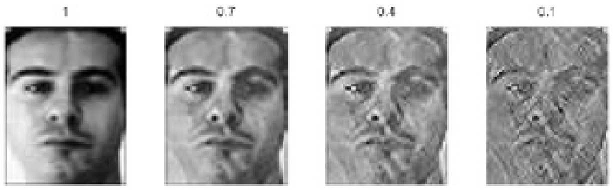
$$0 \leq \alpha \leq 1$$

it is seen that the rank of FSVDR 'B' is r, i.e., identical to the rank of A as the B matrix is fractional raised the values are inflated retaining the rank of the matrix constant.

The $u_i v_i^T$, i = 1, 2, . . . , r form a set of uv^T which are similar to the base images for the SVD approach.

It is observed that the intrinsic characteristic of A, the rank, is retained in the FSVD approach. In fact it has the same uv^T like base images as A, and considering the fact that these base images are the components to compose A and B, the information of A is effectively been passed to B.

As FSVD approach uses a fractional parameter, α which inflates the lower SV, the effect of α on the above-illustrated image is been presented below,



(a) (b) (c)

(d)

Fig 3.2: Face image under variation of fractional factor (a) α = 1, (b) α = 0.7, (c) α = 0.4, (d) α = 0.1

from the observation it could be observed that:

(1) The FSVD is still like human face under lower SV.

(2) The FSVD deflates the lighting condition in vision. Taking the two face images (c) and (d) under consideration, when α is set to 0.4 and 0.1, from the FSVD alone, it is difficult to tell whether the original face image matrix A is of left light on or right light on.

(3) The FSVD reveals some facial details. In the original face images (a) presented, neither the right eyeball of the left face image nor the left eyeball of the right face image is visible, however, when setting α to 0.4 and 0.1 in FSVD, the eyeballs become visible.

In the case of FSVD thus the fractional parameter and it's optimal selection is an important criterion in making the face recognition process more accurate.

V.Database description: We carry out experiments on three renowned face databases: AR, FERET and YALE

The AR database consists of over 4000 color images of 126 person's faces (70 men and 56 women). Each person has 26 different images which were grabbed in two different sessions separated by two weeks, and 13 images in each session were recorded. The 13 images are, respectively, of neutral expression, smile, anger, scream, left light on, right light on, both light on, occlusion by glasses and left light on, occlusions by glasses and right light on, occlusions by glasses and both light on, occlusion by scarves and left light on, occlusions by scarves and right light on, occlusions by scarves and both light on. illustrates the 26 image faces under different facial variations

from one subject in AR face database. In our experiments here, we use a subset of the AR face database provided and preprocessed by Martinez This subset contains 2600 image faces corresponding to 100 person (50 men and 50 women), where each person has 26 different images under the aforesaid conditions. The original resolution of these image faces is 165 × 120. Here, for computational convenience, we resize them to 66×48, and the gray level values are rescaled to [0 1]. As can been seen the AR face database is very challenging. Here, we carry out four independent experiments, AR1, AR2, AR3 and AR4, where the training and testing samples are listed in Table 1. From we can observe that: AR1 evaluates the classification performance over time with variations in expression and lighting conditions; AR2 tests the classi-fication performance in case of occlusions by wearing glasses; AR3 evaluates the classification performance in case of occlusions by wearing scarves; AR4 tests the classification perfor-mance over time with great variations in expression, lighting conditions and occlusions such as wearing glasses and scarves.

Table 1
Data partition on AR face database

Category	Training	Testing
AR1	a, b, c, d, e, f, g	n, o, p, q, r, s, t
AR2	a, b, c, d, e, f, g	h, i, j
AR3	a, b, c, d, e, f, g	k, l, m
AR4	a, b, c, d, e, f, g, h, i, j, k, l, m	n, o, p, q, r, s, t, u, v, w, x, y, z

Table 1

The FERET database is one of the most well-known face recognition benchmarks. The Color FERET database contains a total of 11 338 facial images corresponding to 994 subjects, and the Gray FERET contains a total of 14 051 grayscale im-ages corresponding to 1209 subjects. Here, we carry out experiments on the hardest subset of FERET Tests September 1996, whose testing samples have great facial variations in illumina-tion. More specifically, we employ the gallery set that contains 1196 face images as training set and the fafc set that has 194 face images as testing set. The face images are preprocessed according to the CSU Face Identification Evaluation System with a resolution of 75 × 65. The challenges of this FERET subset are: (1) a large number of subjects (1196) in the training set, (2) one training sample per class and (3) great illumination variations in the testing set. Due to difficulty of this subset, we follow the CSU Face Identification Evaluation System to report the Rank *k* classification, where the testing sample is considered to be correctly classified so long as it belongs to the same class as one of its *k* nearest neighbor (NN) samples in the training set.

The YALE face database contains 165 gray level face im-ages of 15 persons. There are 11 images per subject, and these 11 images are, respectively, under the following different facial expression or configuration: center-light, wearing glasses, happy, left-light, wearing no glasses, normal, right-light, sad, sleepy, surprised, and wink. In our experiment, the images are cropped to a size of 50 × 50, and the gray level values of all images are rescaled to [0 1]. shows the 11 images of one person from this database. On YALE face database, we perform two different experiments, YALE1 and YALE2, where the training and testing samples are given in. From and we know that: YALE1 evaluates the clas-sification performance in case of different lighting conditions and wearing glasses; and YALE2 evaluates the classification performance in case of some distinct expressions and lighting conditions.

VI.Conclusion

In this paper a method for face recognition is suggested based fractional singular value decomposition method. The method is observed to give better result compared to the existing SVD based face recognition. It is observed that the fractional factor applied onto the SVD features result in high accuracy in estimation under lower value assumption and deviates when increased. The facial variation effect is observed to be minimized in case of FSVD approach as compared to SVD based face recognition.

feature for different input samples for its robustness to face recognition

VII.Acknowledgments

I would like to thank the anonymous reviewers for their constructive comments that enhance this paper, providing the AR database, the FERET database, and the YALE database.

VIII.References:

1. Bruce, P.J.B. Hancock, A.M. Burton, and Face Recognition: From Theory to Applications, Springer,

2. A. Goldstein, P. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey, Pattern Recognition

LAB-VIEW BASED LINEAR FILTERING APPROACH TO DIGITAL DTMF DETECTION USING GOERTZEL ALGORITHM

Mr. Prasanna Kumar M.K.[1], Mr.Gandhimathinathan.A.[2], Mr.Prashanth Barla[3]

MR.PRASANNA KUMAR[1]	**MR.GANDHIMATHINATHAN[2]**	**MR.PRASANTH BARLA[3]**
Lecturer, ECE Dept	Lecturer, ECE Dept	Lecturer, ECE Dept
V.C.E.T, Puttur	S.J.C.E, Mangalore	S.C.E.M, Mangalore
prasanna251983@yahoo.com	**agandhimathinathanbe@gmail.com**	**p**rashanthbarla@gmail.com

Abstract

A linear filtering approach to the computation of Discrete Fourier Transform (DFT) is very useful when only a selected number of values of DFT are desired, but the entire DFT is not required. A DFT based algorithm known as Goertzel algorithm is based on linear filtering approach which can be used for digital Dual Tone Multifrequency (DTMF) detection. In this paper a complete DTMF system (generation as well as detection) is implemented along with Goertzel second order IIR filter in Lab VIEW. Lab VIEW is the widely used graphical programming environment which allows designing systems in an intuitive block based manner in shorter times as compared to the text based programming languages. Basically, this paper demonstrates the design of second order Goertzel IIR filter transfer function starting from DFT algorithm and the ease with which complete DTMF system is implemented using Lab VIEW graphical programming environment. Examples are provided to demonstrate the detection of the corresponding DTMF frequency components for the key pressed

Keywords: DTMF, DFT, GOERTZEL, Lab VIEW graphical programming, IIR, FFT, PCM.

Abbreviations:

DFT	= Discrete Fourier Transform
IIR	= Infinite Impulse Response
DTMF	= Dual Tone Multi Frequency
FFT	= Fast Fourier Transform
PCM	= Pulse Code Modulation
DSP	= Digital Signal Processing
VLSI	= Very Large Scale Integration
LAB-VIEW	= Laboratory Virtual Instrumentation Engineering Workbench.

I. Literature Review:

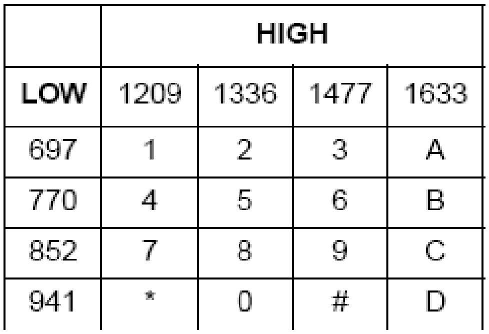
The Goertzel algorithm is a digital signal processing (DSP) technique for identifying frequency components of a signal, published by Dr. Gerald Goertzel in 1958. While the general Fast Fourier transform (FFT) algorithm computes evenly across the bandwidth of the incoming signal, the Goertzel algorithm looks at specific, predetermined frequencies. A practical application of this algorithm is recognition of the DTMF tones produced by the buttons pushed on a telephone keypad. It can also be used "in reverse" as a sinusoid synthesis function, which requires only 1 multiplication and 1 subtraction per sample

II. Introduction:

The first touch tone telephone installation was in 1963. DTMF signaling uses voice band tones to send. Address signals and other digital information from pushbutton Telephones and other devices such as modems and fax machines. Analog DTMF detection is done using band pass filter banks with center frequencies at the DTMF signal frequencies. Analog receivers have wide tolerances to compensate for distortion caused by aging transmitters, variations in keying characteristics, and transmission line distortion. These distortions compound the problem of digital

DTMF detection. The introduction of digital pulse code modulation (PCM) switches in 1976 signaled the beginning of the end of the analog telephone network. In the past 20 years telephone networks have been rapidly moving from totally analog to totally digital. In digital switching systems it is desirable to treat all signals uniformly, bringing all signals through A/D converters and switching them through the PCM system. Therefore the need for digital DTMF detection is justifiable to avoid the costs of hardware and D/A conversion needed to use analog detectors.

With the constant advances in VLSI driving DSP costs downward, it is economically sound to replace analog detectors with their digital counterparts which are more reliable, maintenance cost effective, and spatially minimal. Several techniques for digital DTMF detection have been used, but most designers have settled on either Digital filtering or discrete Fourier transform (DFT). In digital filtering, DTMF signals are passed through Digital band pass filters centered at the signaling frequencies (shades of analog). The power at each frequency is then measured repeatedly to detect the DTMF tones. A DSP then interprets and translates them for the proper switching. The DFT is the technique that is outlined in this paper. As can be seen in Table 1, the DTMF signaling frequencies are very closely spaced. It is obvious that the bandwidth of the filter used for detection must be narrow enough to avoid any bleeding of adjacent frequencies. An even more limited bandwidth is introduced when one considers that some of today's DTMF signals generators (phones, modems, etc.) will determine whether DTMF signals are present or not is done using an adaptive level detection scheme.



	HIGH			
LOW	1209	1336	1477	1633
697	1	2	3	A
770	4	5	6	B
852	7	8	9	C
941	*	0	#	D

Table 1: DTMF Frequencies and Channel assignments

Bell Core specifies that for adaptive threshold, the signal frequency must be present at a power level of 9dB greater than that of the other signal frequencies in the same group (high or low, see Table 1). Once a valid level is found, it is desirable to know how long the signal is present at that level. Timing is another important issue when doing DTMF detection. The duration of the signal is critical to the accuracy of the detector. Bell Core specifies several timing constraints: signal duration, inter digit time, and cycle time. According to Bell Core, a DTMF detector should recognize tones of 40ms or greater. Alternatively, the detector can recognize tones with durations as low as 23ms. Any signal with a duration < 23ms must be rejected. This specification is instrumental for ensuring robustness to noise and speech.

III. Goertzel Algorithm:

A. The Algorithm:

The algorithm we used to compute the Fourier transform X (w) is based substantially on Goertzel algorithm. This algorithm models the computation of the DFT as a linear filtering operation. This operation takes the form of a parallel bank of resonators, where each resonator selects one of the frequencies of the DFT. The output of each of theses filters at n=N yields the value of the DFT at the Frequency wk = 2*PI*k/N. The advantages of this approach over other algorithms are 1) it is computationally more efficient. 2) The value of the DFT can be computed at any frequency desired. The extent to which the Goertzel algorithm is more efficient than other algorithms (such as Fast Fourier Transforms) depends on the number of frequencies at which the DFT is to be computed. Each iteration of the Goertzel algorithm requires one real multiplication and two real additions. If the value of the DFT is required at M points, then the total cost for computing the DFT is M*N multiplications and 2*M*N additions. For values of M < Log N, Goertzel algorithm is less expensive computationally than an FFT Algorithm.

Goertzel algorithm may be used to compute the value of the DFT at any frequency, and with any value of N. This is critical in DTMF detection because the resolution between Frequencies decreases as the value of N decreases. If the FFT were used, the value of the DFT could be computed Only at N equally spaced frequencies, which may or may

Not correspond to the frequencies of interest.

B. The Modifications:

As previously stated, the algorithm presented here is based heavily on Goertzel algorithm; several substantial modifications were made, however, to improve the overall computational efficiency of the DTMF detector. 1) First, we recognize that we are interested in computing the DFT at a very small set of frequencies. We create an array of size eight containing only the values of k corresponding to the eight frequencies with which we are interested. The algorithm was then modified to compute the DFT at only these frequencies. 2) Secondly, we recognized that the values of k need not be restricted to integers. This results directly from the fact that X (w) is continuous in w by allowing k to take on floating point values, the DFT can be computed at precisely the frequency of interest, within round off error, regardless of the value of N. This single modification eliminated altogether the problems associated with frequency resolution. 3) A third improvement came from the realization that Goertzel algorithm was designed to handle a complex input data sequence x (n). However, the real and complex portions of the computations are separated in the algorithm. Because samples of a real signal are real valued, many of the computations were unnecessary and were eliminated. This effectively halved the number of computations required for each iteration. 4) A fourth modification that was made was to pre compute the phase factors Wk. This saves sixteen trigonometric evaluations per DFT computation, at a cost of the storage of sixteen floating point numbers. 5) Finally, the algorithm was modified to extract the input sequence x (n) from a circular buffer. The original Goertzel algorithm assumes that the data is stored in a vector with x (0) located in element 0 of the vector. Linear storage of the data proved to be awkward. The entire vector had to be reconstructed each time the DFT was to be recomputed.

C. Structure of second order IIR Goertzel filter:

$$v(n) = 2\cos(\tfrac{2\pi k}{N})v(n-1) - v(n-2) + x(n)$$

$$\mathbf{y(n) = v(n) - w_N^k}v(n-1)$$

$$W_N^k = \exp\left(-j\frac{2\pi k}{N}\right).$$

$$\left|y_k[N]\right|^2 = v_k^2[N] + v_k^2[N-1] - 2\cos\left(\frac{2\pi k}{N}\right)v_k[N]v_k[N-1]$$

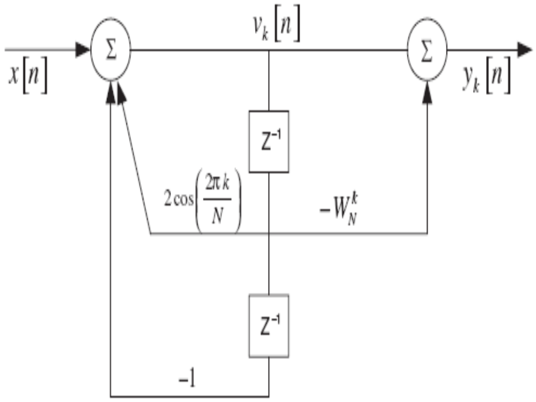


Figure 1: Structure of second order IIR Goertzel filter

The above figure 1 represents the structure of second order Goertzel filter. As indicated in figure 2, seven Goertzel filters are used in parallel to form a DTMF detection system. Each Goertzel filter is designed to detect a DTMF tone. The output from each filter is squared and fed into a threshold detector, where the strongest signal from low and high frequency groups are selected to identify a pressed digit on keypad.

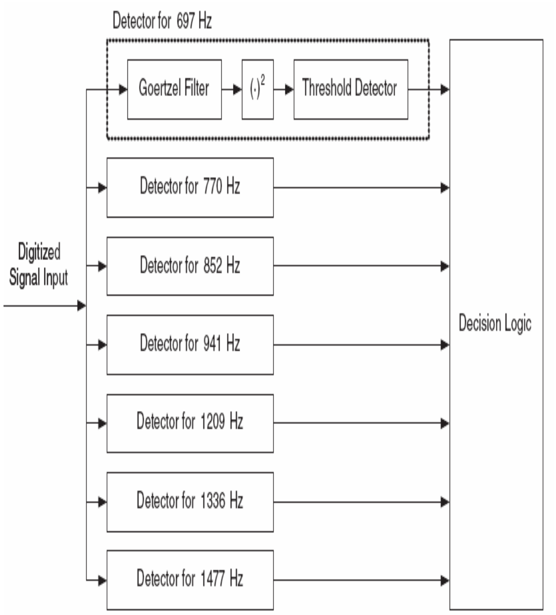


Figure2: DTMF receiver system

D. Calculations of bin ('k') values:

The value of the constant 'k' determines the tone that we are trying to detect and is given by

$$k = N \times \frac{f_{tone}}{f_s}$$

Where ftone=frequency of the tone

fs =sampling frequency

For N=205 and fs=8 kHz, calculated 'k' values are as given in table 2.

Frequency (Hz)	k
697	18
770	20
852	22
941	24
1209	31
1336	34
1477	38
1633	42

Table 2: Bin ('k') values

E. Calculations of the filter coefficients:

Once we have 'k' values we can calculate the coefficients 2cos (2*π*k/N) for individual 'k' values. The table 3 shows the calculated values of the coefficients which are fed to the circular buffer

Frequency(Hz)	coefficient
697	1.703275
770	1.635585
852	1.562297
941	1.482867
1209	1.163138
1336	1.008835
1477	0.790074
1633	0.559554

Table3: Calculation of Coefficients

IV. Lab VIEW Implementation of DTMF system

A. DTMF Generation: DTMF is generated with a help of 4*3 keypad. The keypad is connected to sine wave generators for generating two sine waves for corresponding key pressed. Later two sine waves are added

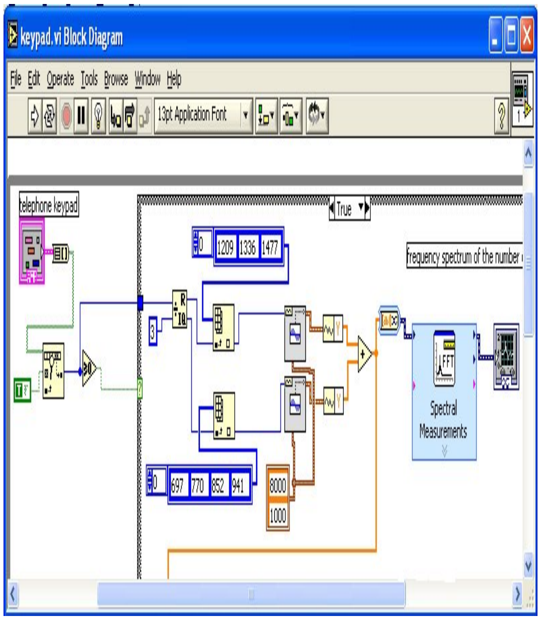


Figure 3: DTMF generation

B. Goertzel filter: The structure of the second order Goertzel filter is discussed above and the same is implemented in lab view.

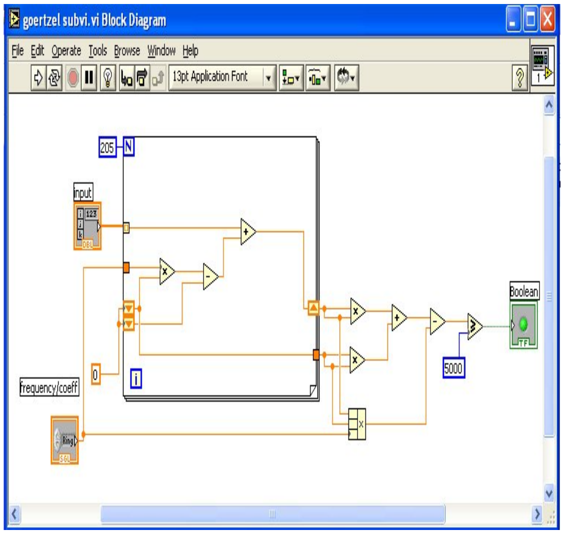

goertzel subvi.vi Block Diagram
File Edit Operate Tools Browse Window Help
13pt Application Font
205
N
input
frequency/coeff
0
i
5000
Boolean

Figure 4: Second order Goertzel IIR filter

C. Creating Sub VI of Goertzel filter: The filter above is made as a subroutine and called 8 times to detect for 8 different frequencies. Below figure shows the sub VI as block 1

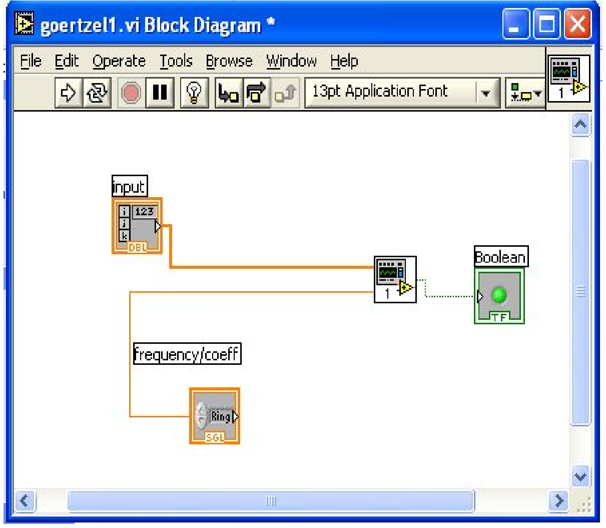

goertzel1.vi Block Diagram *
File Edit Operate Tools Browse Window Help
13pt Application Font
input
Boolean
frequency/coeff
Ring

Figure 5: Creating Sub VI of Goertzel filter

D. Complete DTMF system: In the complete DTMF system the DTMF generator part is cascaded with the DTMF receiver part. DTMF receiver part is designed according to figure 2 discussed above.

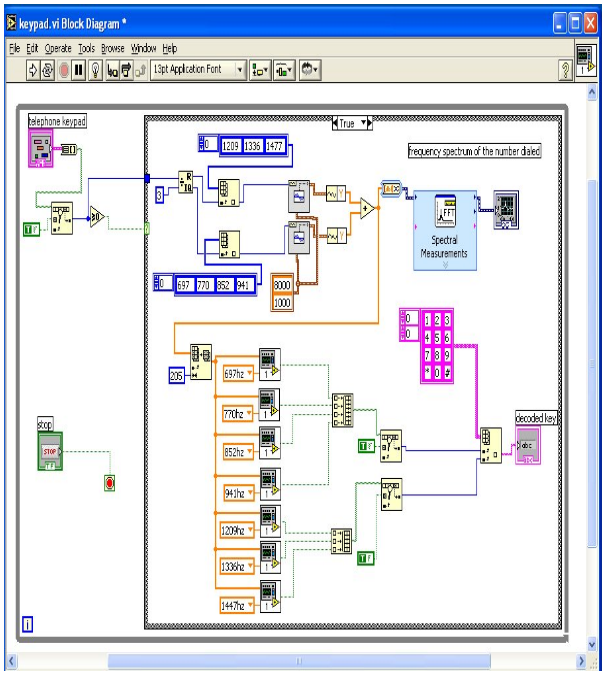
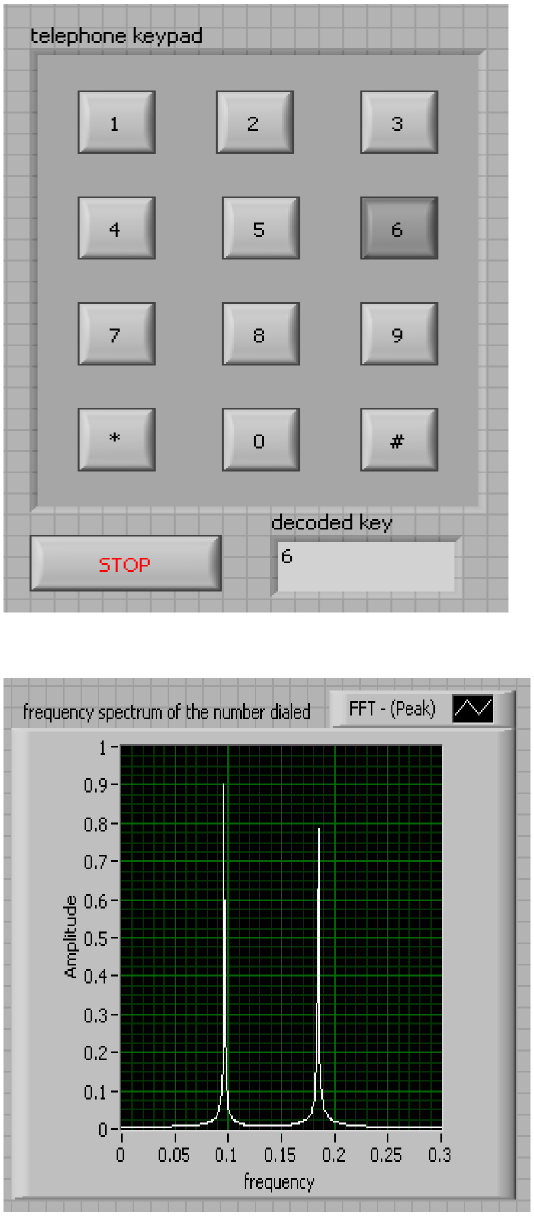


Figure 6: Complete DTMF System

V. Result:

The DTMF system implemented in lab view is tested for its functionality with different keys as shown in the diagram. One such result is as shown below. Here the key pressed is 6 as it is latched after release and the same key is decoded and given in the digital display. We can also observe the frequency spectrum of the corresponding key pressed.



decoded key

6

Figure 7: Decoding of the corresponding key pressed and the frequency spectrum of the corresponding key

VI. Advantages:

1. Needs only one real coefficient/DTMF frequency

2. Takes little memory space and executes fast

3. Unlike FFT it does not wait for complete data set processes each sample as it arrives

4. With FFT value of N is limited to power of 2 for efficiency

5. In Goertzel algorithm any N may be used

VII. Conclusion:

All the DTMF frequencies are transmitted and decoded using Goertzel algorithm. This algorithm requires less memory space and less calculations. Goertzel algorithm is best suited for the application where only few DFT samples are required.

References:

[1] Emmanuel I feachor, “Digital Signal Processing-A practical approach”, Pearson education, 2004

[2] N. Kehtarnavaz and N. Kim, Digital Signal Processing System-Level Design Using Lab VIEW, Elsevier, 2005.

[3] National Instruments, *Lab VIEW User Manual*, Part Number 320999E-01.

[4] N. Kehtarnavaz and C. Gope, “DSP System Design Using Lab view and Simulink: A Comparative Evaluation,” **Proceedings of ICASSP**, vol. 2, 2006, pp. 985-988.

HYBRID OFFLINE SIGNATURE VERIFICATION

Salma Khatoon , Prof. V.C.Patil
Salma446@gmail.com, patilvc@rediffmail.com,
Ballari Institute of Technology & Management ,BELLARY

Abstract: Signature Verification for skilled and random forgeries. In this model prior to extracting the features, we preprocessed the signatures in the database. Preprocessing consists of i) Normalization ii) Noise reduction iii) Thinning and Skelitazition, for feature set extraction which consists of grid feature, global features such as signature height-to-width ratio (Aspect ratio), Maximum Horizontal Signature verification is the biometrics identification method which is legally accepted and used in many commercial fields such as e-business, access control and so on. In this paper we propose a hybrid Off-line Signature Histogram and Maximum Vertical Histogram, Horizontal Center and Vertical center of the signature, End points of the signature, Signature area, pen lifting feature. The objective of the work is to reduce two vital parameters False Acceptance Rate (FAR)

and False Rejection Rate (FRR) normally used in any signature verification scheme. In and comparative analysis has been made with existing schemes.

Keywords: Offline Signature, Centroid, Histogram, Normalization, Skelitazition, Feature point, Euclidean distance model, FAR,FRR.

I.INTRODUCTION

The use of signatures has been one of the most convenient methods for the identification and verification of human beings. Signatures are particularly usefor identification because each person's signature is highly distinct from

imposter's attempts. Biometric authentication has become a popular

research topic due to its wide application in created by professional imposters or persons who have various fields including prevention of signature fraud in financial transaction.

Generally speaking, signature verification can be divided into two groups: I) On-line II) Off-line.

I)On-line signature verification involves more electronic equipment and it uses signatures captured by pressure-sensitive tablets that extract dynamic properties of a signature in addition to its shape. Dynamic features include the number and order of strokes, the overall speed of the signature, the pen pressure at each point etc. and make the signature more unique and more difficult to forge. On-line signature verification is usually suffers the following two drawbacks a) heavy computational load and b)warping forgeries.

II)Off-line signature verification involves less electronic equipment and the features for off-line verification are much simpler. In this only the pixel image can be evaluated. As compared to on-line signature verification systems, off-line systems are difficult to design as many desirable characteristics such as the order of strokes, the velocity and other dynamic information are not available in the off-line case. The verification process has to wholly rely on the features that can be extracted from the trace of the static signature images only. Although difficult to design, off-line signature verification is crucial for determining the writer identification as most of the financial transactions in present times are still carried out on paper. Therefore, it becomes all the more essential to verify a signature for its authenticity. The design of any offline signature verification system generally requires the solution of five sub problems: data acquisition, pre-processing, feature extraction, comparison process and performance evolution. For achieving this one could either trace or imitate the signature by hard way.

Types of Forgeries

The forgeries involved in handwritten signatures have been categorized on their characteristic features. We have also attempted to classify the various kinds of forgeries into the following types:

Skilled forgery- undoubtly the most difficult of all forgeries is by professional imposters or persons who have experience in copying the signature. Casual forgery-the signer imitates the signature in his own style without any knowledge of the spelling and does not have any prior experience. The imitation is preceded by observing the signature closely for a while.

Random forgery-the signer uses the name of the victim in his own style to create a forgery known as the simple forgery or strokes, the overall speed of the signature, the pen pressure at random forgery. This forgery accounts for the majority of the forgery cases although they are very easy to detect even by the naked eye.

Motivation

The most important advantage of off-line signature verifications that it is easy to implement and efficient in identification and fraud detection. Nowadays it is widely used in routine activities like banking transactions and access control. In our proposed model we consider the Maximum Horizontal and maximum Vertical Histogram, Horizontal and vertical center of the signature, End points of the signature or pen lifting points and signature area gives the crucial points to differentiate between genuine and forgery.

Contribution

Signatures in the database are preprocessed prior to feature extraction for both the reference and testing signatures. In this paper we proposed a model that compares the extracted feature set of genuine signature with testing signature feature set by Euclidean distance method. Validation result depends on the minimum Euclidian distance between reference and testing signature.

Problem definition:

Given test signature and large signature database, the objective is to verify the authenticity of the test signature by comparing with the database using HOSV algorithm.

Organization: This paper is organized into following sections. Section 2 is an overview of related work. The HOSV model describe in Section 3. Section 4 discusses the algorithm for HOSVGF system. Experimental results of the system are in section5 on and Conclusions are contained in Section6.

II. RELATED WORK

Samuel Audit et al., [1] distinguished, Offline signature verification using virtual support vector machines. Offline signature verification is the art of properly classifying between one's real signature and reasonably good forgeries from after -the -fact (scanned or otherwise captured) images. They implemented a technique of signature verification using virtual support vector machines. A testing procedure was devised and the results were analyzed and written. When compared Support Vector Machine classification without virtual signatures, it was found that the use of virtual Support Vector Machine with invariant rotation and translation transformation has a small detrimental effect on the error rate, but reduces the false rejections rate at the expense of the false acceptance rate which increases significantly.

Madasu Hanmandlu et al., [2] distinguished, automatic Verification is a well-established and an active area of research with numerous applications such as bank check verification, ATM access, etc. they proposed a novel approach to the problem of automatic off-line signature verification and forgery modeling that employs the takagi-sugeno model. Signature verification and forged detection are carried out using angle features extracted box approach. Each feature gathered from all samples because of the variations in hand-written signature. Two cases are considered .In the first case the coefficient of the rule are fixed so as to yield a simple form of takagi-sugeno model and in the second case the co-efficient are adopted. In this each rules constituted by a single rule encompassing all feature. In second only a signal rule encompassing all the features is considered. Here again, two cases depending on whether coefficient of the consequent part are fixed or adopted are considered.

Shih-yin Ooi et al., [3] described that the offline signature verification rest on the hypothesis that each writer has similarity among signature samples with small distortion and scale variability. They proposed a novel method to increase the accuracy in biometric matching which was termed as biometric strengthening. They reported 1.1% equal error rate (EER) over the independent database on random forgery, while casual forgery on ERR is 1.2% ,their experiments showed that the biometric strengthening reduces the false acceptance rate (FAR) and false rejection rate(FRR) by increasing the disparity between the features of the two persons which tends to tolerate more interpersonal variance which can reduce the FRR without increasing the probability of false accepts.Therefore They
proposed a novel process called as biometric strengthening to increase the accuracy of the system. Given the robustness of the algorithm and the fact that only concern on global features, optimum results are obtained when the optimum results are obtained when the algorithm is applied to independent database of 1000 signatures from 50 writers and 5 forgers.

Stephane Armand et al., [4] described that the Signatures continue to be an important biometric for authenticating the identity of human beings. They described an effective method to perform off-line signature verification using unique structural features extracted from the signature's contour. A novel combination of the modified direction feature and additional distinguishing features such as the centric, surface area, length and skew are used for classification. A resilient back propagation neural network and a radial basis function network were compared in terms of verification accuracy. Using verification rates of 91.21% and 88.0% were obtained using radial basis function and resilient back propagation respectively. The principle objective was to investigate the efficiency of the enhanced version of the modified direction feature extractor for signature verification.

III MODEL

In this section, grid ,pen lifting and global feature measurement of Horizontal Histogram, vertical Histogram, Horizontal and vertical centre, Edge points of the signature, Signature Area, Aspect ratio, Block diagram, Results are discussed.

Grid Feature

Grid segmentation procedures have been used extensively in the off-line signature verification approach. The skeletonized image is divided into 120 rectangular segments (15x8), and for each segment, the area (the sum of foreground pixels) is calculated. The results are normalized so that the lowest value (for the rectangle with the smallest number of black pixels) would be zero and the highest value (for the rectangle with the highest number of black pixels) would be one. The resulting 96 values form the grid feature vector.

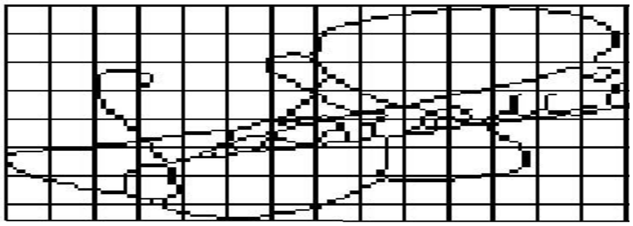


Figure. Simple Grid with Image

Pen lifting point. : points where the pen is lifted, strokes and vertex are calculated. Start vertex and end vertex ,angle, normalization etc . These features will be used to evaluate the similarity cost between online and offline signatures.

Global feature

1. Max Horizontal Histogram and Max vertical Histogram: The horizontal histograms are calculated for each row and the row which by the highest value is taken as max horizontal histogram. The vertical histograms are calculated for each column and the column which has the highest value is taken as max vertical histogram.

2. Horizontal and vertical centre of the signature are calculated using the formula

3. Edge points of the signature: Edge point is the pixel which has only one neighbor, which belongs to the signature, in 8-neighbours.

4. Signature Area the density of signature is obtained by signature area, which is number of signature pixels.
5. Aspect ratio: it is the ratio of Signature height to Signature width. Signature height and width may vary, but aspect ratio of individual will remain approximately same.

Block diagram of HOSV:
Figure 1 gives the block diagram of Hybrid Offline Signature verification system

Training stage

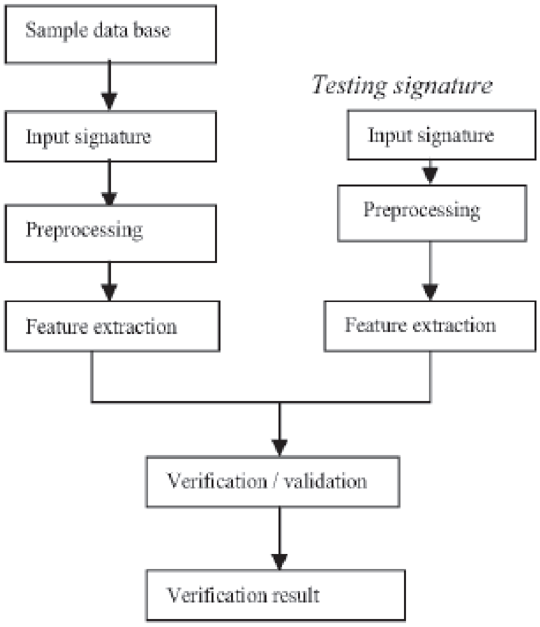


FIGURE 1: Block diagram of HOSV system.

Pre processing:
The pre processing step is applied both in training and testing phases. Signatures are scanned in gray. The purpose in this phase is to make signature standard and ready for feature extraction. The preprocessing stage includes the following.

1. Normalization
2. Noise reduction
3. Skeletonization

1..Normalization
A signer may use an arbitrary base line while writing the signature. We normalize the positional information by calculating an angle θ of corrective rotation about the centroid
(x,y) such that rotating the signature by θ brings it back to a uniform baseline. We calculate θ by maximizing the deviation of data in one direction, example: the x direction. Normalization is
important because it establishes a common ground for image compression.
The mean µx of the x series is calculated by using Equation (3)

$$\mu_{x-}\frac{\sum_{t}^{T} xt}{T} \tag{3}$$

And deviation of x series from 1tx is calculated by using Equation(4)

$$\sigma_{x-}\sqrt{\frac{\sum_{t}^{T}(x_t-\mu_x)^2}{T}}$$

In order to maximize the deviation, we need only to maximize the Equation (5)

$$\sum_{t}^{T}\left(x_t^{*}-\mu_x\right)^2 \tag{5}$$

Where x * indicates a rotated x value. A rotation about a point (in this case we choose to rotate about the co ordinate ($\mu_x.\mu_y$) as this preserves the centroid) This can be expressed as Equation(6)

$$x_t^{*}=(x_t-\mu_x)\cos(\theta)+(y_t-\mu_y)\sin(\theta)+\mu_x \tag{6}$$

We now define function f (0) by substituting the Equation (6)in Equation (5).

$$f(\theta)=\sum\left[(x_t-\mu_x)\cos(\theta)+(y_t-\mu_y)\sin(\theta)+\mu_x-\mu_x\right]^2 \quad \ldots \tag{7}$$

$$=\sum a_t^2\cos^2(\theta)+2a_tb_t\cos(\theta)\sin(\theta)+b_t^2\sin^2(\theta)$$

$$=\cos^2(\theta)\sum_{t}^{T}a_t^2+2\cos(\theta)\sin(\theta)\sum_{t}^{T}a_tb_t+\sin^2(\theta)\sum_{t}^{T}b_t^2$$

$$=\cos^2(\theta)p+2\cos(\theta)\sin(\theta)Q+\sin^2(\theta)R \tag{8}$$

Where $a_t = x_t - \mu_x$ and b_t, $=y_t - \mu_y$.

Taking the derivative of Equation (8) yields Equation(9).

$$f'(\theta)=2Q\cos^2(\theta)-2P\cos(\theta)\sin(\theta)$$
$$+2R\cos(\theta)\sin(\theta)-2Q\sin^2(\theta) \tag{9}$$

The roots obtained for Equation (9) are expressed as Equation (10) and Equation(11)

$$\pm\cos^{-1}\left(\pm\frac{1}{\sqrt{2}}\sqrt{\frac{1+(P-R)}{\sqrt{P^2+4Q^2-PR+R^2}}}\right) \tag{10}$$

and

$$\pm\cos^{-1}\left(\pm\frac{1}{\sqrt{2}}\sqrt{\frac{1+(R-P)}{\sqrt{P^2+4Q^2-PR+R^2}}}\right) \tag{11}$$

We adopt the smallest value for θ from Equations (10) and Equation (11) which will result in a maximum value of f(θ).

2. Noise reduction:

Dirt on cameras or scanner lens, imperfection in the scanner lighting etc introduces the noise in the scanned signature images. A filtering function is used to remove the noises in the image, which works like a majority function that replaces each pixel by its majority function. A noise reduction filter is applied to the. binary scanned image. The goal is to eliminate single white pixel on black background and single black pixel on white background.

3.Skeletonizaton:

Skeletonizaton makes the extracted features invariant to image Characteristics like quality of pen and paper. A simplified version of a skelitization technique is described as follows Step 1: Mark all the points of signature that should be removed (black pixel that have at least 8 neighbor and at least 2 black 8 neighbor pixel) Step2: The marked points are examined for removal along the Contour lines of the signature image, and remove these as there removal will not cause a break in the resulting pattern Step3: If at least one point being deleted, the above steps are repeated until the Skeletalized image is obtained.

*Classifie*r:

All the extracted features are used by a classifier that compares the extracted feature with the number of prototype with known identity. Euclidian-distance based K-nearest neighbor classifier is used. This classifier determines all nearest neighbors to each input feature vector and opts for the class that is most often represented. In case of tie, the class with the smallest sum of distances chosen. The Euclidian-distance(Ed) is measured using Equation(12)

$$Ed = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \qquad (12)$$

Comparison:
The test signature extracted features are compared with guanine Signature extracted features based on measured Euclidian-distance of both test and guanine signatures.

IV. ALGORITHM: HOSV SYSTEM

Table 1 gives algorithm for HOSV system in which reference set of signatures features are compared with features of testing signatures using the Euclidean distance.

Table 1:HOSV algorithm

```
Input: Test Signature , Reference signature from database
Output: Verified signature.
    i.  Pre Processing the Reference
            (a).Normalization
            (b). Noise reduction
            (c).Skelitization
    ii. Grid feature
        Pen lifting feature
        Global feature Extraction
            (a). Max Horizontal Histogram and Max vertical Histogram
            (b).Horizontal and vertical center of the signature are calculated using the formula.
            (c) Edge point number of the signature
            (d) Signature Area
            (e)Aspect Ratio
            iii. Steps from 1 to 2 are repeated for testing signature
        iv. Comparison using Euclidian distance method.
```

V. EXPERIMENT

For experiment 21 persons are considered, and for each person 15 genuine signatures are taken at different timing and 10 skilled forgery samples are taken for each person. The data base has 315 genuine samples and 210 skilled forgery samples. For each sample features are extracted and based on Euclidian-distance validation done. The performance of the proposed system is given in terms of type-I (False rejection of genuine signature)and type-Il error(False acceptance of forge signature)

Type I error: False rejection of genuine signature is the ratio of number of genuine signature rejected and total number of genuine signature

Type II error: False acceptance of forge signature is the ratio of number of forged signature accepted and total number of forged signature

Experimental results

TABLE 2:Results of HOSV(HYBRID OFFLINE SIGNATURE VERFICATION)

Features used	**Type-I**	**Type-II**
Global	6.4%	4.6%

TABLE 3: COMPARISION OF ERRORS OF SKILLED FORGERIES OF FEMOSV AND HOSV

Errors	**FEMOSV**	**HOSV**
Type-I	11.26%	6.4%
Type-II	13.66%	4.6%

VI. CONCLUTION

We propose a robust off line signature verification system in which we consider the global features like Maximum Horizontal Histogram, Maximum vertical Histogram, Horizontal and vertical center of the signature, Edge point of signature, Signature area and Aspect ratio, the performance of the proposed system is given in terms of Type-I and Type-Il error which are better compared to the existing systems.

REFERENCE

[1] Samuel Audet, Peyush Bansal, Shirish Baskaran, "Offline Signature Verification Using Virtual Support Vector Machines,"

[2] Madasu Hanmandlu, Mohd Hafizuddin, Mohd. Yusof, Vamsi Krishna Madasu."Off-line Signature Verification And Forgery Detection Using Fuzzy Modeling," Australian
Conference on Artificial Intelligence 2003,pp. 1003-1013

[3] Shih-Yin Ooi, Andrew Beng-Jin Toeh and Thian-Song Ong. "Offline Verification Through Biometric Strengthening," workshop on Automatic Identification Advanced Technologies, pp.226-231, June 2007

[4] Stephane Armand, Michael Blumenstein and Vallipuram Muthukkumaraswamy "Offline signature verificationusing the Enhanced modified Direction Feature and Neuralbased classification,"International conforance on Neural network, pp.684-691, October 2006

[5] G. Rigoll, A. Kosmala " A Systematic comparision between on-line and off-line methods for Signature Verification with hidden markov models," Fourteenth International conformance on pattern recognition,' Vol.2,' pp.1755-1757, August 1998 [6] Edson J R Justino1, Flavio Bortolozzi1, Roert Sabourin1'2 signature "Off-line Signature verification Using HMM for Random, Simple and Skilled Forgeries, Sixth international conformance on document analysis and recognition, pp.1031-1034, September 2001

[7] V. Di Lecce, G. Dimauro, A. Guerriero, S. Imperdovo, G Pirlo, A. Salzo, L. Sarcinella. "Selection Of Reference Signatures For Automatic Signature Verification." Fifth International conformance on document Analysis and Recognition, pp.597-600, September 1997 [8] Ibrahim S. I. Abuhaiba. "Offline Signature Verification Using Graph Matching," Fifteenth international Conformance on pattern Recognition, vol.2, pp.851-854.

[9] Kai Huang and Hong Yan. "Signature Verification using Fractal Transformation," fifteenth International conformance on pattern Recognition, Vol.2, pp.851-854, September 2000

[10]Ramanujan Kashl, Winston Nelson " Signature Verification: Benefits of multiple tries," proceeding of performance of the Eighth International workshop on Frontiers in Handwriting Recognition, pp.341-356, August 2002

[11] Javed Ahamed MaharI, prof. Dr. Mumtaz Hussain Mahar', Muhammad Khalid Khan2" Comparative Study of Feature Extraction Methods with K-NN for Off-Line Signature verification,' Second international conformance on Vector emerging Technologies, pp.115-120, November 2006

[12]Sudarshan Madabusi, vivek Srinivas, sudharsan Bhaskaran, Muthukumaran Balasubramanian "On-line and off-line Signature Verification Using Relative Slope Algorithm." Procedings of the 2005 IEEE Intenational workshop on measument system for home land security, contraband detection and personal safety, pp.1 1-15, March 2005

Compression of Speech Signals using Cosine PacketTransform

G.Vijaya Durga[*],K.Veera Swamy[#] ,Y.V.Bhaskar Reddy[$]

[*]P.G Student,ECE Dpt.,QIS College of Engineering and Technology

[#]Professor & Pricipal, QIS College of Engineering and Technology

[$]Associate Professor in ECE Dpt.,QIS College of Engineering and Technology

email: pvdurga40@gmail.com mobile : 9247581274

Abstract – In this paper a new algorithm for compressing the speech is presented.This is an adaptive algorithm which reduces computatational complexity of a system by using packet decomposition In this paper different speech signals are taken and compression ratio of methods using Wavelet Transform, Discrete Cosine transform, Wavelet Packet Transform and proposed adaptive algorithm using Cosine Packet Transform is calculated The mean compression ratio is calculated for all the methods and compared. The implemented results show that the proposed compression algorithm gives the better performance for speech signals.

Keywords: Discrete Cosine Transform, Discrete Wavelet Transform, Wavelet Packet Transform, Cosine Packets, adaptive thresholding.

I.INTRODUCTION

Speech signals has unique properties that differ from a general audio/music signals. First, speech is a signal that is more structured and band-limited around 4 kHz. These two facts can be exploited through different models and approaches and at the end, make it easier to compress. Today applications of speech compression involve real time processing in mobile satellite communications, cellular telephony, internet telephony, audio for videophones or video teleconferencing systems, among others. Other applications include also storage and synthesis systems used, for example in voice mail systems, voice memo wristwatches, voice logging recorders and interactive PC software[1]. The idea of speech compression is to compress speech signal to take up less storage space and less bandwidth for transmission. To meet this goal different methods for compression have been designed and developed by various researchers [2-7]. The speech compression is used in digital telephony, in multimedia and in the security of digital communications. Before the introduction of Packet based transform techniques, audio coding techniques used DFT and DCT with window functions such as rectangular and sine-taper functions. However, these early coding techniques have failed to fulfil the contradictory requirements imposed by high-quality audio coding. For example, with a rectangular window the analysis/synthesis system is critically sampled, the overall number of the transformed domain samples is equal to the number of time domain samples, but the system suffers from poor frequency resolution and block effects, which are introduced after quantization or other manipulation in the frequency domain. Overlapped windows allow for better frequency response functions but carry the penalty of additional values in the frequency domain, thus not critically sampled. Discrete Cosine Packet Transform is currently the best solution, which has satisfactorily solved the paradox.

Speech compressions are done by either based on linear prediction or based on orthogonal transforms methods. On the basis of the classical papers written by Shannon, [8] and Kolmogorov, [9], recently was highlighted a strong connection between the systems proposed in many lossy compression standards and the harmonic analysis, [10]. All these systems use orthogonal transforms. The algorithm described in this paper belongs to the second category. Unfortunately there is no any fast algorithm for the computation of orthogonal transform This is the reason why in practice other orthogonal transforms are used. The quality of compression system can be appreciated with the aid of his rate dist distortion function. A compression system is better than another if, at equal distortions, it realizes a higher compression rate. The maximization of compression rate can be done, if a good selection of orthogonal transform be made.

This paper is organized as follows. The mathematical model for speech signal and the description about Discrete Cosine Transform is presented in Section II. With necessary mathematical modeling, the proposed adaptive algorithm for speech compression is explained in Section III. In section IV the developed algorithm is tested for various speech signal samples and comparison is made with Wavelet Transform Cosine Transform and Wavelet Packet Transform. Finally section V concludes the paper with some discussions

II. Mathematical Model

Mathematical model of speech signal

Every spoken word is a sequence of tons with different intensities, frequencies and duration. Every ton is a sinusoidal signal with a certain amplitude, frequency and duration Therefore it is possible to represent any speech signal in to a sinusoidal model. A mathematical description of this model is given by

$$x(t)=\Sigma_{i=1}^{Q(t)} Ai\ \cos\theta_i(t) \quad (1)$$

Where Ai, θ_i ,t are amplitude, frequency and time duration of the particular incident respectively. Every term of this sum is a signal with double modulation So this is not a stationary signal. But frequently the speech is regarded like a sequence of stationary signals. Dividing the speech signal into a sequence of stationary signals, each of them having duration inferior to 25 ms, a sequence of stationary signals is obtained. On each segment the speech model can be of the form

$$Xs(t)= \Sigma_{i=1}^{n} Ai\ \cos\omega_i(t) \quad (2)$$

This decomposition is very similar with the decomposition of the signal xs(t) into a cosine packet The energy of the signal xs(t) can be computed using the following relation

$$Ex= \Sigma_{i=1}^{n} | Ai |^2 \quad (3)$$

The Discrete Cosine Transform

The most common DCT [11] definition of a 1-D sequence of length N is

$$Ci=\alpha_i =\Sigma_{i=1}^{n-1} f(x) \cos[\Pi(2x+1)i/2N] \quad (4)$$

for i = 0,1,2,…,N-1.

Where $\alpha_i= 1/\sqrt{2}$ for i=0

$\alpha_i=2/\sqrt{N}$ for $i\neq0$

it is clear from (1) that for i=0

$$C(i=1)=(\sqrt{1}/N)\ \Sigma_{x=1}^{n-1} f(x)$$

Thus, the first transform coefficient is the average value of the sample sequence. In literature, this value is referred to as the DC Coefficient. All other transform coefficients are called the AC Coefficients

III.PROPOSED ALGORITHM

The proposed adaptive algorithm for speech compression using Cosine Packet Transform is shown in Fig 1. The speech signal to be compressed is converted in to packets with finite duration. The Discrete Cosine Transform is applied to each packet and transformed coefficients are computed. The coefficients are extracted and fed into the adaptive threshold detector to nullify the inferior coefficient for better compression.

Selection of best packets

The main reason to choose the Packet Cosine transform is cost functional used for the best packet. This transform is an adaptive one. The result of its utilization in a given application can be optimize using the best packet selection procedure. This is a very efficient procedure which is able to enhance very much quality of a given signal processing method.The most used is the entropy but its utilization do not realizes the maximization of the compression rate. The optimal cost functional for compression is that realizing the minimization of the number of coefficients superior to a given threshold t,C_i using the cost functional C_i coefficients superior to the threshold t are obtained.This is a minimal number because it was obtained using the appropriate cost functional for the selection of the best packet. Increasing the threshold value t, the number C_i decreases and the compression rate increases Hence, the threshold detector must be an adaptive one. Another parameter of the DCPT who must be considered for the optimization of the compression is its number of iterations

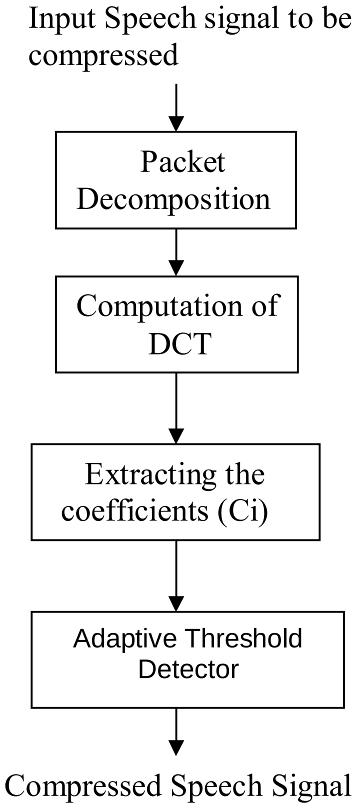

Input Speech signal to be compressed
Packet Decomposition
Computation of DCT
Extracting the coefficients (Ci)
Adaptive Threshold Detector
Compressed Speech Signal

Fig 1. Flow diagram for the proposed adaptive algorithm

Adaptive Threshold Detector

One of the most important processes of the proposed compression algorithm is the threshold detector. The main role of this process is to nullify all the coefficients obtained from the Cosine Packet Transform smaller to a threshold value. This is in fact the compression mechanism. This process is an adaptive system, which automatically choose the threshold value

Let us assume that the distortion parameter of a compression system is a ,a<1, N is the number of samples of signal to be processed and Ex is the energy of the input speech signal, then the threshold value is defined as

$T=\sqrt{a}Ex/N$ (7)

The constant a can be related with the signal to noise ratio of the input signal x(t) and is defined as

$b=-10\log_{10}a$ (8)

From the above equation a is given by

$a_n=10^{-b/10}$ (9)

where a_n is nothing but lower bound of a . Using eqns (7) & (9) the lower bound value for the threshold can be obtained as

$t_n = \sqrt{10}^{\,b/10}Ex/N$

For the threshold a value t_n, superior to an output signal to noise ratio superior to b will be obtained. Unfortunately the exact value of Ex will not be known a priori. This is the reason why an adaptive algorithm for the election of the threshold value is recommended. The flow diagram of adaptive threshold detector is shown in Fig 2. The energy of the input signal to be compressed is computed and the value b is initialized. The threshold value is calculated using eqn 10. The threshold value is increased starting from this value. At every iteration the value Ex is computed. If this value is higher than b then the extracted coefficient is compared with threshold value t. If it is less then the threshold value then the corresponding coefficients is replaced with zero value otherwise the coefficients value is maintained the same. The proposed adaptive algorithm is stopped when for the first time the value Ex becomes smaller than b. .

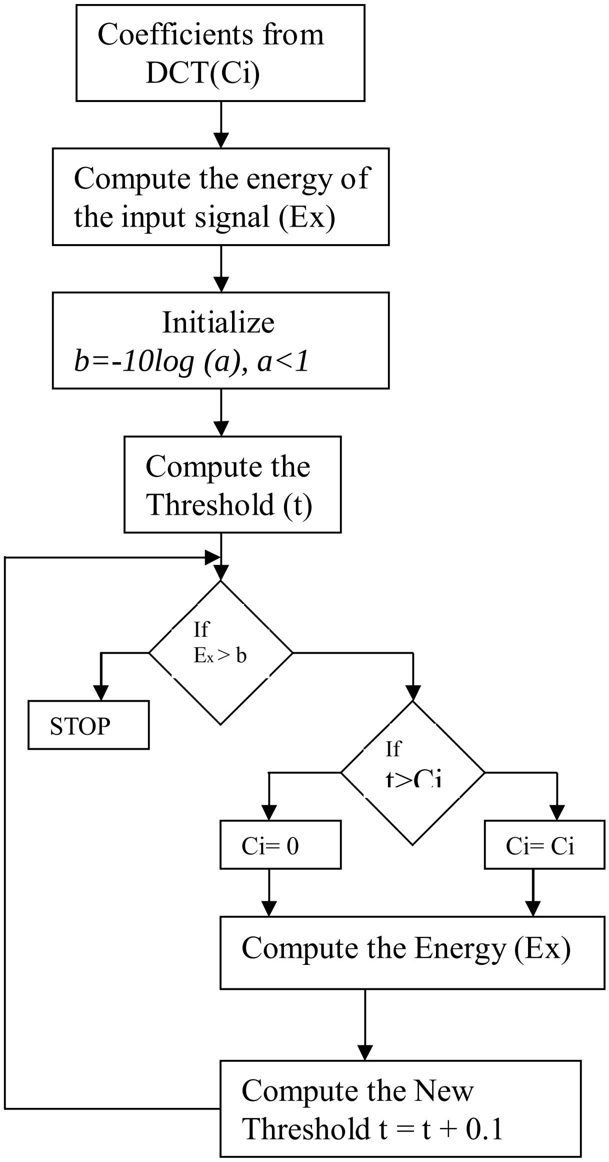

Coefficients from DCT(Ci)
Compute the energy of the input signal (Ex)
Initialize
b=-10log (a), a<1
Compute the Threshold (t)
If $E_x > b$
STOP
If t>Ci
Ci= 0
Ci= Ci
Compute the Energy (Ex)
Compute the New Threshold t = t + 0.1

Fig 2. Flow diagram for the adaptive threshold detector

IV. SIMULATION RESULTS AND DISCUSSION

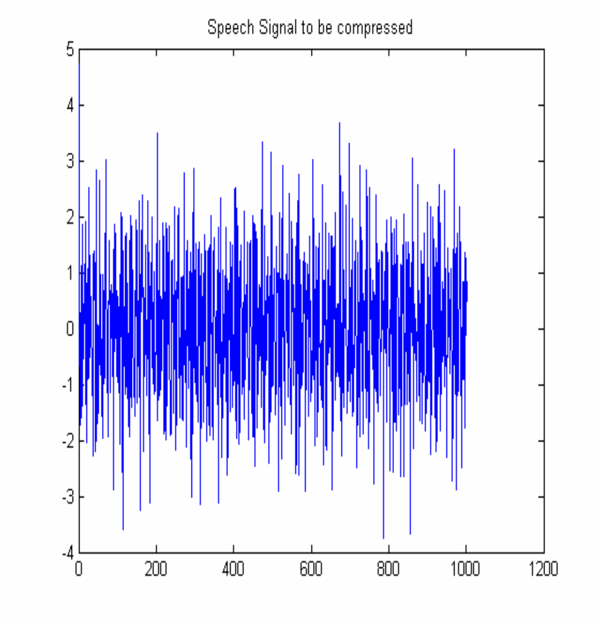
The various speech signal sample is simulated using MATLAB . The generated speech sample is shown in fig. 3.

The generated speech signal is segmented in to 15 packets with 512 samples (the duration of each block being inferior to 25 ms) per packet

The Discrete Cosine Transform is

Fig 3. Speech Signal Sample

computed for each packets using eqn 4. The transformed coefficients are extracted for further processing. The energy of the input signal is computed and the threshold value is calculated using eqn 10. The value of input energy is compared with b. If it is higher then each and every transformed coefficient value is compared with threshold value. The inferior coefficients are nullified The new energy of the signal is calculated and compared with b. If energy is lower than b the above process is repeated for new threshold value otherwise the compression process is stopped.



For 10 different speech signals, compression is performed using Discrete Cosine Transform, Discrete Wavelet Transform Wavelet Packet Transform and the proposed adaptive algorithm. The compression ratios achieved through these methods are tabulated for various speech signal sample.

TABLE I COMPARISION OF COMPRESSION RATIO

Speech Signal Sample	DWT	DCT	WPT	Proposed Adaptive algorithm
6.0825	8.9413	12.1784	12.1163	6.0825
32.6634	47.8547	65.4811	65.636	32.6634
32.7957	46.9867	65.6898	65.4774	32.7957
29.7397	43.1513	59.6003	60.4	29.7397
25.8621	37.5097	51.8826	50.1416	25.8621
26.8487	39.7978	53.7361	55.0038	26.8487
35.6182	37.7567	51.2569	52.4244	35.6182
23.8238	34.7395	47.6895	48.7892	23.8238
30.3963	44.4662	60.9127	60.503	30.3963
30.9598	44.8917	62.027	63.3549	30.9598

The Table I shows the comparison of compression ratio for various methods. Analyzing the Table, the good performance of the proposed adaptive algorithm can be observed. The proposed algorithm gives the better compression ratio for most he the speech samples. The comparison of compression ratio for speech signal sample from 1 to 10 shown in Fig 4.

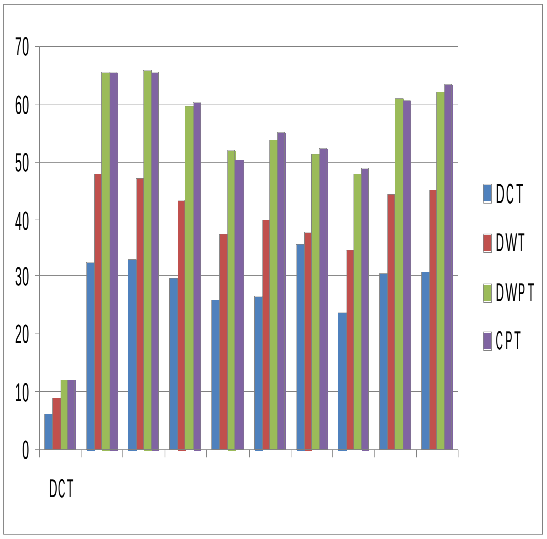


Fig 4. Comparison of Compression ratio (Speech signal sample 1-10)

V. CONCLUSION

A new compression method based on adaptive threshold detector is proposed and tested. The simulated results show that the proposed algorithm gives the better compression ratio as compared with other methods. Using this method, a mean compression rate of 28.275, was obtained in the simulation report. This value is superior to mean compression rate, of other methods. Using fast DCT algorithm, the proposed method can be implemented on a Digital Signal Processor. The proposed system is a good alternative to the speech compression systems based on the linear prediction approaches.

REFERENCES

[1]. R. W. Yeung, “A First Course in Information Theory,” New York: *Kluwer Academic/Plenum Publishers*, 2002.

[2]. A.Gersho, “Advances in Speech and Video Compressions,” *Proceedings of the IEEE*, vol. 82, pp. 900-918, June 1994.

[3]. J.L.Flanagaran, M.R.Schroeder, B.S.Atal, R.E.Crocherie, N.S.Jayant and J.M.Tribolet, “Speech Coding,” *IEEE Transactions on Communications*, vol. 27, pp.710-737, April 1979.

[4]. P.Noll, “Wideband Speech and Audio Coding,” *IEEE Communications Magazine,* pp. 34-44, Nov. 1993.

[5]. K. Sayood and J. C. Borkenhagen, “Use of residual redundancy in the design of joint source/channel coders,” *IEEE Transactions on Communications*, 39(6):838-846, June 1991.

[6]. Edler, B., “Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions,” (in German), *Frequenz,* vol.43, pp.252-256, 1989.

[7]. Q. Memon, T. Kasparis, “Transform Coding of Signals Using Approximate Trigonometric Expansions”. *Journal of Electronic Imaging*, Vol. 6, No. 4, October 1997, pp. 494-503.

[8]. C. E. Shannon, .A mathematical theory of communications,. *Bell System Technical Journal*, vol. 27, pp. 379.423, 623.656, 1948.

[9]. A. N. Kolmogorov, .On the Shannon theory of information transmission in the case of continuous signals,. *Trans. IRE*, vol. IT-2, pp. 102.108, 1956.

[10]. D. L. Donoho, M. Vetterli, R. A. Devore, and I. Daubechies, .Data compression and harmonic analysis,. *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2435.2476, 1998.

[11]. N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. C-32, pp. 90-93, Jan. 1974.

ENCRYPTION AND DECRIPTION USING SCAN METHOD

Sumith.K.S[1], Arjun.H[2]
[1] student e & c dept. Kalpataru Institute of Technology, Email:sumith.say@gmail.com
[2] student e & c dept. Kalpataru Institute of Technology,Email:harjun6@gmail.com

ABSTRACT

This paper proposed an image encryption and decryption process. Its encryption method is based on SCAN patterns generated by the SCAN methodology.

The SCAN is a language-based two-dimensional Spatial -accessing methodology which can efficiently specify and generate a wide range of scanning paths. Then scanning paths sequence fill in original image. Note that the scanning paths with random code generating procedure, which produces the encryption keys in a very many ways; so come to the quite secret system. This paper presents a brief overview of SCAN, encryption and decryption, and test results of the methodology.

1. INTRODUCTION

With the ever-increasing growth of multimedia applications, security is an important issue in communication and storage of images, and encryption is one the ways to ensure security. Image encryption has applications in inter-net communication, multimedia systems, medical imaging, telemedicine, and military communication. There already exist several image encryption methods. They include SCAN-based methods, chaos-based methods, tree structure-based methods, and other miscellaneous methods. However, each of them has its strength and weakness in terms of security level, speed, and resulting stream size metrics. We hence proposed the new encryption method to overcome these problems.

The proposed image encryption method is based on rearrangement of the pixels of the image. The rearrangement is done by scan patterns that generated by the SCAN methodology. The scanning path of the image is a random code form, and by specifying the pixels sequence along the scanning path. Note that scanning path of an image is simply an order in which each pixel of the image is accessed exactly once. Such the encryption also involves the specification of set secret scanning paths. Therefore, the encryption needs a methodology to specify and generate a larger number of wide varieties of scanning paths effectively.

2. ENCRYPTION

An original message is known as the plain text, while the coded message is called cipher text. The process of converting plain text to cipher text is known as *enciphering* or *encryption*. Restoring the plain text from the cipher text is *deciphering* or *decryption*.

A symmetric encryption scheme has five ingredients.

Plain text: This is the original intelligible message or data that is fed into the algorithm as input.Encryption algorithm: The encryption algorithm performs various substitutions and transformations on the plain text.Secret key: the secret key is also input to the encryption algorithm. The key is a value independent of the plaintext. The algorithm will produce a different output depending on the specific key being used at the time. The exact substitutions and transformations performed by the algorithm depend on the key.Cipher text: This is scrambled message produced as output. It depends on the plain text and the secret key. For a given message, two different keys will produce two different cipher texts. The cipher text is an apparently random stream of data and it stands, is unintelligible. Decryption algorithm: This is essentially the encryption algorithm run in reverse. It takes the cipher text and secret key and produces the original text.
Encryption provides a foundation upon which communications security may be built. Increasing reliance upon computer networks for the support of financial, industrial, transportation, and other systems, coupled with growing concerns for data integrity and security, has been driving forces behind the current civilian interest in data encryption. Interestingly, encryption techniques have been used for thousands of years to protect information, yet research is still ongoing in the field. Encryption may be used to achieve the following

- To *prevent* unauthorized access to transmitted or stored data(disclosure)
- To *prevent* the analysis of data traffic(disclosure)
- To *detect* any modification of the data stream (including data destruction)
- To *detect* the denial of transmission service
- To *detect* unauthorized connections (authentication)

Within a computer communications environment, encryption techniques provide mechanisms for authentication as well as secure communications.

3. TYPES OF ENCRYPTION

i. Data encryption:

The process of converting plain data to cipher data is known as data encryption

ii. Audio encryption:1

In many applications, the audio sequences need confidentiality during transmission. Sometimes it is enough to apply the naive approach, but in many instances this is too computationally expensive (for example, in small mobile devices). As far as the security is concerned, perhaps the most important type of audio data is speech. Unlike in the case of music files and similar entertainment audio sequences, in many applications speech requires substantial level of security

iii. Video and image encryption:

Information security is becoming more and more important with the progress in the exchange of data for electronic commerce. The image encryption is the most direct way to protect digital images, due to the redundancy and correlation of the image information, image encryption is different from traditional cryptology.

During the last decade, the use of computer networks has grown spectacularly, and this growth continues unabated. Almost all networks are being installed, interconnected, and connected to the global internet. Today more and more information has been transmitted over the internet. The information is not only text, but also audio, image, and other multimedia. Images have been widely used in our daily life. However, the more extensively we use the images, the more important their security will be. For example, it is important to protect the diagrams of army emplacements, the diagrams of bank building construction, and the important data captured by military satellites.

4. SCAN LANGUAGE

The "SCAN" language includes an alphabet consisting of primitive scanning techniques, as, letters, and a simple grammar to manipulate and combine the alphabet symbols by generating new scanning patterns (words) from simple ones.

The development of SCAN provides an efficient approach to the problem of modeling and generating all accessing algorithmic patterns of an image of NxN. There are two concepts underlying and motivating the specification of the SCAN language:

The concept of hierarchical decomposition of the image into square regions

The concept of recursive composition of two or more basic accessing techniques in such a way, so that to expand uniformly over the image levels SCAN words convey information about both the

decomposition to be applied and type of basic scanning techniques to be composed.

The SCAN is a formal language-based two-dimensional spatial accessing methodology which can represent and generate a large number of wide varieties of scanning paths. The SCAN is a family of formal languages such as Simple SCAN, Extended SCAN, and Generalized SCAN, each of which can represent and generate a specific set of scanning paths. Each SCAN language is defined by a grammar and each language has a set of basic scan patterns, a set of transformation of scan patterns and a set of rules to recursively compose simple scan patterns to obtain complex scan patterns. Note that this set of basic scan patterns can be extended or reduced as needed by a specific application. There are 6 transformations of scan patterns. They are identity, horizontal reflection, and vertical reflection, rotation by 90, 180, and 270°, and compositions of these transformations. The rules for building complex scan patterns form simple scan patterns are specified by the production rules of the grammar of each specific language. The encryption specific SCAN language uses four basic scan patterns. They are continuous raster C, continuous diagonal D, continuous orthogonal O, and spiral S.

Each basic pattern has eight transformations numbered from 0 to 7. For each basic scan pattern, the transformations 1, 3, 5, 7 are reverses of transformations 0, 2, 4, 6, respectively. The basic scan transformations are shown in Figure 4.1. Since most images require different scanning indifferent sub regions, the encryption specific SCAN language allows an image region to be recursively partitioned into four sub regions, and each sub region to be scanned independently. When an image region is partitioned, the order in which the four sub regions are scanned is specified by a partition pattern. The partition patterns are letter B, letter Z, and letter X, each of which has eight transformations.

The SCAN alphabet consists of 15 SCAN items and is by the set ∑:

∑= {r,c,d,e,a,i,o,l,h,y,w,z,b,x,s}

The symbols of ∑ are called SCAN *letters,* and correspond to scan orders (algorithms) which are illustrated in figure 4.3(for an image of size 4*4).For a complete specification of the underlying scan algorithms. SCAN syntax is simple. Statements of SCAN are simple composed by SCAN terms via a unique operator. Each SCAN term represents a basic scanning technique adapted to an image size. SCAN sentences are of the form

$w_t= L_1n_1\# L_2 n_2\#.......... L_t n_t$

Where, L_i є∑, n_i=power of 2, $n_1n_2........ n_t$=n.

The terms L_in_i in a SCAN word are separated by the punctuation mark "#", which also plays the role of the connection operator between the SCAN items. In the semantic level, the combination of the fundamental algorithms represented by the SCAN letters is performed in a hierarchical fashion to provide composite algorithmic accessing patterns, which are called SCAN *patterns.*

Following by basic scan patterns and partition patterns to produce concept, we use a random code generating produce the SCAN word and to define encryption key. The SCAN word contain scan and partition patterns. The scan partition word hasc0~c7,d0~d7, o0~o7, s0~s7, r0~r7, a0~a7, e0~e7, y0~y7,w0~w7,b0~b7,z0~z7,x0~x7. The partition word has B0~B7, Z0~Z7 and X0~X7.

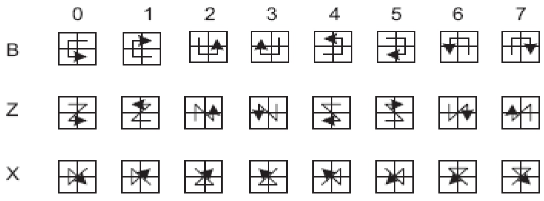


Figure 4.1 Partition Patterns

The basic idea of the proposed encryption method is to rearrange the pixels of the image and change the pixel values. The rearrangement is done by a set of scanning patterns (encryption keys) generated by an encryption-specific SCAN language which is formally defined by the grammar $G = (\Gamma, \Sigma, A, \Pi)$ Where non-terminal symbols $\Gamma = \{A, S, P, U, V, T\}$, Terminal symbols Σ = {c, d, o, s, r, a, e, m, y, w, b, z, x; B, Z, X, ( ), space, 0, 1, 2, 3, 4, 5, 6, 7},
Start symbol is A, and Production rules is given by
A → S | P,
S → UT,
P → VT (A A A A),
U → c | d | o | s | r | a | e | a | i | l | h | y | w | b | z | x
V → B | Z | X
T → 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7

The semantics of this encryption-specific SCAN language is described next.

(a) A → S | P means process the region by scan S or partition P.

(b) S → UT means scan the region with scan pattern U and transformation T.

(c) P → VT(A A A A) means partition the region with partition pattern V and transformation T, and process each of the four sub regions in partition order using As from left to right.

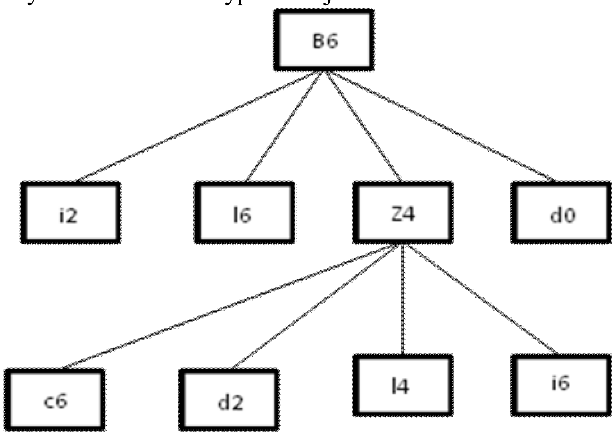
(d) U → c | d | o | s | r | a | e | i | l | h | y | w | b | z | x means scan with continuous raster or diagonal or continuous orthogonal or spiral out or raster or right orthogonal or diagonal parallel or horizontal symmetry or diagonal symmetry or diagonal secondary or block or zeta or xi, respectively.

(e) V → B | Z | X means partition with letter B or letter Z or letter X , respectively.

(f) T → 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 means use one of the eight transformation with scan or partition. For partition, these transformations are shown in Figure 4.1. For scan, these transformations are defined as follows. For all scan patterns, 0 means identity transformation,2 means 90° clockwise rotation. For scan patterns c, o, s, a, e, m, y, w, b and x, 4 means 180. Clockwise rotation and 6 means 270.clockwise rotation. For scan patterns r and z, 4 means vertical reflection and 6 means vertical reflection followed by 90° clockwise rotation. For scan pattern d, 4 means 90° clockwise rotation followed by horizontal reflection and 6 means 180° clockwise rotation followed by vertical reflection. For all scan patterns, 1, 3, 5, and 7 are reverses of scanning paths specified by 0, 2, 4, and 6, respectively.

As an example, consider the scan key B6 (s3 o4 Z4 (c7 d0 o1s7) d2) for a 16 × 16 image. The scanning path which corresponds to this scan key is shown in Figure B. The image is first partitioned into four sub regions using B6 partition order. These four sub regions are scanned using s3, o4 and Z4 (c7 d0 o1s7) d2. The second sub region is further partitioned into four sub regions using Z0 partition order and these four sub regions are scanned using c7 d0 o1and s7.

The scanning path is corresponding to the SCAN word constructed as follows. The SCAN word defines encryption key can achieve encryption objective.



Encryption keys is B6 (i2 l6 Z4(c6 d2 l4 i6) d0)

Figure 4.2.The SCAN word diagram

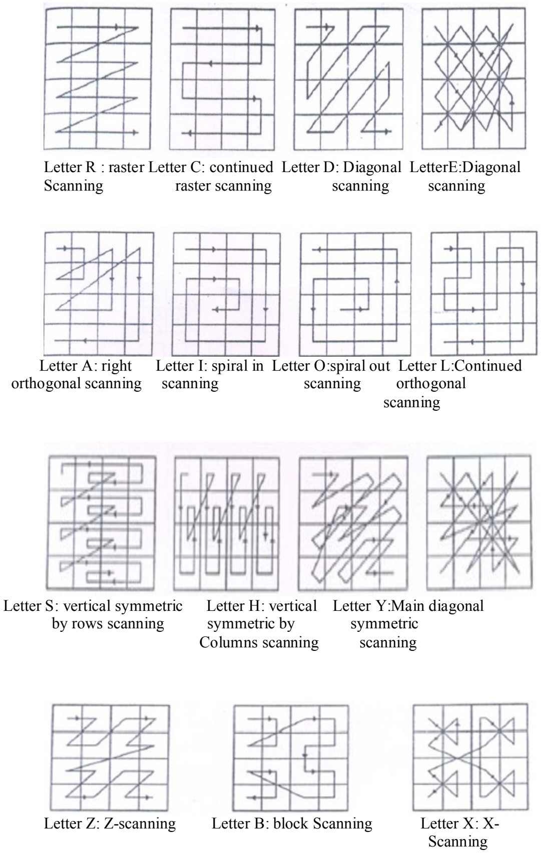

Letter R : raster Scanning
Letter C: continued raster scanning
Letter D: Diagonal scanning
LetterE:Diagonal scanning
Letter A: right orthogonal scanning
Letter I: spiral in scanning
Letter O:spiral out scanning
Letter L:Continued orthogonal scanning
Letter S: vertical symmetric by rows scanning
Letter H: vertical symmetric by Columns scanning
Letter Y:Main diagonal symmetric scanning
Letter Z: Z-scanning
Letter B: block Scanning
Letter X: X-Scanning

Figure 4.3 Graphical representations of the Scan Letters

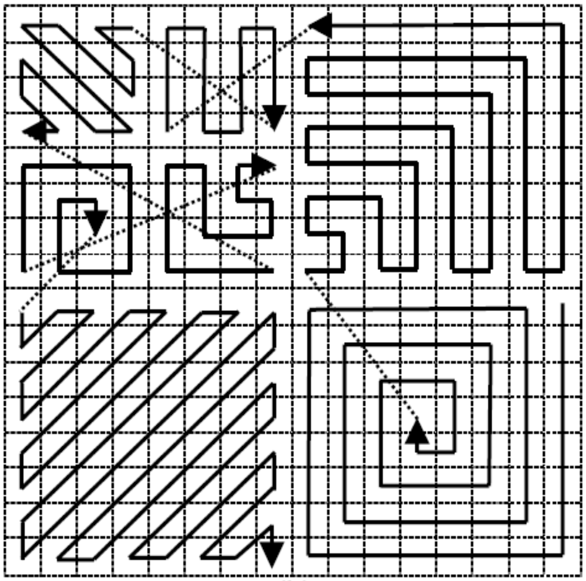
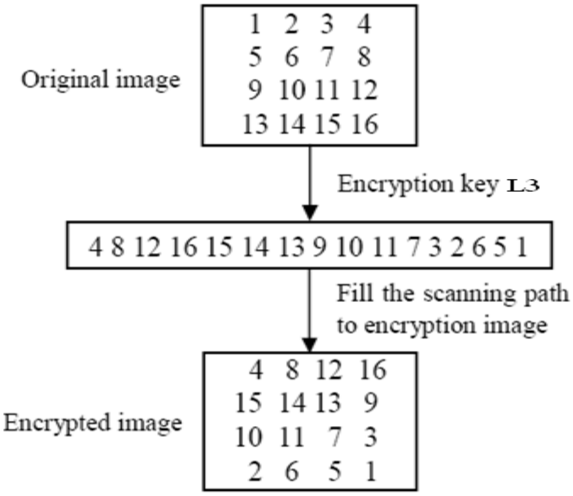


Figure4.4 The scanning path for 16 X 16 size image using the key B6 (i2 l6 Z4(c6 d2 l4 i6) d0)

5. ILLUSTRATION OF ENCRYPTION.

For a given 2D 2^n X 2^n, 2 ≥n ≥9 images, the encryption algorithm transforms it into one dimensional strings of length 2^{2n} firstly. Then each arrangement strings of length are encrypted using random generating encryption keys. Figure 5.1 illustrates how we encrypt a 4 x 4 image. Transform one –dimensional string of length 16 using an encryption key. Then wide string data ccording filled the new 4× 4 encryption image.



Original image

1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16

Encryption key L3

4 8 12 16 15 14 13 9 10 11 7 3 2 6 5 1

Fill the scanning path to encryption image

Encrypted image

4 8 12 16
15 14 13 9
10 11 7 3
2 6 5 1

Figure 5.1. Illustration of encryption

6. TEST RESULTS

The proposed encryption methodology was implemented in software using MATLAB 7.1 Figure 6.1 to 6.4 shows the 256 ∗ 256 gray-scale banana and coffeemaker image. The process encryption image is compliance the encryption key. It is clear that the SCAN methodology image encryption and decryption achieves an excellent encryption.

From previous mention, it is clearly known that we have 3 ∗ 8 possible partition and 4 ∗ 8 possible scan patterns. Due to which we can randomly select the partition and/or scan pattern. Thus, a high volume of secret keys is achieved as shown in Table 1.

Table1.Possible number of encryption keys

Size of the image that is to be encrypted	Number of possible encryption key pairs is greater than
4*4	32
8*8	$(3*8)*(4*8)^{4}$
16*16	$(3*8)^{5}*(4*8)^{16}$
32*32	$(3*8)^{21}*(4*8)^{64}$
64*64	$(3*8)^{85}*(4*8)^{256}$
128*128	$(3*8)^{341}*(4*8)^{1024}$
256*256	$(3*8)^{1365}*(4*8)^{4096}$
512*512	$(3*8)^{5461}*(4*8)^{16384}$

The Encryption key is ‘B0 (c1d3o1s4)'

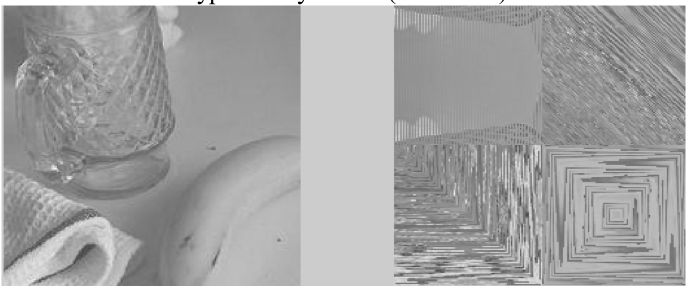


Figure 6.1: Original Banana image and corresponding encrypted image

The Encryption key is ‘B0 (c1d3o1s4)'

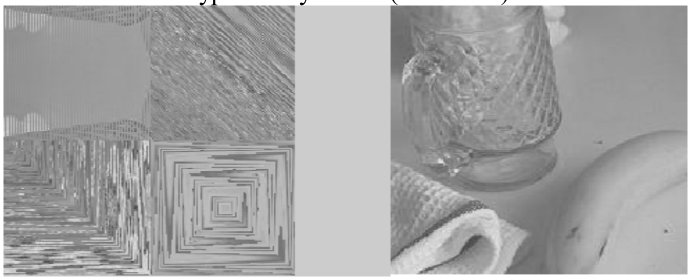


Figure 6.2: encrypted banana image and corresponding decrypted image.

The Encryption key is ‘Z7 (s6o2c7d4)'

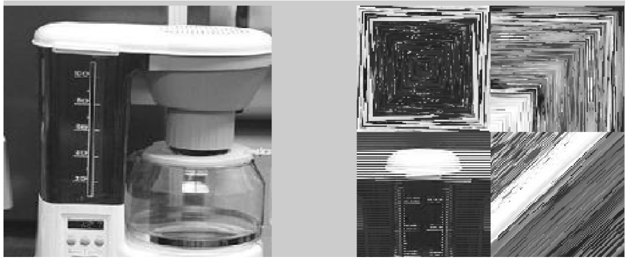


Figure 6.3: Original Coffeemaker image and corresponding encrypted image

The decryption key is ‘Z7 (s6o2c7d4)'

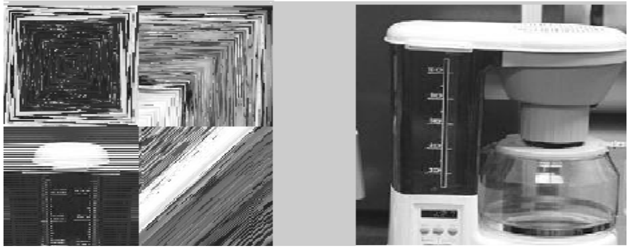


Figure 6.4: Encrypted Coffeemaker image and corresponding decrypted image

7. CONCLUSIONS

In our paper we implemented a image encryption using SCAN methodology. The encryption method is based on rearrangement of the pixel. The pixel is arrangement dependant on the encryption key. A number of scanning patterns, however, has been used for some image processing methods, such as implementation of pyramid and tree data structures, detection of objects in binary images, extraction and reconstruction of simple objects from binary images Moreover, the SCAN methodology is a useful tool for image decomposition. The salient features of our SCAN methodology image encryption and decryption can be summarized as:

- Loss less encryption of image.
- Secrete key have variable length to produce a large number of possible secrete keys.
- The security can be increased using more several encryption loops.

8. FUTURE SCOPE OF WORK

Extensions of this work could be the investigation of new scanning patterns applied efficiently to image processing. Such texture synthesis using SCAN words matrix elements rearrangement for parallel storage and manipulation.

In future work will hardware implement using DSP chip, and accelerate the process speed. Although the encryption is very much possible, but we analysis the gray-scale image the same form each encryption image. So can using progressive encryption achieve uniform distribute gray-scale. And very difficult observe form some feature.

9. REFERENCES:

1. Rafael C.Gonzalez, Richard E.Woods , “Digital Image Processing”, second Edition, Pearson Education, 2002 Reading.
2. Anil K. Jain, “Fundamentals of Digital Image Processing”, Prentice Hall Information & System Sciences Series.
3. William Stallings, “Cryptography and Network Security”(3rd Edition)
4. N. Bourbakis, C. Alexopoulos, “Picture data encryption using SCAN patterns,” Pattern Recognition, vol. 25, no. 6, pp. 567-581, 1992.
5. S. Maniccam, N. G. Bourbakis, “Image and video encryption using SCAN patterns,” Pattern Recognition, no. 37, pp. 725-737, 2004.
6. N-Bourbakis, C.Alexopoulos, “A Fractal Based Image Processing Language–Formal Modeling,” Pattern Recognition Journal, vol 32, no 2, 1999, pp 317-338.
7. J. Scharinger, “Fast encryption of image data using chaotic Kolmogorov flows,” Electronic Imaging, vol. 17, no. 2, pp. 318-325, 1998.8
8. X. Li, “Image compression and encryption using treestructure,” Pattern Recognition, no. 18 no. 11, pp. 1253-1259, 1997.

BIODATA:

SUMITH. K. S
7th Sem
Electronics and Communication
Kalpataru Institute of Tehnology
Tiptur-572002

ARJUN.H
7 th Sem
Electronics and Communication
Kalpataru Institute of Tehnology
Tiptur-572002

HARDWARE SYNTHESIS OF MODULES ENCOUNTERED IN JPEG ALGORITHM.

BY

PRASHOB NAIR, AMEYA KUVELKAR, SHANMON PHILIP,RAVIKRISHNA DHAVALIKAR & PROF. NITESH NAIK

GOA COLLEGE OF ENGINEERING

FARMAGUDI-GOA

ABSTRACT:

Reconfigurable computing has become increasingly important over the last few years. A lot of research has been carried out on the advantages of using reconfigurable hardware as an addition to the conventional processor.

This projects implements the JPEG compression scheme using reconfigurable hardware. The FPGA used in the implementation is the vertex development board. A number of challenges needed to be met to design and implement a JPEG decoder in hardware rather than in software running on a microprocessor. JPEG coding normally requires many floating-point calculations. Since these types of calculations are not efficiently implemented in custom hardware they were replaced by scaled fixed-point approximations. Also the JPEG decoding algorithm requires a substantial amount of memory. To reduce the memory requirements of the JPEG decoding only the core algorithm, which works on relatively small blocks of data, is to be implemented.

INTRODUCTION:

Data compression is the technique to reduce the redundancies in data representation in order to decrease data storage requirements and hence communication costs. Reducing the storage requirement is equivalent to increasing the capacity of the storage medium and hence communication bandwidth. Thus the development of efficient compression techniques will continue to be a design challenge for future communication systems and advanced multimedia applications.

In this paper, we discuss the JPEG codec using a Field Programmable Gate Array (FPGA). The FPGA will be programmed with the JPEG en/decoder design and will receive input PPM images from a serial communication link with a computer system and send the decoded output images back to the computer for viewing. For the purpose of describe digital circuits with which the FPGA will be configured Hardware Description Languages (HDLs) is to be used Vis -a- Vis Verilog. The verilog model is designed for the hardware implementation of the JPEG algorithm. through its port interface.

The module the basic building block in Verilog is used for the Hardware Description of the various blocks encountered in the algorithm. Typically, modules are instantiated and executed in the order the steps are encountered in the algorithm. A module so described provides the necessary functionality to the higher-level block

OVERVIEW OF JPEG ALGORITHM:

Steps in jpeg compression algorithm consist of:

- Divide the file into 8 X 8 blocks.
- Transform the pixel information from the spatial domain to the frequency domain with the Discrete Cosine Transform.
- Quantize the resulting values by dividing each coefficient by an integer value and rounding off to the nearest integer.
- Look at the resulting coefficients in a zigzag order.
- Do a run-length encoding of the coefficients ordered in this manner. Follow by Huffman coding.

HARDWARE REALIZATION OF THE ABOVE STEPS:

The hardware realization is obtained by using a hardware description language such as verilog.

The verilog module named "Compressor" is the basic encoding element which is described and is the top-level block (in a top-down design methodology). This module consists of a collection of lower-level design blocks of the jpeg algorithm.

Modules in the design

Module CONTROL () Top most block which reads input from memory and passes to the Compressor () circuit.

Module Compressor ( ) The block which prepares the image block and communicates with others**.**

`Module Level_shift` ( ) Shifts the range of the elements of the image matrix

`Module FDCT` ( ) Calculates DCT**.**

`Module Mat_Mul` ( ) Calls Matrix Multiplication logic.

`Module B_Mul` ( ) Calls Booth Multiplier.

`Module Quantize` ( ) Performs quantization of matrix.

`Module Mat_Divide` () Calls Matrix division logic.

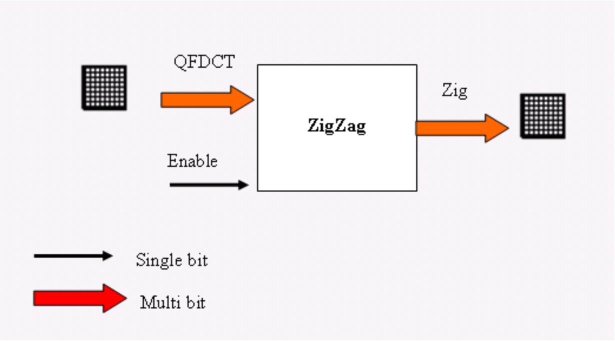
`Module Divide` () Calculates division of two binary nos.

`Module ZigZag` ( ) Performs Zig-Zag reading of matrix.

`Module RLE` ( ) Performs RLE encoding on array of data.

ZIG ZAG

Description on the zig zag block:



The output from Quantize (), is fed into this block so that data can be read in Zig Zag fashion.

ZIG ZAG verilog module is coded in State Machine Design Technique. This makes it Synthesizable. This particular block is instantiated by Module Quantize (), and is triggered by it.

Description of the zigzag verilog code:

- The Module ZigZag ( ) acts on two

dimensional array of register "QFDCT" created by Module Quantize ()

- The output is one dimensional array named "Zig"
- Snk_cnt is memory pointer register for "Zig". ele_cnt, k, i, j, z is used as counting registers.

States Encountered in the Circuit.

0. The idle state .No operations are performed. But it checks for input signal Enable. The circuit operates when Enable =1 i.e. next state is reached.
1. All registers reset to 0 except cnt_snk2 = 36, cnt_snk3=29.
2. If itern is even ,{ RowA, ColA, RowB, ColB }←{ (itern), 0, 7, and (7-itern) }
   If itern is odd, {RowA, ColA, RowB,ColB} ← {0, itern, (7-itern), 7}.
   The Zig [cnt_snk3] ←img [RowA],[ColA] .
3. *Cnt_snk3 ← {Cnt_snk3 + 1}.* And *cnt_snk* is reset.
4. The Zig [cnt_snk] ←img [RowA][ColB] .
   The Zig [cnt_snk3] ←img [RowA][ColB] .
   {cnt_snk, cnt_snk3 } ←incremented by 1.
5. If itern is even,{ColA, ColB} ← {incremented by 1} and {RowA, RowB} ← {decremented by 1}.
   If itern is odd,{*RowA*, RowB} ← {incremented by 1} and {ColA, ColB} ← {decremented by 1}.
6. ele_cnt is incremented by 1.
7. If ele_cnt is greater than the value (itern +1) then GOTO state 9 else GOTO state 5.
8. Itern is incremented by.
9. If itern is greater than 6 then GOTO state else GOTO state 2.
10. The Zig at location pointed by 43 is inputted with value from img which is at location 7 and 0, circuit then goes back to its initial state 0.

Algorithmic State Machine:

The figure representing the algorithmic state machine for the module is shown below.

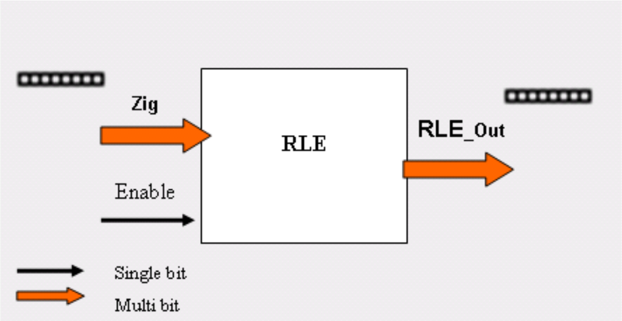
RUN LENGTH ENCODING:

Description of Run length Encoding Block:

Run length encoding is coded in State Machine Design Technique. This makes it Synthesizable. This particular block uses output data from Module ZigZag ( ), and is instantiated by it., and is triggered with the Enable signal.

Description of Run Length Encoding Code:

- The Module RLE (), acts on a one dimensional array of register "Zig" which has image data in it, generated by module ZigZag ( ).



- The Output is stored in another one dimensional array "RLE_OUT".
- Two registers LOC & C are used internally as memory pointer (element location) and counter respectively.

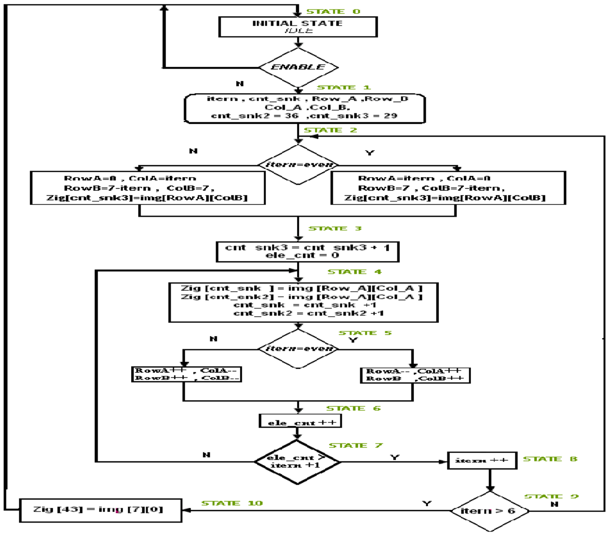
DESIGN:

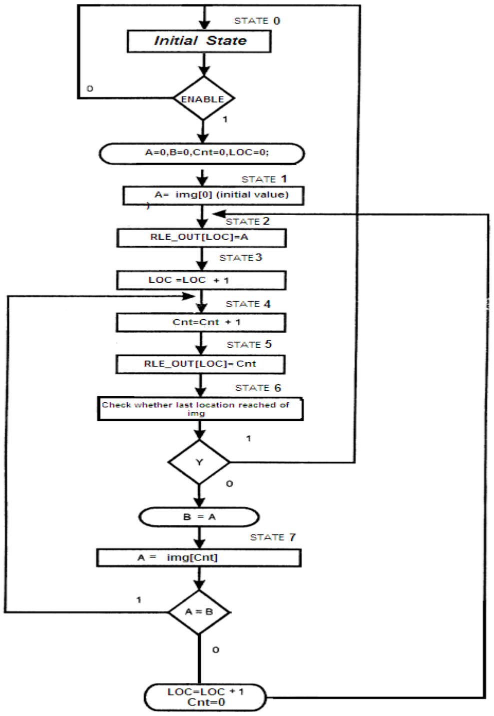
The binary- sequential state encoding style with zero-idle, is used for different transition states.

States Encountered in the Circuit and the control flow.

0. The idle state .No operations are performed .But it checks for input signal Enable.The circuit begins operation when Enable signal is given i.e. next state is reached.
1. Registers loc & C are reset to 0;
2. First element of Zig is accessed and stored in RLE_OUT.
3. LOC is then incremented by 1 (LOC = LOC+1).
4. C is incremented by 1
5. The C value is stored in RLE_OUT pointed by LOC.
6. The code checks whether “Zig” has its last data element being processed. If yes……… then go to INITIAL (IDLE) state, else …….. A’s content is saved in temporary reg B.
7. New “Zig” data is stored in A, then A and B is compared .If A = B ………Go to 5: If A != B ………loc=loc+1; C=0; Go to 2;

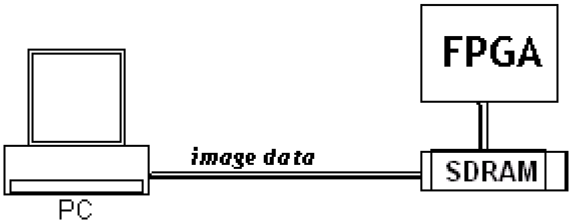
Algorithmic State Machine:





DESIGN PROTOTYPE:

Hardware Design



The figure shows the implementation of the architecture on Hardware. The PC is used to store an image into a memory, in this case SDRAM. The FPGA will then receive these data and execute the hardware programs configured in it.

The FPGA is configured with three Compressor modules for Red, Green and Blue data values. Hence the FPGA is partitioned to contain the three processing units.

The Compressor modules are the same except the data on which they work on.

Hence we encounter the concurrent nature in the code.

The concurrent parts of the code are implemented thus by using parallel modeling.

The APPLICATION:

The prototype can be used to implement and test compression algorithms .Not only of image compression but data compression as a whole. The algorithms can be modified and tested with ease due to the reconfigurable property of the FPGA.

IMPLEMENTATION RESULTS:

The image in the "raw" PPM format has its data stored in the SDRAM by using the PC. The RGB values are stored at in Alternate Locations in the memory.

For the codes written behavioral as well as state machine approach is followed. The codes then are tested in Xilinx, Release 7.1i - xst H.38.The HDL Synthesis is performed and RTL schematic is obtained. The FPGA device is programmed with the bit file. The compressed data is then stored back into another memory device.

CONCLUSION

In this paper, we have presented the design of hardware modules for JPEG algorithm. We have also presented a prototype design which

□Demonstrate the feasibility of reconfigurable hardware

□Get a feasible and reliable system of connecting an FPGA board with a computer system

REFERENCES

- Verilog® HDL: A Guide to Digital Design and synthesis, by Samir Palnitkar.
- Object Oriented Programming with C++, by Balagurusamy.
- www.wikipedia.org
- Tim Trunta, and Jesse Trunta, “An Introduction to JPEG compression”. http://online.redwoods.cc.ca.us/instruct/darnold/laproj/Fall2001/TheTrutnaBoys /JPEG_LaTeX.pdf
- www.jpeg.org
- http://www.faqs.org/faqs/compression-faq/part2/section-6.html
- D. A. Huffman, “A Method for the Construction of Minimum-Redundancy Codes”
- www.vlsibank.com
- Wallace Gregory K “The JPEG Still Picture Compression Standard”
- Arcangelo Bruna “Principles of Image compression”
- www.xilinx.com

Video Adaptation of the JPEG2000 Standard

Sanjeeva kumar Harihar, P.G.Student, University B.D.T. College of Engineering, Davanagere
sanjeevkumarharihar.ubdt@gmail.com
N. Manja Naik , Asst.Prof , Dept of Electronics & Communication Engg, UBDT College of Engg, Davangere
manjunn3@yahoo.com
Manjula.N.Harihar, Faculty, Department of Electronics and Communication Engg, Jain University
manjulaharihar@gmail.com
Santosh Nejakar, Guest Faculty, Dept of Electronics & Communication Engg, Govt Engg College, Haveri

***Abstract* — The Video adaptation of the JPEG2000 core is a JPEG encoder that forms a high performance solution for image and video compression applications. The JPEG-E can encode over 30 frames/sec of 4:3 HDTV, 1440x1152, 4:2:0 even on FPGA devices.**

Compliance with the baseline ISO/IEC 10918-1 JPEG standard makes the JPEG encoder core ideal for any cross platform application such as consumer digital cameras, camcorders, copiers, printers, scanners and remote surveillance systems.

Apart from baseline JPEG streams, the core is capable of producing non-standard motion JPEG streams. Furthermore, bandwidth-constraint applications may benefit from the included bit-rate control block. The core can be implemented using ASIC technology.

***Keywords* — MJ2K (Motion Jpeg 2000), DCT (Discrete Cosine Transform), RLE (Run length code), Huffman, Parser, Intra-frame, Inter-frame coding. ASIC (Application Specific Integrated Circuit).**

I. INTRODUCTION

JPEG is a compression standard that is used universally for digital photographic images. It is used in virtually all Digital Still Cameras (DSC) and for almost all Internet images. In addition, Motion JPEG is used in important applications from professional video editing to capture cards in a PC.

The JPEG standard was produced by an international group of experts - the Joint Photographic Experts Group under the ISO/IEC/ITU standards organizations. JPEG is a sister organization of MPEG.

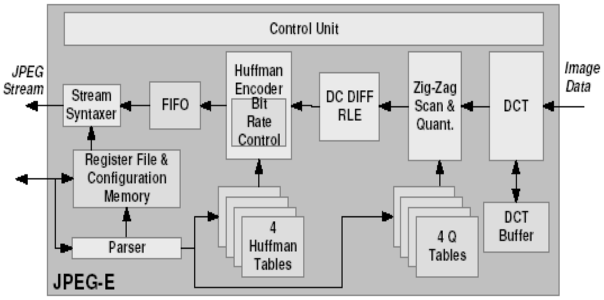


Figure-1. Block diagram of JPEG Encoder

In Motion JPEG 2000 each frame is coded individually. Intra-frame coding allows for random access and reduced complexity. In MPEG the encoding uses a series of frames – inter-frame coding. This allows for improved compression efficiency but the coding technique is more complex. The JPEG-E core block diagram is as shown in Figure-1. Image samples in any color format are fed to the JPEG-E in a MCU block by MCU block, raster scan order. The original version of the core is capable of processing up to 4 image components in any number of scans. The core compresses the image samples according to the programmed Quantization and Huffman tables (ref), and it produces an ISO/IEC 10918-1 compatible data stream.

Each 8x8 image block is frequency transformed by a forward DCT into the domain of 2-dimensional DCT (2D-DCT) basis images. The DCT concentrates the energy or information content in the left upper corner of the 8x8 block. The output DCT coefficients are then scanned in a Zigzag order and quantized according to programmed Quantization tables.

The quantization prepares the blocks for efficient coding. Each DCT coefficient block is divided by an 8x8 quantization matrix (one of the 4 possible Quantization Tables that supported in JPEG Baseline). Typical quantization matrices used in telecommunication applications turn small coefficients and coefficients representing finer detail into zeros.

The quantized DCT coefficients are then fed to the differential coding and run length coding unit that produces the Run-Amplitude pairs for the Huffman encoder. The Huffman encoder accepts symbols from RLE unit, which are further compressed by Huffman encoder, which encodes them according to one of the four possible programmed Huffman tables.

Finally, Huffman compressed data stream is written to the output FIFO. The stream syntax unit selects between data stored in the configuration memory and produced entropy coded data to assemble the final JPEG stream; it also adds the necessary stuffing bytes, and if configured to do so, inserts restart markers.

The core operation can be controlled via the control registers, accessible via the control interface, while it broadcast its state and configuration via registers accessible via the status interface.

MJPEG2000 or MJ2K is a video adaptation of the JPEG2000 standard for still photos. It treats a video stream as a series of still photos, compressing each individually, with no interframe compression and because it uses no interframe compression, it is ideal for editing. The

JPEG2000 standard is the official successor to JPEG and will eventually replace the older JPEG standard for high-quality image compression.

DCT (Discrete Cosine Transform)

The 8 × 8 DCT processes an 8 × 8 block of samples to generate an 8 × 8 block of DCT coefficients, as shown in Figure-2. The input maybe samples from an actual frame of video or motion- compensated difference (error) values, depending on the encoder mode of operation. Each DCT coefficient indicates the amount of a particular horizontal or vertical frequency within the block.

DCT coefficient (0, 0) is the DC coefficient, or average sample value. Since natural images tend to vary only slightly from sample to sample, low frequency coefficients are typically larger values and high frequency coefficients are typically smaller values

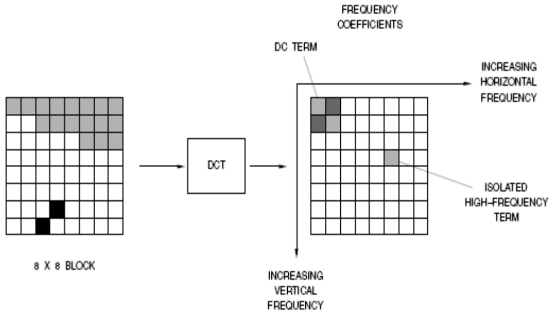

FREQUENCY COEFFICIENTS
DC TERM
INCREASING HORIZONTAL FREQUENCY
DCT
ISOLATED HIGH-FREQUENCY TERM
8 X 8 BLOCK
INCREASING VERTICAL FREQUENCY

Figure – 2. Concept of Discrete Cosine Transform

Quantization

The 8 × 8 block of DCT coefficients is quantized, limiting the number of allowed values for each coefficient. This is the first lossy compression step. Higher frequencies are usually quantized more coarsely (fewer values allowed) than lower frequencies, due to visual perception of quantization error. This results in many DCT coefficients being zero, especially at the higher frequencies

Zigzag Scanning

The quantized DCT coefficients are rearranged into a linear stream by scanning them in a zigzag order. This rearrangement places the DC coefficient first, followed by frequency coefficients arranged in order of increasing frequency. This produces long runs of zero coefficients.

Run Length Coding

The linear stream of quantized frequency coefficients is converted into a series of [run, amplitude] pairs, run indicates the number of zero coefficients, and [amplitude] the nonzero coefficient that ended the run

II. FEATURES

Baseline ISO/IEC 10918-1 JPEG Compliance

- Programmable Huffman Tables (2 DC, 2 AC) and Quantization tables (4).
- Supports all possible scan configurations and all JPEG formats for input/output data
- Any image size up to 64k x 64k

Non-Standard Features

- Motion JPEG encoding

Ease of Integration

- Single clock per input sample for encoding
- Fully programmable through standard JPEG stream marker segments
- Automatic headers generation

III. PROBLEM STATEMENT AND APPROACH

The Baseline JPEG standard has limitation that is the APP and COM marker segments can have a length up to 128 bytes each; up to 4 image components are supported. Sampling factors 3 and above 4 are not supported.

The Video adaptation of the JPEG2000 core can be modified to support more than 128-byte APP and COM markers and/or more than 4 image components upon request.

Programming of encoding options is achieved by feeding the core with a “configuration stream” via its JpegIN port. The configuration stream contains standard JPEG marker segments that define the Huffman tables, Quantization tables, frame, and scan format.

The core preserves its configuration until a new configuration stream is fed to it. Once configured, the core is ready to encode image samples.

Image samples in any color format are fed to the core in a MCU block by MCU block, raster scan order. The core compresses the image samples according to the programmed Quantization and Huffman tables, and it produces an ISO/IEC 10918-1 compatible data stream.

IV. OPERATION MODES

1. Configuration Mode

During the configuration mode the core receives marker segments interface, that define the Huffman and Quantization tables (DQT and DHT marker segments), the frame format (SOF0), the length of the restart interval (DRI), the next scan configuration (SOS), and also the comment and application marker segments (COM and APP).

The DQT, DHT, SOF0, DRI and SOS marker segments are decoded and used for the self configuration of the core, while the COM and APP marker segments are internally stored so that they are available for extracting in the JPEG stream whenever needed.

The format of the configuration stream must be as follows:

Marker	Code	Description
SOI	FFD8h	Start of Image
EOI	FFD9h	End of Image
DQT	FFDBh	Quantization Table(s)
DHT	FFC4h	Huffman Table(s)
DRI	FFDDh	Restart Interval
SOF0	FFC0h	Frame Definition
SOS	FFDAh	Start of Scan
COM	FFFEh	Comment
APPn	FFF0h to FFEFh	Application Segment, n=0-F

Figure–3. JPEG marker segments that participate in the configuration stream.

Format of the configuration stream must be as follows:

SOI	(Start of Image)
APPn	(Application Segment: Optional)
COM	(Comment Segment: Optional)
DQT	(Quantization Table(s) Segment(s))
SOF0	(Start of Frame (JPEG Baseline) Segment)
DHT	(Huffman Table(s) Segment(s))
DRI	(Restart Interval Segment Optional)
SOS	(Start of Scan Segment(s))
EOI	(End of Image)

2. Encoding Mode

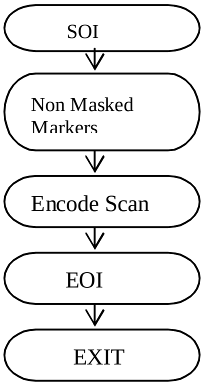


Figure – 4. Flow control under single-scan encoding

As in encoding mode the core receives image samples on an MCU by MCU, raster scan order via the Pixel-In interface and outputs a Baseline ISO/IEC 10918-1 JPEG stream via the Jpeg-Out interface.

The control flow under encoding mode is illustrated in Figure -4. The core extract SOI marker then encodes the frame samples and outputs the entropy-coded segments via the Jpeg-Out interface. After the entire scan is encoded the core extracts the EOI marker and leaves the encoding mode. Core supports both single-scan and multiscan.

3. Motion Encoding Mode

In this operation mode, the core receives pixels from each frame in a single scan and on an MCU by MCU, raster scan order via the Pixel-In interface. The core outputs a stream that is structured as follows:

SOI Marker
Non-masked markers
SOS Marker
ECS (frame 1)
SOS Marker
ECS (frame 2)
.
.
.
SOS Marker
ECS (frame n)
EOI Marker.

The encoder encodes the frame samples and outputs the entropy-coded segments via the Jpeg-Out interface. After the encoding of a frame is completed, the core automatically outputs a SOS marker and it becomes ready to encode next frame's pixels.

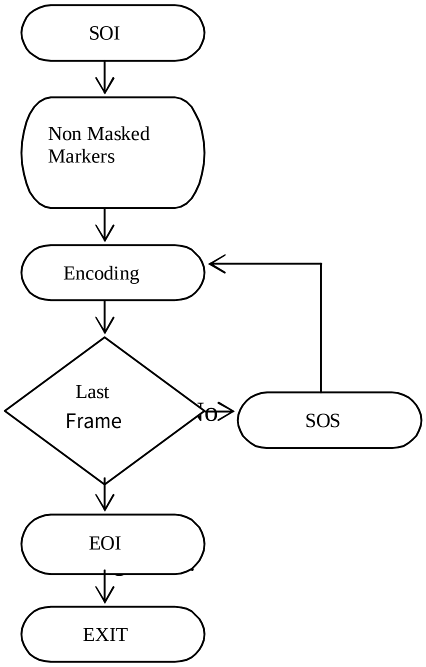


Figure-5 Control flow under motion encoding

V. PIN DIAGRAM

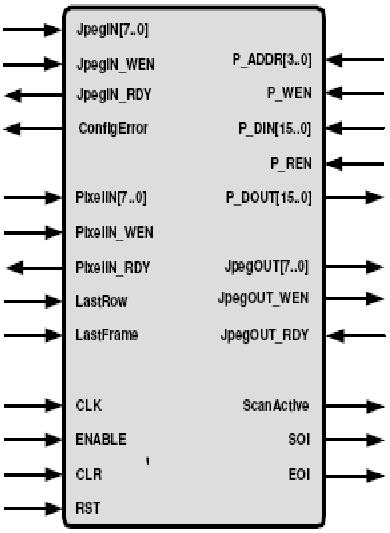


Figure-6. Pin diagram of Encoder

The Pixel-In Interface:

The pin diagram of encoder is as shown in Figure-6. The image samples are fed to the core via the PixelIN port. This stream must also be accompanied by a write-enable signal, PixelIN_WEN, which masks all valid data. The PixelIN_RDY output from the core controls the flow of data on the PixelIN bus and, if necessary, acts as a request to halt the input of data for an arbitrary time interval.

When the core is configured to produce DNL markers, the Last Row signal must mask at least one sample of the last MCU row of a scan. Finally, for Multi-Scan or Motion JPEG encoding the Last Frame signal needs to mask at least one sample of the last scan or last scan of last frame respectively.

The Jpeg-Out Interface:

The compressed JPEG data are sent to the output though the Jpeg-Out interface, when the JpegOUT_RDY input to the core permits it. Valid output data in the JpegOUT bus are masked with the JpegOUT_WEN output signal. If a request to stall the output data flow is identified – JpegOUT_RDY is de-asserted – the core stops extracting pixel data. WEN output is not deasserted, it is a responsibility of the external application to get the data only when the JpegOUT_RDY signal is high. Once a SOI marker is extracted in the JPEG out stream the SOI port is asserted high for one clock cycle.

In the same manner, after an EOI marker is extracted, the EOI pin is asserted for one clock cycle. The ScanActive signal is asserted after the extraction of a SOS marker to indicate that the core will produce the entropy-coded scan data from now on.

VI. CONCLUSION

The JPEG-E core is a JPEG encoder that forms a high performance solution for image and video compression applications. Probably the fastest core in market, the JPEG-E can encode over 30 frames/sec of 4:3 HDTV, 1440x1152, 4:2:0 even on FPGA devices.

The core can be modified to support more than 128-byte APP and COM markers and/or more than 4 image components upon request. Programming of encoding options is achieved by feeding the core with a configuration stream.

Apart from baseline JPEG streams, the core is capable of producing non-standard motion JPEG streams.

VII. REFERENCES

[1] Industry standards and specifications ISO/IEC IS 10918-1
ITU-T Recommendation T.81

[2] JPEG2000 compatible lossless coding journal of image
And coding of floating-point data. (Research Article),
International journal of image video Processing | January 1,
2007 Usevitch , Bryan E.

[3] Glenn Pearson & Michael Gill"An evaluation of motion Jpeg2000 for video Archiving" Proc Archiving 2005.

[4] S. W. Golomb (1966); IEEE Trans Info Theory "Run-length encodings"

[5] Chengie Tu, Jie Liang and Trac D "Adaptive Runlength Coding" IEEE signal processing Letters, Vol 10, NO.3, March 2003

[6] Andrew B & Watson, "Image Compression Using the Discrete Cosine Transform"

Implementation of Pervasive Virtual Workplace System using RFB and H.323 protocols

Anand kumar M

Department of Electronics and Communications, Visveswaraya Technological University
MTech (II year), Networking and Internet Engg, SJCE, Mysore
anand_mhalli@yahoo.co.in ph: 9739459183

***Abstract*— We propose a pervasive and efficient architecture to implement a Virtual Workplace system to control remote server/s. Rapid improvements in network bandwidth, cost, and ubiquity combined with the security hazards and high total cost of ownership of personal computers have created a growing market for thin-client computing. We introduce a remote display system architecture with integrated voice communication for high-performance thin-client computing in both LAN and WAN environments. Connections to any PC or server inside a network is established within just a few seconds. We should be able to remotely control our partner's PC as if we are sitting right in front of it. The system should also help to transmit the keyboard and mouse events from one computer to another. Using the advantages of H.323 protocol, two-way voice communication can be established to increase the interactivity. Better security is provided by employing VNC authentication at frame buffer level. Zlib Run Length Encoding which combines zlib compression, tiling, palettisation and run-length encoding is used before a message is transmitted.**

***Keywords*— ZRLE, H.323 protocol, Remote Frame Buffer, thin-client, Virtual Network Computing**

I. Introduction

Personal computing came into a much more distributed and pervasive environment in the last decade, and this trend will keep on in the future decades. In the last two decades, the centralized computing model of mainframe computing has shifted toward the more distributed model of desktop computing. But as these personal desktop computers become prevalent in today's large corporate and government organizations, the total cost of owning and maintaining them has become unmanageable. This problem is exacerbated by the growing use of mobile laptop computers and handheld devices to store and process information, which poses additional administration and security issues.

Thin client computing detaches processing logic and GUI display in the computer systems, and it provides an approach for terminals or devices with restricted capacity to borrow power from higher performance computers across the network. RFB is a simple protocol for remote access to graphical user interfaces. Because it works at the framebuffer level it is applicable to all windowing systems and applications, including X11, Windows and Macintosh. RFB is the protocol used in VNC (Virtual Network Computing). H.323 is a powerful protocol to implement two-way voice transmission. VNC using RFB protocol is core part of the system.

There are many challenges in this system design, one of which is the user interface system. In this paper, we introduce, a remote desktop architecture which provides virtually integrated but physically distributed desktop environment for the thin-client users. It integrates various application interfaces from diverse service nodes into one virtual desktop environment, and presents the virtual desktop to the ultra-thin user client.

In the reminder of this paper, we describe our work in more detail. Section 2 introduces the system architecture for Virtual Workplace. Section 3 describes the ZRLE encoding methodology. Section 4 gives the basics of H.323 protocol. Section 5 describes design and implementation of the system. Section 6 gives the details of technology used to develop the system. Related work is described in section 7. Finally, we conclude the full paper and introduce our future work in section 6.

II. System Architecture

The system architecture is based on a thin-client model in which all persistent state is maintained by the server. The remote endpoint where the user sits (i.e. the display plus keyboard and/or pointer) is called the RFB client or viewer. The endpoint where changes to the framebuffer originate (i.e. the windowing system and applications) is known as the RFB server.

The protocol also makes the client stateless. If a client disconnects from a given server and subsequently reconnects to that same server, the state of the user interface is preserved. Furthermore, a different client endpoint can be used to connect to the same RFB server. At the new endpoint, the user will see exactly the same graphical user interface as at the original endpoint. In effect, the interface to the user's applications becomes completely mobile. Wherever suitable network connectivity exists, the user can access their own personal applications, and the state of these applications is preserved between accesses from different locations. This provides the user with a familiar, uniform view of the computing infrastructure wherever they go.

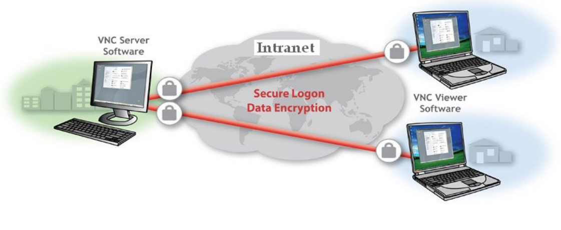


Fig.1 Secure VNC architecture

Within this basic architecture shown in Fig.1, one important design consideration is the choice of commands used to encode display information for transmission from the server to the client. The choices range from encoding high-level graphics calls to sending raw pixel data.

A. Display Protocol

The display side of the protocol is based around a single graphics primitive: "put a rectangle of pixel data at a given x,y position". At first glance this might seem an inefficient way of drawing many user interface components. However, allowing various different encodings for the pixel data gives us a large degree of flexibility in how to trade off various parameters such as network bandwidth, client drawing speed and server processing speed.

B. Input Protocol

The input side of the protocol is based on a standard workstation model of a keyboard and multi-button pointing device. Input events are simply sent to the server by the client whenever the user presses a key or pointer button, or whenever the pointing device is moved.

C. Security

VNC authentication is to be used and protocol data is to be sent unencrypted. The server sends a random 16-byte challenge:

No. of bytes	Type [Value]	Description
16	U8	challenge

The client encrypts the challenge with DES, using a password supplied by the user as the key, and sends the resulting 16-byte response:

No. of bytes	Type [Value]	Description
16	U8	response

The protocol continues with the SecurityResult message.

III. ZRLE Encoding

ZRLE stands for Zlib Run-Length Encoding, and combines zlib compression, tiling, palettisation and run-length encoding. On the wire, the rectangle begins with a 4-byte length field, and is followed by that many bytes of zlib-compressed data. A single zlib "stream" object is used for a given RFB protocol connection, so that ZRLE rectangles must be encoded and decoded strictly in order.

No. of bytes	Type [Value]	Description
4	U32	length
length	U8 array	zlibData

The zlibData when uncompressed represents tiles of 64x64 pixels in left-to-right, top-to-bottom order, similar to hextile. If the width of the rectangle is not an exact multiple of 64 then the width of the last tile in each row is smaller, and if the height of the rectangle is not an exact multiple of 64 then the height of each tile in the final row is smaller.

ZRLE makes use of a new type CPIXEL (compressed pixel). This is the same as a PIXEL for the agreed pixel format, except where true-colour-flag is non-zero, bits-per-pixel is 32, depth is 24 or less and all of the bits making up the red, green and blue intensities fit in either the least significant 3 bytes or the most significant 3 bytes.

In this case a CPIXEL is only 3 bytes long, and contains the least significant or the most significant 3 bytes as appropriate. bytesPerCPixel is the number of bytes in a CPIXEL.

Each tile begins with a subencoding type byte. The top bit of this byte is set if the tile has been run-length encoded, clear otherwise. The bottom seven bits indicate the size of the palette used - zero means no palette, one means that the tile is of a single colour, 2 to 127 indicate a palette of that size. The possible values of subencoding can be found in [1].

IV. H.323 Protocol Basics

H.323 defines the interworking of
– call signalling,
– call control,
– and media stream protocols,
in order to build a packet-based multimedia communications system.

H.323 [2] further describes the network components that are used to build up such a communications system. H.323 can be seen as an "umbrella standard" which aggregates standards for multimedia conferencing over packet-based networks.

D. H.323 Components

- Terminal
  - Video/audio/data client
- MCU
  - Conference co
  - Content mixing
- Gateway
  - Protocol translation
- Gatekeeper
  - Address resolution
  - Admission control
- Terminals, MCUs, and Gateways are called H.323 Endpoints
- An endpoint is ISDN "callable"

E. Protocol Stack

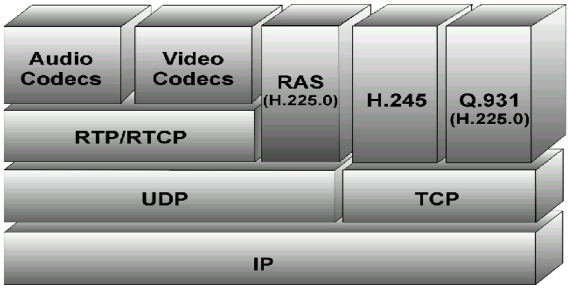


Fig.2 H.323 protocol stack

- Audio codecs (G.711, G.723.1, G.728, etc.) and video codecs (H.261, H.263) compress and decompress media streams

- Media streams transported on RTP/RTCP
    - RTP carries actual media
    - RTCP carries status and control information
    - RTP/RTCP carried unreliably on UDP
    - Signaling is transported reliably over TCP
    - RAS - registration, admission, status
    - Q.931 - call setup and termination
    - H.245 - capabilities exchange

F. Gatekeeper Routed Call Signalling (Q.931/H.245) between client A and client B

Establishing a call between client A and client B:

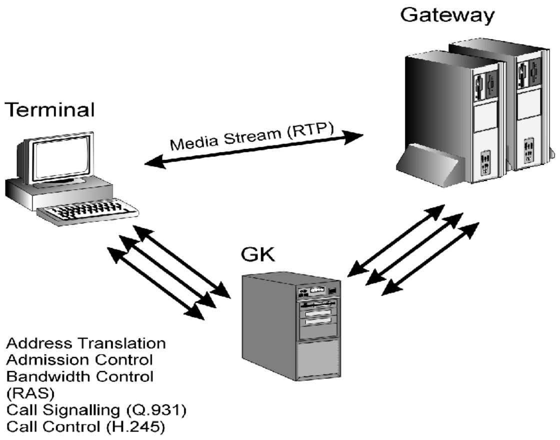


Fig.3 Gatekeeper Routed Call Signalling (Q.931/H.245)

- Discover and register with the gatekeeper - RAS channel Routed call setup between the endpoints through the gatekeeper - Q.931 call signalling
- Initial communications and capability exchange - H.245 call control
- Establish multimedia communication/call services - H.245 call control
- Call termination - H.245 call control & Q.931 call signalling.

V. Design and Implementation

G. Basic Modules

- Load all active System IP Addresses within LAN or WAN:

  Browse for all active IP addresses (Systems which are switched on) in local network and load for administrator console for further monitoring. Automatically detects active IP addresses.
- Remote Process/Service Monitoring:

  Administrator can choose any of the active system in the network to monitor their processes and services running in that particular machine. Also admin would be given an option to kill unwanted process/s running in those remote systems.
- Uni-cast/Broadcast Message:

  Administrator has an option to send message to a single system or multiple systems at a given point of time within the network.
- File transfer from local system to remote system:

Administrator would also be given an option to transfer files to/from his system to remote system. Here administrator would be able to view the complete file system of remote system to browse for.

H. Advanced Module

- Remote System Computing with desktop sharing, that is getting real time display status with video streaming using RFB Protocol.

  Using RFB protocol [1], remote desktop is shared to administrator console for monitoring/controlling. Here the remote desktop is viewed as a continuous stream of images like a video rather been displaying just as a still image. With such a protocol in use, the advantage on our system would be to access the remote systems desktop as a live video or as if we are sitting right in front of it.

This module would allow even to passes all keyboard and mouse events from administrator system to remote system, i.e. the admin can have control over remote systems keyboard and mouse. This module has to even take care of display resolutions which may differ from system to system over a network and has to overcome the mouse movement coordination issues.

- Integrating Voice over IP (VoIP) Option.

  Using the advantages of Voice over IP protocol, the system can also communicate with the remote desktop user with voice data. This module allows our system user to communicate voice data under 2-way live communication.

- Recording remote desktop to a video file

  This module could be used to record the remote desktop shared video into a file and store the same under the administrators file system.

VI. TECHNOLOGY USED

- **FRONT END:** VISUAL STUDIO 2005 WINFORMS (WITH FRAME WORK 2.0)
- **PROGRAMMING LANGUAGE:** C# (C HASH/C SHARP)

Concepts from Technology:

1) Controls under Winforms.Net
2) Multi Threading
3) File Handling
4) Socket Programming
5) Remoting under C#
6) H.323 protocol for voice communication
7) RFB protocol for remote desktop streaming

VII. RELATED WORK

There are some recently developed systems similar to Virtual Workplace system. HP SoftUDC [3] is a software-based cost-effective and flexible solution to the quest for utility computing. The main difference between SoftUDC and Virtual Workplace is that, SoftUDC mainly concern the problem of isolation and migration basing on virtualization, where as Virtual Workplace mainly concern the problem of integration basing on virtualization. Stanford Collective [4] is a system that delivers managed desktops to personal computer (PC) users. The main difference between Stanford Collective and Virtual Workplace is that Collective's main motivation is to achieve better security and lower cost of management, and its server side is a centralized architecture. HP Interactive Grid [5] is a grid computing architecture designed to support graphical interactive sessions. The fundamental difference between our Virtual Workplace and HP Interactive Grid is that we mainly focus on the distribution features of the future personal computing environment and we address the challenges brought by such features, such as GUI merging, virtual user space construction, etc. Moreover, the system above only provides remote KVM devices, and no other devices, such as USB storage or CDROM are provided.

VIII. CONCLUSION AND FUTURE WORK

We introduce Virtual Workplace System: a remote desktop architecture for the virtual personal computing. Virtual Workplace is a computing paradigm proposed to accommodate the more and more secured and pervasive personal computing environment with voice communication utility. The main innovative works are: a remote desktop architecture, which provides an integrated desktop window-style user environment in a distributed computing environment; a framebuffer based GUI merging mechanism, which is perfect for GUI merging under distributed computing environment; a virtual desktop manager, which acting as both an resource manager and a security control tool for the virtual desktop; H.323 protocol which provides two-way voice communication feature for the virtual system.

According to our experiments, the architecture brings better resource utilization and load balance in personal computing environment, and it performs perfect in display latency, overhead, robustness, and system scalability. Compared with other similar systems, Virtual Workplace shows it superiorities in scalability, efficiency, flexibility, and resource utilization.

The research on virtualization is still in an initial stage, and there are many works to be done in the future. Firstly, more kinds of devices should be supported on the client side. Personal computing environment consists of not only GUIs and local storage interfaces, but also audio devices, printing services, etc. In the next stage work, more kinds of devices should be supported on the user terminals. Secondly, a more flexible rendering strategy is needed.

In current version of Virtual Workplace implementation, we adopt an absolutely server-side GUI rendering strategy. To some stronger clients, this strategy wastes client processing capacity and consumes more network bandwidth. So a more flexible rendering strategy, which can automatically balance the loadings between the server and the client, should be developed. Thirdly, more technologies should be developed to guarantee the system's performance in a more complex network environment.

The experiments in this paper are made in a stable network environment. When the system is ported to a more complex environment, its performance, such as latency, overhead, robustness and scalability, will be greatly affected. So many technological measurements should be developed to guarantee the system's performance in a more complex environment.

REFERENCES

[1] Tristan Richardson, "The RFB Protocol," AT&T Labs Cambridge Whitepaper, March, 2005.

[2] www.imtc.org

[3] Mahesh Kallahalla, Mustafa Uysal, Ram Swaminathan, David Lowell, Mike Wray, Tom Christian, Nigel Edwards, Chris Dalton, Frederic Gittler, "SoftUDC: A Software-Based
Data Center for Utility Computing," IEEE Computer, November 2004.

[4] Ramesh Chandra, Nickolai Zeldovich, Constantine Sapuntzakis, Monica S. Lam, "The Collective: A Cache-Based System Management Architecture," Proceedings of the 2nd Symposium on Networked Systems Design and Implementation, May 2005.

[5] Vanish Talwar, Sujoy Basu, Raj Kumar, "An Environment for Enabling Interactive Grids," Proceedings of
the 12th IEEE International Symposium on High Performance Distributed Computing, 2003.

[6] J. S. Bruner, "The Act of Discovery," Harvard Educational Review, vol. 31, 1961, pp. 21-32.

[7] R. Schank and A. Kass, "A goal-based scenario for high school students," Communications of the ACM, vol. 39, 1996, pp. 28-29.

[8] YANG SJ, NIEH J, KRISHNAPPA S, MOHLA A, SAJJADPOUR M,
"WEB BROWSING PERFORMANCE OF WIRELESS THIN-CLIENT COMPUTING," PROCEEDINGS OF THE TWELFTH INTERNATIONAL
CONFERENCE ON WORLD WIDE WEB, 2003.

[9] P.T. ARES, "THE NETWORK BLOCK DEVICE",
HTTP://WWW2.LINUXJOURNAL.COM/ARTICLE/3778,
MAY,2000.

[10] HTTP://WWW.GNOME.ORG/.

[11] BRODERSEN, R.W, "THE NETWORK COMPUTER AND ITS FUTURE,
"IEEE SOLID-STATE CIRCUITS CONFERENCE, SAN FRANCISCO, FEB.
1997.

[12] H. OKADA, K. KATO, T. IKEGAMI, Y. TATSUMI, AND T. ASAHI. PROPOSAL OF A PC REMOTE CONTROL SYSTEM BY MOBILE DEVICES. IN IPSJ SIG NOTES, VOLUME 93 OF HUMAN INTERFACE, INFORMATION PROCESSING SOCIETY OF JAPAN, MAY 2001, 1–6. (IN JAPANESE).

Retrieval of Original Image and Quality Analysis Using MGA based Wavelet Decomposition

N. Ajay Kumar
M.Tech- IV Sem (DECS)
S.K.T.R.M.C.E, A.P.
Email: nara_ajay402@yahoo.com

L. Manjunath
Dept. of ECE, S.K.T.R.M.C.E,
Email: l_manjunathrao@rediffmail.com

Abstract— **The current standard is the wavelet-domain natural imagestatistics model (WNISM) fails to consider the statistical correlations of wavelet coefficients in different subbands and the visual response characteristics of the mammalian cortical simple cells. In addition, wavelet transforms are optimal greedy approximations to extract singularity structures, so they fail to explicitly extract the image geometric information, e.g., lines and curves. In this paper, to target the aforementioned problems in IQA,we develop a novel framework for IQA to mimic the human visual system (HVS) by incorporating the merits from multiscale geometric analysis (MGA), contrast sensitivity function (CSF), and the Weber's law of just noticeable difference (JND). In the proposed framework, MGA is utilized to decompose images and then extract features to mimic the multichannel structure of HVS. Additionally, MGA offers a series of ransforms including wavelet, curvelet, bandelet, contourlet, wavelet-based contourlet transform (WBCT), and hybrid wavelets and directional filter banks (HWD), and different transforms capture different types of image geometric information. CSF is applied to weight coefficients obtained by MGA to simulate the appearance of images to observers by taking into account many of the nonlinearities inherent in HVS. JND is finally introduced to produce a noticeable variation in sensory experience. Thorough empirical studies are carried out upon the LIVE database against subjective mean opinion score (MOS) and demonstrate that the proposed framework has good consistency with subjective perception values and the objective assessment results can well reflect the visual quality of images,**

Index Terms—**Contrast sensitivity function (CSF), human visual system (HVS), image quality assessment (IQA), just noticeable difference (JND), multiscale geometric analysis (MGA), reduced-reference (RR).**

I. INTRODUCTION

THE objective of *image quality assessment* (IQA) is to provide computational models to measure the perceptual quality of an image. In recent years, a large number of methods have been designed to evaluate the quality of an image, which may be distorted during acquisition, transmission, compression, restoration, and processing (e.g., watermark embedding). The past five years have demonstrated and witnessed the tremendous and imminent demands of IQA methods in various applications, including evaluating and optimizing image processing algorithms and systems. Existing IQA methods can be categorized into subjective and objective methods. Results of a subjective method are directly given by human observers, so it is probably the best way to assess the quality of an image. This is because human observers are the ultimate receivers of the visual information contained in an image. However, subjective IQA methods are expensive and time consuming, so they cannot be easily and routinely performed for many scenarios, e.g., real time systems. The latter depends on the quantified parameters which are obtained from metrics to measure the image quality. Metrics are usually obtained from either reference or distorted images to reflect a number of image characteristics To evaluate the quality of a distorted image, FR methods usually provide the most precise evaluation results in comparing with NR and RR. Conventional FR IQA methods calculate pixel-wise distances, e.g., *peak signal-to-noise ratio* (PSNR) and *mean square error* (MSE), between a distorted image and the corresponding reference, but they have not been in agreement with perceived quality measurement widely .

To evaluate the quality of a distorted image in real time systems, NR methods [4], which output evaluation results without the corresponding reference, have been designed. However, all these methods rely on strong hypotheses, i.e., they are designed for one or a set of predefined distortions, so there is a big gap between NR methods and real scenarios.

To provide a compromise between FR and NR, RR methods which become popular in recent years, have been designed for IQA by employing partial information of the corresponding reference. Based on results in natural image statistics,Wang *et al.*proposed the *wavelet-domain natural image statistic metric* (WNISM), which achieves promising performance for image visual perception quality evaluation. Based on this fact, WNISM measures the quality of a distorted image by the fitting error between the wavelet coefficients of the distorted image and the Gaussian distribution of the reference.

In this paper, to target the aforementioned problems in WNISM, to further improve the performance of RR IQA, and to broaden RR IQA related applications, a novel HVS driven framework is proposed for IQA. This framework is constructed by pooling *multiscale geometric analysis* (MGA), *contrast sensitivity function* (CSF) [7], and the Weber's law of *just noticeable difference* (JND) [7] together. The new framework is consistent with HVS: MGA decomposes images for feature extraction to mimic the multichannel structure of HVS, CSF re-weights MGA decomposed coefficients to mimic the nonlinearities inherent in HVS, and JND produces a noticeable variation in sensory experience. This framework contains a number of different ways for IQA because MGA offers a series of transforms including wavelet [5], curvelet ,bandelet [6], contourlet , *wavelet-based contourlet transform* (WBCT) [3], and *hybrid wavelets and directional filter banks* (HWD) and different transforms capture different types of image geometric information. Extensive experiments based on LIVE database against subjective *mean opinion score* (MOS) have been conducted to demonstrate the effectiveness of the new framework.

2. MULTISCALE GEOMETRIC ANALYSIS

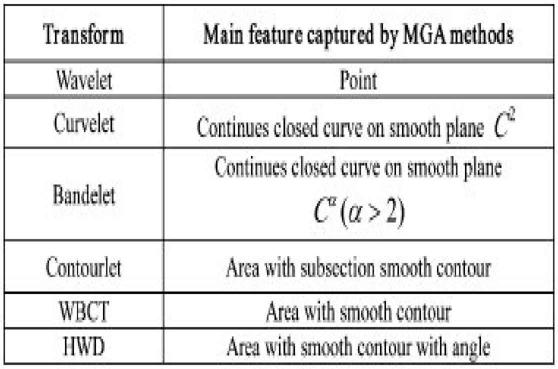
Wavelet transform [5] have been successfully applied in a wide variety of signal processing tasks, e.g., speech signal compression and voice-based person identification, because it is an optimal greedy approximation to extract singularity structure for 1-D piecewise smooth signals. Although the 2-D extension, i.e., 2-D wavelet transform, can also be applied to image processing relevant applications, e.g., compression, de-noising, restoration, segmentation, and structure detection, it can only deal with the singularity problem of point. To our knowledge, image contains rich and varied information, e.g., texture and edges, and wavelet transform is not effective in dealing with directional information. Therefore, it is essential to develop a directional representation framework for precisely detecting orientations of singularities like edges in a 2-D image while providing near optimal sparse representations.

Multiscale geometric analysis (MGA) [6] is such a framework for optimally representing high-dimensional function. It is developed, enhanced, formed and perfected in signal processing, computer vision, machine learning, and statistics. MGA can detect, organize, represent, and manipulate data, e.g., edges, which nominally span a high-dimensional space but contain important features approximately concentrated on lower dimensional subsets, e.g., curves. MGA contains a large number of tools and gets wavelet transform involved as a special case.

For IQA, we need to find MGA transforms, which perform excellently for reference image reconstruction, have perfect perception of orientation, are computationally tractable, and are sparse and effective for image representation. Among all requirements for IQA, effective representation of visual information is especially important.. However, MGA transforms can capture the characteristics

of image, e.g., lines, curves, cuneiforms and the contour of object. As mentioned in Table I, different transforms of MGA capture different features of an image, and complement to each other.

TABLE I
MAIN FEATURES CAPTURED BY DIFFERENT MGA TRANSFORMS



Transform	Main feature captured by MGA methods
Wavelet	Point
Curvelet	Continues closed curve on smooth plane C^2
Bandelet	Continues closed curve on smooth plane $C^\alpha(\alpha > 2)$
Contourlet	Area with subsection smooth contour
WBCT	Area with smooth contour
HWD	Area with smooth contour with angle

3. MULTISCALE GEOMETRIC ANALYSIS-BASED IMAGE QUALITY ASSESSMENT

MGA contains a series of transforms,which can analyze and approximate geometric structure while providing near optimal sparse representations. The image sparse representation means we can represent the image by a small number of components, so little visual changes of the image will affect these components significantly. The objective of this framework is providing IQA results for distorted images, which have good consistency with subjective perception values. This framework incorporates merits from three components, i.e., MGA, *contrast sensitivity function* (CSF), and theWeber's law of *just noticeable difference* (JND),to model the process of image perception.Fig. 1 shows the framework for IQA and it works with the following stages: 1) anMGA transform, e.g., curvelet, bandelet,contourlet, WBCT, and HWD, is utilized to decompose both the reference image at the sender side and the distorted image at the receiver side, 2) CSF masking is utilized to balance subbands coefficients in different scales obtained by the MGA transform.With this stage, we can simulate the appearance of images to observers by taking into account many of the nonlinearities inherent in HVS, 3) JND produces a noticeable variation in sensory

experience, 4) a histogram is constructed for image representation,each bin of the histogram corresponds to the amount of visual sensitive coefficients of a selected subband, and finally the normalization step is applied to the histogram, and 5) the IQA result is the transformed city-block distance between the normalized histograms of the reference and distorted images.

3.1 MGA-Based Feature Extraction

In this paper, a number of MGA transforms, which are curve let, band let, contour let, WBCT and HWD, are considered for image decomposition and feature extraction. Moreover, wavelet is utilized as the baseline for comparison. In this framework, MGA is utilized to decompose images and then extract features to mimic the multichannel structure of HVS.

- **Wavelet transform**: It [5] is popular to analyze signals in both time and frequency domains simultaneously and adaptively. By using multiscale operation, it extracts effective features to represent signals, especially for nonstationary signals (as mentioned in Table I, it extracts singularity structure of an image). In our experiments, three level wavelet transform with "Daubechies" filters, is applied to decompose the image into nine highpass subbands and a lowpass residual subband. Coefficients in all nine highpass subbands are preserved for further processing,e.g., CSF masking and JND. As shown in Fig. 2, selected subbands are marked with white dashed boxes and numerals.

- **Contourlet transform**: It is constructed via filter banks and can be viewed as an extension of wavelets with directionality. In the proposed framework, the image is decomposed into three pyramidal levels. Based on the characteristics of directional filter banks for decomposition, coefficients in half of the directional subbands are selected for further processing.shown in fig.3.

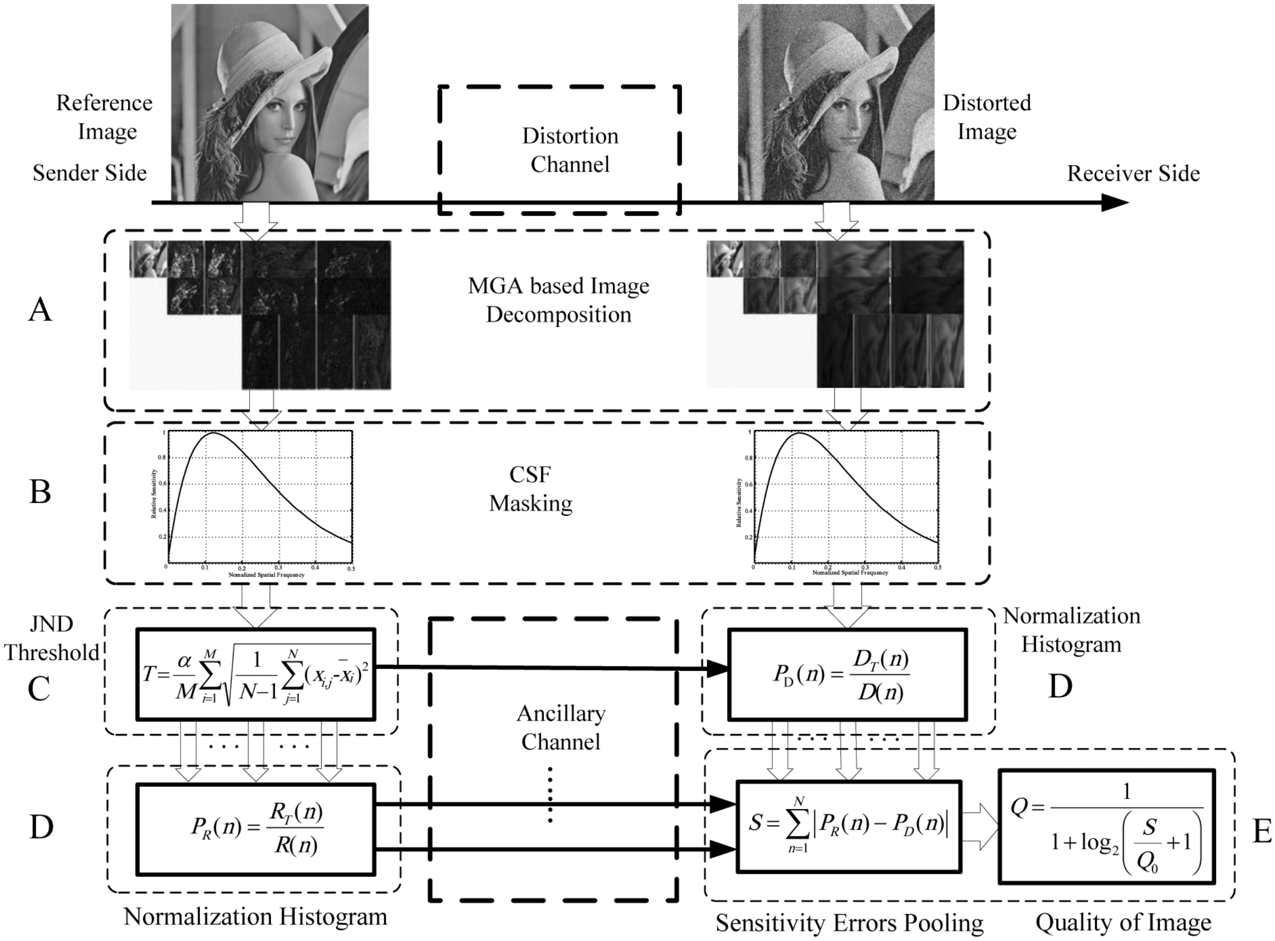

Reference Image
Sender Side
Distortion Channel
Distorted Image
Receiver Side
A
MGA based Image Decomposition
B
Relative Sensitivity
Normalized Spatial Frequency
CSF Masking
JND Threshold
C
$T=\frac{\alpha}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(x_{i,j}-\bar{x}_i)^2}$
Ancillary Channel
Normalization Histogram
$P_D(n)=\frac{D_T(n)}{D(n)}$
D
D
$P_R(n)=\frac{R_T(n)}{R(n)}$
Normalization Histogram
$S=\sum_{n=1}^{N}\left|P_R(n)-P_D(n)\right|$
$Q=\frac{1}{1+\log_2\left(\frac{S}{Q_0}+1\right)}$
E
Sensitivity Errors Pooling
Quality of Image

Fig. 1. Multiscale geometric analysis-based image quality Analysis framework.

- **WBCT**: It [3] is an enhanced version of contourlet transform based on two stages of filter banks that are nonredundant and perfect for reconstruction. As mentioned in Table I, it captures areas with smooth contours. Based on these filters, an image is decomposed into 48 high frequency directional subbands and a lowpass residual subband.Coefficients in half of the subbands at each fine scale are selected for further processing.
- **HWD Transform**: Based on these filters, an image is decomposed into 32 high frequency directional subbands and a lowpass residual subband. Coefficients in half of subbands at each fine scale are selected for further processing.

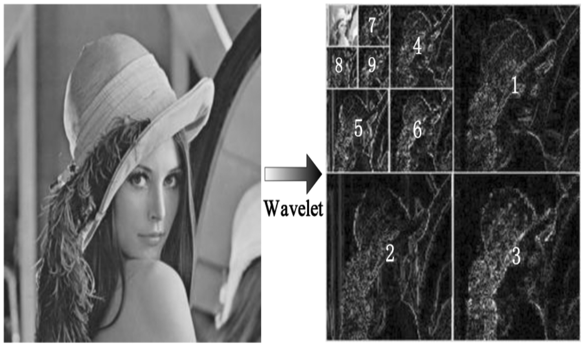


Fig. 2. Wavelet transform-based image decomposition.

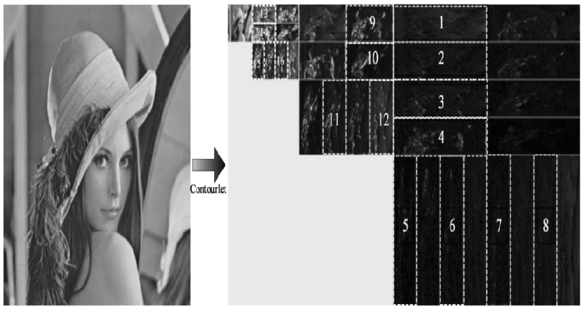


Fig. 3. Contourlet transform-based image decomposition.

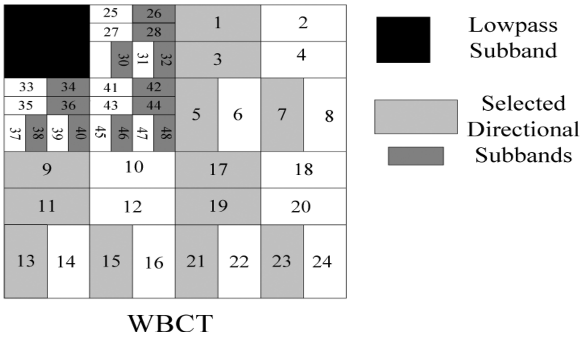


Fig. 4. WBCT-based image decomposition.

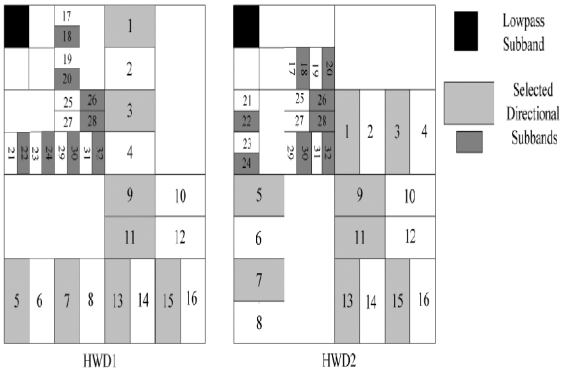


Fig. 5. HWD transform-based image decomposition.

3.2.CSF Masking

$$T = \frac{\alpha}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{N_i-1}\sum_{j=1}^{N_i}(x_{i,j}-\bar{x}_i)^2}$$

MGA is introduced to decompose images and then extract features to mimic the multichannel structure of HVS, i.e., HVS works similar to a filter bank (containing filters with various frequencies). CSF measures how sensitive we are to the various frequencies of visual stimuli, i.e., we are unable to recognize a stimuli pattern if its frequency of visual stimuli is too high..Because coefficients in different requency subbands have different perceptual importance, it is essential to balance the MGA decomposed coefficients via a weighting scheme, CSF masking.In this framework, the CSF masking coefficients are obtained by *modulation transfer function* (MTF),i.e.,

$$H(f) = a(b + cf)\exp(-cf)^d \qquad (1)$$

where , the center frequency of the band, is the radial frequency in cycles/degree of the visual angle subtended, is the normalized spatial frequency with units of cycles/pixels,and is the sampling frequency with units of pixels/degree.

According to [4], , and are 2.6, 0.192, 0.114, and 1.1,respectively.

The sampling frequency is defined as

$$f_s = \frac{2 \cdot v \cdot \tan(0.5^\circ) \cdot r}{0.0254} \qquad (2)$$

where is the viewing distance with units of meter and is the resolution power of the display with units of pixels/inch. In this framework, is 0.8 meter (about 2–2.5 times height of the display),the display is 21 inches with the resolution of 1024 768,and pixels/inch.

3.3. JND Threshold

Because HVS is sensitive to coefficients with the larger magnitude,it is valuable to preserve visually sensitive coefficients.JND, a research result in psychophysics, is a suitable way for this function.In our framework, MGA is introduced to decompose an image and highpass subbands contain the primary contours and textures information of the image,CSF masking makes coefficients have similar perceptual importance in different frequency subbands, and then JND is alculated to obtain a threshold to remove visually insensitive coefficientsThe lower the JND threshold is, the more coefficients are utilized for image reconstruction and the better visual quality of the reconstructed image is. Therefore, the normalized histogram reflects the visual quality of an image. Here, the JND threshold is defined as in equ.(3) below

where is the $X_{t j}$ is the jth coefficient of the ith subband in the finest scale and x_i, is the mean value of the th subband coefficients, M is the amount of selected subbands in the finest scale, N_i is the amount of coefficients of th subband, and is a tuning parameter corresponding to different types of distortion.

3.4. Normalized Histogram for Image Representation

By using JND threshold *T*, we can count the number of visually sensitive coefficients in the *n* th selected sub-band and define the value as C_T(n)., which mean the number of coefficients in the *n*th selected sub-band which are larger than *T* obtained. The number of coefficients in the *n*th selected sub-band is C(n). Therefore, for a given image, we can obtain the normalized histogram with *L* bins (*L* sub-bands are selected) for representation and the *n*th entry is given

$$P(n) = \frac{C_T(n)}{C(n)}. \qquad (4)$$

3.5.Sensitive Errors Pooling

Based on (4), we can obtain the normalized histograms for both the reference and the distorted images as P_R (n) and P_D (n) respectively. In this framework, we define the metrics of the distorted image quality as

$$Q = \frac{1}{1 + \log_2\left(\frac{S}{Q_0} + 1\right)} \qquad (5)$$

where S is the city-block distance between P_D (n) and, P_R (n) and is a constant used to control the scale of the distortion measure. In this framework, we set as Q_0 0.1. The log function is introduced to reduce the effects of large S and enlarge the effects of small S, so that we can analyze a large scope of S conveniently.

4. ALGORITHM

STEP 1: The images in the database and Query image are converted into a data file.

STEP 2: In this stage MGA is introduced to decompose the query images and then extract features to mimic the multi-channel structure of HVS, i.e., HVS works similar to a filter bank. In this project we are using wave let transform to decompose the Query image and then extract the co efficient in different sub bands.

STEP 3: Next we utilize Contrast sensitivity masking(CSF)which measures how sensitive we are to the various frequencies of visual stimuli, i.e., we are unable to recognize a stimuli pattern if its frequency of visual stimuli is too high. Because coefficients in different frequency sub bands have different perceptual importance, it is essential to balance the MGA decomposed coefficients via a weighting scheme, CSF masking.

STEP 4: Using JND Threshold we produce a noticeable variation in sensory experience. In this stage we determine Inverse Discrete Wavelet transform

STEP5: Next we reconstruct the original image from the coefficients which are obtained after CSF masking. we display the compression ratio query image and recovery score and single and multi level decomposition structures.

5. EXPERIMENTAL RESULTS:

Fig 7.1 a sample query image

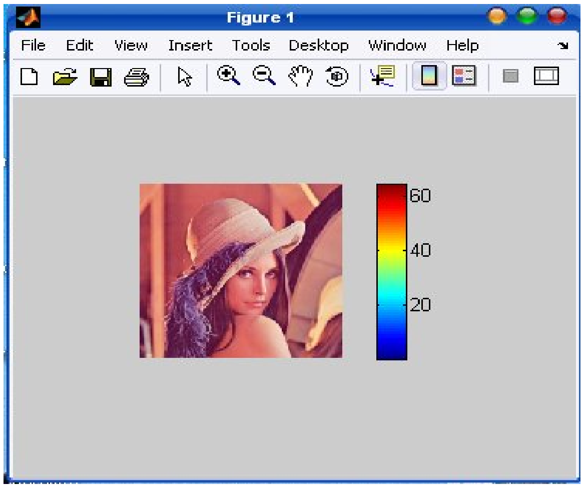


Fig 7.2 Single level Decomposition of query image

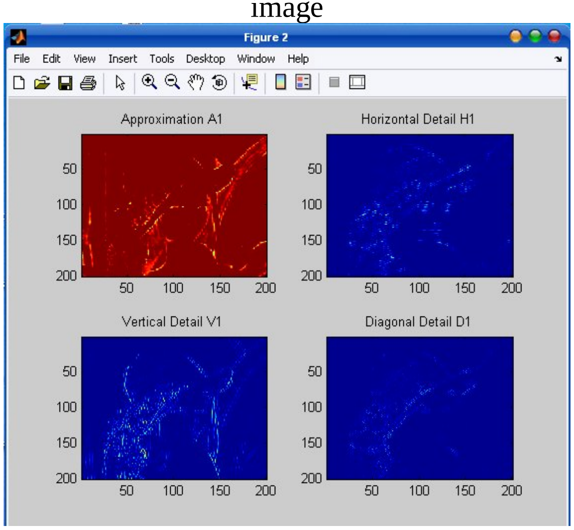
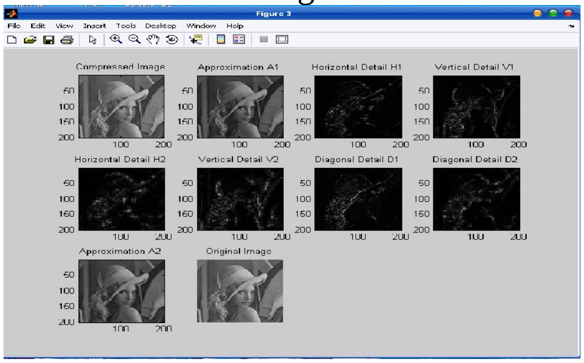


Fig. 7.3 Multilevel Decomposition of query image



5. CONCLUSION

Although the proposed framework has good consistency with subjective perception values and the objective assessment results can well reflect the visual quality of images, there are still some issues deserve to be further investigated in the future.In the future, we will further the proposed MGA-based framework to VQA by integrating the merits of the spatial temporal information and the human perception.

REFERENCES

[1] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–213, 2002..

[2] B. Chitprasert and K. R. Rao, "Human visual weighted progressive image transmission," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 1040–1044, Jul. 1990..

[3] R. Eslami and H. Radha, "Wavelet-based contourlet transform and its application to image coding," in *Proc. IEEE Int. Conf. Image Processing*,Piscateway, NJ, 2004, pp. 3189–3192..

[4] X. Li, "Blind image quality assessment," in *Proc. IEEE Int. Conf.Image Processing*, New York, 2002, vol. 1, pp. 449–452.

[5] S. Mallat, "A theory for multiresolution decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7,pp. 674–693, Jul. 1989..

[6] E. Le Pennec and S. G. Mallat, "Image compression with geometrical wavelets," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC,Canada, 2000, pp. 661–664..

[7] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf,"An objective video quality assessment system based on human perception,"*SPIE Human Vision, Visual Processing, and Digital Display IV*, 1993.

VNC-based Real Time Multimedia System with integrated Bi-Direction Audio and Video

Anand kumar M

MTech (II year), Networking and Internet Engg, SJCE, Mysore

Department of Electronics and Communications, Visveswaraya Technological University

anandmhalli@gmail.com

***Abstract*— Recently, VNC (Virtual Network Computing) has become one of the popular remote access techniques for Internet applications such as remote distant learning and collaborative application systems. However, there is no any multimedia functions have been implemented to enhance the performance.**

In this paper, I present a framework based on remote frame buffer (RFB) and real-time transport protocol (RTP) to provide real-time video and audio transmissions for multimedia applications. The framework can be applied to build distance collaborative environments with multimedia and RTP features efficiently.

In this implementation, I have used different compression criteria to evaluate the pro-posed system architecture. After adopting different types of compression algorithms and encoding length to estimate the bandwidth of Internet connection. Then I could obtain a proper way for my system to transfer. The experimental results show that the good transformation quality.

As a result, the proposed VNC based multimedia system is not only very suitable for a collaborative environment but also highly suited for multimedia transmissions since its real-time characteristics. Finally, it is worth noticing that the presented VNC based multimedia architecture is especially suited for embedded systems, intelligent homes, and could be used for mobile hand-held devices due to its thin-client properties.

***Keywords*— Virtual Network Computing (VNC), Real-Time Transport Protocol (RTP), Multimedia, Audio, Video**

I. Introduction

With the rapid development of the Internet, the broadband network is widespread in the industries, schools, and homes. As a result, most application functionality moves back to the server and leaves only the user interface on the desktop of personal computers. Such share-screen technology can realize a mechanism for remote sharing and controlling, especially for the distant collaborative systems [20] [19] [4] [3] [11] [6].

There are a lot of products using the share-screen technology, such as distant service of Windows 2000, PC Anywhere, Remote Administrator, Virtual Network Computing (VNC) and NetMeeting [19]. VNC is the most popular mechanism among all due to its open source property and can be run on different platforms. Therefore, it is a suitable choice for constructing the collaborative environment [13] [16]. VNC based collaborative application systems enable a server's desktop be shared by a group of VNC clients for collaboration. The cooperative participants can use the VNC virtual desktop to share and control the server's desktop. Recently, there have been few attempts to add video or audio mechanism to improve such distant collaborative systems [5] [18] [10] [4] [14] [15]. Thus, my research is motivated by the observation of lack of support for remote sharing like video or audio transmission.

In this master thesis, I propose a high efficiency and high-quality VNC based multimedia architecture, which is optimized by taking advantage of certain properties of the RTP (Real-Time Transport Protocol) algorithm and its implementations [8]. Based on the remote frame buffer property of VNC, I optimize and add both of audio and video functions by ignoring some features of hardware acceleration and using the mixer devices as sound input and output. Experimental results show that the presented VNC based multimedia architecture retains the good transformation quality for transmission in general environment of network.

This master thesis is organized as follows. Chapter 2 briefly introduces the back-ground of the VNC, remote frame buffer (RFB) and its operations. In Chapter 3 we first give an introduction for the Real-Time Transport Protocol (RTP) and Real-Time Transport Control Protocol (RTCP). In Chapter 5 we will present the proposed VNC based multimedia system. Finally, Chapter 6 concludes this master thesis with future work.

II. Virtual Network Computing

In this Chapter I briefly review the VNC background and discuss the thin client features. I also discuss the remote frame buffer protocol which makes the client stateless. Later an example discussing about the operations of the server and the clients is given. Lastly I give some applications as examples which are based on VNC technology.

A. *The VNC Background*

In computing, Virtual Network Computing (VNC) is a graphical desktop sharing system which uses the Remote Frame Buffer (RFB) protocol to remotely control another computer [17]. It transmits the keyboard and mouse events from one computer to another, relaying the graphical screen updates back in the other direction, over a network.

VNC is platform-independent; it means that VNC viewer on any operating system usually connects to a VNC server on any other operating system. There are clients and servers for almost all GUI operating systems and for Java. In addition, multiple clients may connect to a VNC server at the same time. Popular uses for this technology include remote technical support and accessing files on one's working computer from one's home computer, or vice versa.

VNC was originally developed at the Olivetti Research Laboratory in Cambridge, England. The original VNC source code and many modern derivatives are open source under the GNU General Public License.

B. *Thin Client Protocol*

A thin client (sometimes also called a lean client) is a client computer or client software in client-server architecture networks which depends primarily on the central server for processing activities, and mainly focuses on conveying input and output between the user and the remote server. In contrast, a thick or fat client does as much processing as possible and passes only data for communications and storage to the server.

The Virtual Networking Computing (VNC) system is a thin client system. Like all such systems, it reduces the amount of state maintained at the user terminal. VNC viewers are exceedingly thin because they do not store any unrecoverable state at the endpoint. This system contrast is the same as X Windows, and allows arbitrary disconnection and reconnection of the client without effects on the session at the server. Since the client can reconnect at a different location even on the other side of the planet, VNC achieves mobile computing without requiring the user to carry computing hardware.

C. *Remote Frame Buffer Protocol*

RFB (Remote frame buffer) is a simple protocol for remote access to graphical user interfaces.

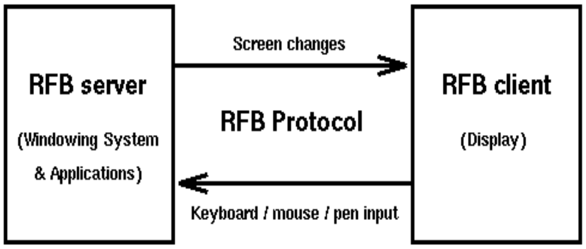

Screen changes
RFB server
(Windowing System & Applications)
RFB Protocol
RFB client
(Display)
Keyboard / mouse / pen input

Fig.1 VNC Architecture

Fig.1 shows the VNC architecture which uses RFB protocol. Because it works at the framebuffer level it is applicable to all windowing systems and applications, including X11, Windows and Macintosh. RFB is the protocol used in VNC (Virtual Network Computing). The remote endpoint where the user sits (i.e. the display plus keyboard and/or pointer) is called the RFB client or viewer. The endpoint where changes to the framebuffer originate (i.e. the windowing system and applications) is known as the RFB server.

RFB is truly a thin client protocol. The emphasis in the design of the RFB protocol is to make very few requirements of the client. In this way, clients can run on the widest range of hardware, and the task of implementing a client can be made very simple.

The protocol also makes the client stateless. If a client disconnects from a given server and subsequently reconnects to that server, the state of the user interface is preserved. Furthermore, a different client endpoint can be used to connect to the same RFB server. At the new endpoint, the user will see exactly the same graphical user interface as at the original endpoint. In effect, the interface to the user's applications becomes completely mobile. Wherever suitable network connectivity exists, the user can access their own personal applications, and the state of these applications is preserved between accesses from different locations. This provides the user with a familiar, uniform view of the computing infrastructure wherever they go.

D. *Operations of VNC Server and VNC Client*

Screen update on the server side adopting the RFB protocol. The server sends small rectangles of the framebuffer to the client. In its simplest form, the VNC protocol can use a lot of bandwidth, so various methods have been devised to reduce the communication overhead. For example, there are various encodings (methods to determine the most efficient way to transfer these rectangles). The VNC protocol allows the client and server to negotiate which encoding will be used. The simplest encoding, which is supported by all clients and servers, is the raw encoding where pixel data is sent in left-to-right scanline order, and after the original full screen has been transmitted, only transfers rectangles that change. This encoding works very well if only a small portion of the screen changes from one frame to the next (like a mouse pointer moving across a desktop, or text being written at the cursor), but bandwidth is very demanding get very high if a lot of pixels change at the same time, such as when scrolling a window or viewing full-screen video.

VNC by default uses TCP ports 5900 through 5906, each port corresponding to a separate screen (5900 to 5906). A Java viewer is available in many implementations such as Rea1VNC on ports 5800 through 5806, allowing clients to interact through, among other things, a Java-enabled web browser. Other ports can be used as long as both client and server are configured accordingly.

Using VNC over the Internet works well if the user has a broadband connection at both ends. However, it may require advanced NAT, firewall and router configuration such as port forwarding in order for the connection to go through. Some users may turn to use instant private networking applications such as Remobo or VPN (Virtual Private Network) applications such as Hamachi to make usage over the Internet much easier. Besides, Remobo also adds an additional layer of encryption for enhanced security.

Note that on some machines, the server does not necessarily have to have a physical display. Xvnc is the Unix VNC server, which is based on a standard X server. Xvnc can be considered as two servers in one; to applications it is an X server, and to remote VNC users it is a VNC server.

Applications can display themselves on Xvnc as if it was a normal X display, but they will appear on any connected VNC viewers rather than on a physical screen.

In addition, the display that is served by VNC is not necessarily the same display seen by a user on the server. On Unix/Linux computers that support multiple simultaneous X11 sessions, VNC may be set to serve a particular existing X11 session, or to start one of its own. It is also possible to run multiple VNC sessions from the same computer. On Microsoft Windows the VNC session served is always the current user session.

E. *VNC Applications*

VNC is commonly used as a cross-platform remote desktop system. For example, Apple Remote Desktop for Mac OS X interoperates with VNC and will connect to a Linux user's current desktop if it is served with xllvnc, or to a separate X11 session if one is served with TightVNC. From Linux, TightVNC will connect to an OS X session served by Apple Remote Desktop if the VNC option is enabled, or to a VNC server running on Microsoft Windows. In this thesis, we adopt ultraVNC that runs on Microsoft Windows system as our experimental environment owing to its popularity and convenience.

III. Real-Time Transport Protocol

Internet is changing: Static content is giving way to streaming video, text is being replaced by music and the spoken word, and interactive audio and video is becoming commonplace. These changes require new applications, and they pose new and unique challenges for application designers [12] [7] [18].

F. *Networking Background*

To transmit the audio data, TCP is not a good choice, it has a lot of features which are unnecessary for audio transmission over the Internet, and which increase the overall delay. On the other hand, UDP itself is too simple to carry audio data over Internet. To transmit audio data should provide information for synchronization, flow and congestion control and identification. As a result, it is considered as the most suitable way to use RTP in the TCP/IP architecture.

G. *Real-Time Transport Protocol*

The Real-time Transport Protocol (RTP) defines a standardized packet format for delivering audio and video over the Internet. It was developed by the Audio-Video Transport Working Group of the IETF and first published in 1996 as RFC 1889 which was made obsolete in 2003 by RFC 3550 [1]. Real time transport protocol can also be used in conjunction with RTSP (Real-Time Streaming Protocol) protocol which enhances the field of multimedia applications.

RTP does not have a standard TCP or UDP port on which it communicates. The only standard that it obeys is that UDP communications are done via an even port and the next higher odd port is used for RTP Control Protocol (RTCP) communications. Although there are no standards assigned, RTP is generally configured to use ports 16384-32767. RTP can carry any data with real-time characteristics, such as interactive audio and video. Call setup and tear-down for VoIP applications is usually performed by either SIP or H.323 protocols. The fact that RTP uses a dynamic port range makes it difficult for it to traverse firewalls. In order to get around this problem, it is often necessary to set up a STUN (Simple traversal of UDP over NATs) server.

RTP was originally designed as a multicast protocol, but has since been applied in many unicast applications. RTP is frequently used in streaming media systems (in conjunction with RTSP) as well as video conferencing and push to talk systems (in conjunction with H.323 or SIP), making it the technical foundation of the Voice over IP industry. RTP goes along with the RTCP and is built on top of the User Datagram Protocol (UDP). Applications using RTP are less sensitive to packet loss, but typically very sensitive to delays. Thus, UDP is a better choice than TCP for such applications.

H. *Real-Time Transport Control Protocol*

Real-time Transport Control Protocol (RTCP) is a sister protocol of the Real-time Transport Protocol (RTP). It is defined in RFC 3550 (which obsoletes RFC 1889).

RTCP provides out-of-band control information for an RTP flow. It partners RTP in the delivery and packaging of multimedia data, but does not transport any data itself. It is used periodically to transmit control packets to participants in a streaming multimedia session. The primary function of RTCP is to provide feedback on the quality of service being provided by RTP.

RTCP gathers statistic information on a media connection such as bytes sent, packets sent, lost packets, jitter, feedback and round trip delay. An application may use this information to increase the quality of service, perhaps by limiting flow or using a different codec.

There are several types of RTCP packets: Sender report packet, Receiver report packet, Source Description RTCP Packet, Goodbye RTCP Packet and Application Specific RTCP packets.

I. *Media Transport using RTP*

Behaviour of a Sender:

A sender is responsible for capturing audiovisual data, whether live or from a file, compressing it for transmission, and generating RTP packets. It may also participate in error correction and congestion control by adapting the transmitted media stream in response to receiver feedback.

The sender starts by reading uncompressed media data - audio samples or video frames - into a buffer from which encoded frames are produced. Frames may be encoded in several ways depending on the compression algorithm used, and encoded frames may depend on both earlier and later data. The next section, Media Capture and Compression, will describe this process.

Media Capture and Compression:

The media capture process is essentially the same whether audio or video is being transmitted: An uncompressed frame is captured, if necessary it is transformed into a suitable format for compression, and then the encoder is invoked to produce a compressed frame. The compressed frame is then passed to the packetization routine, and one or more RTP packets are generated. Factors specific to capture audio and video we only discuss the part of audio capture. Video capture that we adopt the method of VNC remote frame buffer.

Considering the specifics of audio capture, sound being captured, digitized, and stored into an audio input buffer via mixer device. This input buffer is commonly made available to the application after a fixed number of samples have been collected. Most audio capture APIs return data from the input buffer in fixed-duration frames, blocking until sufficient samples have been collected to form a complete frame. This imposes some delay because the first sample in a frame is not made available until the last sample has been collected. If given a choice, applications intended for interactive use should select the buffer size closest to that of the codec frame duration, commonly either 20 milliseconds or 30 milliseconds, to reduce the delay.

Uncompressed audio frames can be returned from the capture device with a range of sample types and at one of several sampling rates. Common audio capture devices can return samples with 8-, 16-, or 24-bit resolution, using linear, u-law or A-law quantization, at rates between 8000 and 96000 samples per second, and in mono or stereo. Depending on the capabilities of the capture device and on the media codec, it may be necessary to convert the media to an alternative format before the media can be used. For example, changing the sample rate or converting from linear to u-law quantization.

Captured audio frames are passed to be encoded for compression. Depending on the codec, state may be maintained between frames - the compression context - that must be made available to the encoder along with each new frame of data. Some codecs, particularly music codecs, base their compression on a series of uncompressed frames and not on uncompressed frames in isolation. In these cases the encoder may need to be passed several frames of audio, or it may buffer frames internally and produce output only after receiving several frames. Some codecs produce fixed-size frames as their output; other produce variable-size frames. Those with variable-size frames commonly select from a fixed set of output rates according to the desired quality or signal content; very few are truly variable-rate.

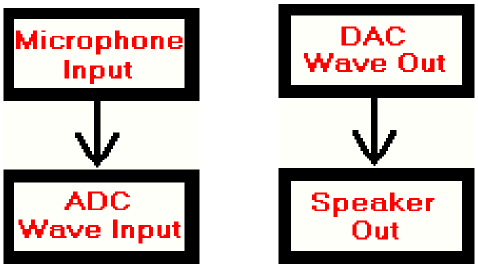
Generating RTP Packets:

As compressed frames are generated, the are passed to the RTP packetization routine. Each frame has an associated timestamp from which the RTP timestamp is derived. If the payload format supports fragmentation, large frames are fragmented to fit within the maximum transmission unit of the network. Finally, one or more RTP packets are generated for each frame, each including media data an any required payload header. The format of the media packet and specification for the codec used. The critical parts to the packet generation process are assigning timestamps to frame, fragmenting large frames, and generating the payload header.

Behaviour of a Receiver:

A receiver is responsible for collecting RTP packets from the network, repairing and correcting for any lost packets, recovering the timing, decompressing the media, and presenting the result to the user. In addition, the receiver is expected to send reception quality reports so that the sender can adapt the transmission to match the network situation. The receiver will also typically maintain a database of participants in a session to be able to provide the user with information on the other participants.

The first step of the reception process is to collect packets from the network, validate them for correctness, and insert them into a per-sender input queue. This is a straightforward operation, independent of the media format. The next section describes this process.



Packet Reception:

A RTP session comprises both data and control flows, running on distinct ports. This means that a receiving application will open two sockets for each session: one for data, one for control. Because RTP runs above UDP/IP, the sockets used are standard *SOCK_DGRAM* sockets, as provided by the Berkeley sockets API on UNIX-like systems, and by Winsock on Microsoft platforms.

The first stage of the media playout process is to capture RTP data packets from the network, and to buffer those packets for further processing. As packets are received, they are validated for correctness, their arrival time is noted, and they are added to a per-sender input queue, sorted by RTP timestamp, for later processing.

In parallel with the arrival of data packets, an application must be prepared to receive, validate, process, and send RTCP control packets.

Adapting the Playout-buffer, Decode, and Playout:

The final stages of the playout process are to decode the compressed media, mix media streams together if there are fewer output channels than active sources, and finally play the media to the user. This section considers each stage in turn.

For each active source the application must maintain an instantiation of the media decoder, comprising the decompression routines along with state known as the compression context. The decoder may be an actual hardware device or a software function, depending on the system. It converts each compressed frame into uncompressed media data, on the basis of the data in the frame and the compression context. As each frame is decoded, the compression context for the source is update.

The presence of accurate state in the decompression context is fundamental for correct operation of the decoder. The codecs will produce incorrect results if the context is missing or damaged. This is most often an issue if some data packets are lost because there will be a frame that cannot be decoded. The result will be a gap in the playout where that frame should have been, but the decompression context will also be invalidated and the following frames will be corrupted.

Depending on the codec, it may be possible to feed it an indication that a frame has been lost, allowing the decoder to better repair the context and reduce the damage on the media stream.

The process by which audio is played to the user is typically asynchronous, allowing the system to paly one frame of audio while processing the next. This capability is essential to normal operation because it allows continuous playback even though the application is busy with RTP and media processing. It also shields the application from variations in the behaviour of the system, perhaps due to other application from applications running on that system.

IV. Analysis of System Architecture

In this thesis, a two-phase study was designed to explore the VNC-based multimedia system. In order to learn more about the design philosophy, in this chapter, we give an analysis of VNC software and media capture through mixer device as the first phase.

J. *Analysis of VNC software*

The remote endpoint where the user sits (i.e. the display plus keyboard and/or pointer) is called the VNC clients or viewers. The endpoint where changes to the framebuffer originate is known as the VNC server. Fig.1 shows server and client delivered the RFB protocol message.

K. *Media Capture via Mixer device*

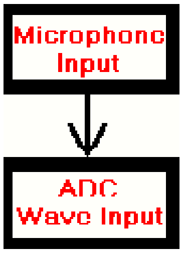


Fig. 2 Microphone input to ADC wave input

We can capture audio and media data via different mixer devices as record input. Let's consider a typical, basic audio card. First of all, if the audio card is capable of recording digital audio, then it typically has a microphone input jack (with some sort of pre-amp), and it also has an analog-to-digital converter (ADC) to convert an analog microphone signal to a digital stream. Therefore, it has two components -the Microphone input component, and the ADC component. The Microphone input is piped into the ADC. So, we can represent the layout with the following block diagram showing two components, with the signal flow between them (i.e., the arrow) in Fig.2.

Fig.3 A typical audio card with playing and recording capability

A typical audio card is also capable of playing back digital audio, so it has a DAC to convert the digital stream back to an analog signal, and also it has a speaker output jack (i.e., with some sort of analog amplifier). Therefore, it has two more components - the DAC component, and the Speaker component. The DAC output is piped to the speakers.

A typical audio card may have some other components. For example, it may have a built-in sound module (i.e., synth) capable of playing MIDI data. The audio output of this component would typically be piped to the speaker output just like the DAC. So, our block diagram now looks like Fig.3.

V. Multimedia system Implementation

This system is based on ultraVNC open source that runs on the Windows operating system [2]. Fig.4 shows the improved ultraVNC architecture by adopting multi-media functions, which is powerful, easy to use and free software that can display the screen of another computer on client's own screen. It allows us to use the client's mouse and keyboard to control the other PC remotely. It means that we can work on a remote compute as if we were sitting in front of it, right from your current location. If we provide computer support, we can quickly access server's computers from anywhere in the world and resolve helpdesk issues remotely. We do not even have to pre-install software or execute complex procedures to get remote helpdesk support. RTP concepts have been introduced in Chapter 3. We adopt the open source project of jori's RTP library (jrtplib) [9], which has been widely used for many multimedia applications, to implement the purposed multimedia system. It is an object-oriented RTP library, written in C++. In our system we can transfer video and audio for remote sharing and remote control fields by jrtp library.

For the video, in order to preserve performance we use the way of capturing the screen images that VNC server used, but the overlay technology designed by Microsoft Widows system makes us hard to capture the video played on the server screen. Here we disable the acceleration option of the graphic card to cope with the problem to obtain a high-quality transmission.

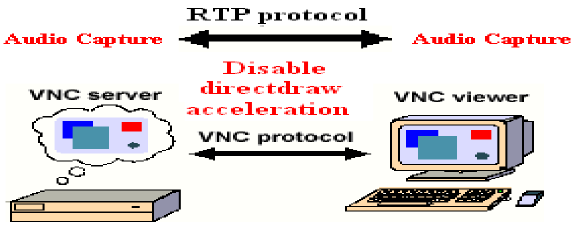


Fig.4 The proposed improved VNC architecture.

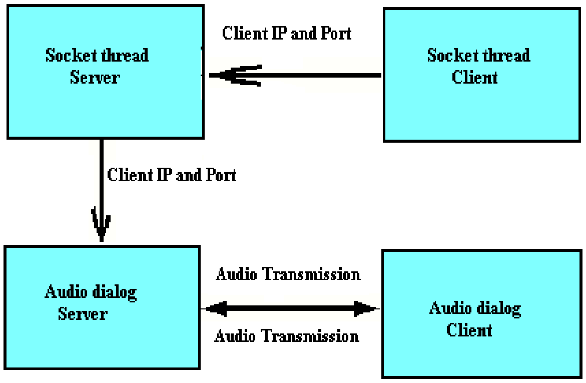


Fig.5 Conceptual flow of audio transmission.

For the audio, Fig.5 shows the audio transmission flow. First, the server creates an audio session thread waiting for clients to connect. When an audio session is created on the client side, it will send the IP address, port number, and audio flag to server side. It should be noted that the transmission will also be started at the same time. After the server received the client's information, it will add the client's information to server's audio session thread. Subsequently, it gives the message of client's information to audio dialog.

The audio dialog will start the audio transmission to the client dynamically. If the client left or destroyed the audio session, the server will remove the client from the transmission list.

Both of server and client have some parameters need to be configured, like sampling rates, encoding types, compression types, and record device options.

For the general parameters, the input and output sample rates could be chosen among 4000, 8000, 11025, 22050, and 44100 bits per second. 8000 bps makes voice sound like telephone. If we want better quality, we could configure 44100 bps that would perform as CD-like quality. 8 or 16 bit of encoding length for Input and output encoding that larger bit length makes voice sound more accurate. The sample interval sets the time interval to record sound data. Port base sets the port to receive the audio data. We can choose different devices for input or output audio data and the compression type could be set in the component fields.

Different compression types, like DPCM compression, Mu-law encoding, LPC compression, GSM 13 kbps compression, and Speex compression, could be chosen for different compression rates. We integrate different compression algorithms in this system that we could choose one of these compression algorithms according to the network transmission bandwidth. Record devices can be configured as microphone or system sound which recording sound from microphone or system sound, respectively. We can choose different configurations according to the network bandwidth.

After setting all basic configurations, we can create an audio session. Note that due to different protocols are employed in video and audio transmissions respectively, synchronization problem should be taken into account. We add the time-stamp information on each video frame and audio packet to solve. Video and audio can be displayed and played on both server and client sides efficiently.

VI. Conclusion and Future Work

This paper proposed a new real-time multimedia system based on VNC technology. We incorporated many useful open source projects and algorithms for this system and proposed new methods for integrating audio and video functions into this VNC-based multimedia system. Although only modeling the ultraVNC on Windows system is studied, the technologies can be also applied to other platforms.

Nevertheless, there are still several works should be done in the future. First, more compression algorithms should be tested and evaluated. Second, this current architecture can be applied to some intelligent home applications. The potential of its use in intelligent home applications clearly needs further exploration. We can implement the architecture on different platforms and make this model platform-independent. Third, remote frame buffer protocol makes video frame played not very smoothly.

I am now proceeding for further development of above works and tests to obtain higher performance of the proposed system for distance collaborative environments with multimedia and real-time features efficiently.

References

[1] *Real-time protocol :* http://www.ietf *org/rfc/rfc3550.t.*

[2] *ultravnc* : http://www.uvnc.com/.

[3] Bourov S. Eckerlin G.-Elsen E. Bacher, R. and Kammering, *Remote control and monitoring of accelerators and detectors in a global facility (gan/gdn),* Proc. 15th IEEE-NPSS Real-Time Conference, April 29 2007-May 4 2007, pp. 1-4.

[4] J. Brooke, T. Eickermann, and U. Woessner, *Application steering in a collaborative environment,* Proc. ACM/IEEE Conference Supercomputing, 15-21 Nov. 2003, pp. 61-61.

[5] C.Y.Y. Cheng and J. Yen, *Virtual learning environment (vie): a web-based collaborative learning system,* Thirty-First Hawaii International Conference on System Sciences, vol. 1, Jan. 1998, pp. 480-491.

[6] P.M. Corcoran, F. Papal, and A. Zoldi, *User interface technologies for home appliances and networks,* IEEE Trans. Consumer Electron. 44 (1998), no. 3, 679-685.

[7] M. Herscher and R. Cox, *Voice programming of numerically controlled machines,* Proc. IEEE International Conference on ICASSP '77. Acoustics, Speech, and Signal Processing, vol. 2, May 1977, pp. 452-455.

[8] Wim Lammotte Jori Liesenborgs and Frank Van Reeth, *Voice over ip with jvoiplib and jrtplib,* IEEE Annual Conference on Local Computer Networks, 2001, pp. 346-347.

[9] Jori's RTP library, http://research.edm.uhselt.be/jori/page/ *index.php.*

[10] Kaplinsky and K.V., *Vnc tight encoder-data compression for vnc,* Proc. 7th International Scientific and Practical Conference of Students

Post-graduates and Young Scientists Modern Techniques and Technology, Feb. 2001, pp. 155-157.

[11] K.V. Kaplinsky, *Vnc tight encoder-data compression for vnc,* Proc. 7th International Scientific and Practical Conference of Students Post-graduates and Young Scientists Modern Techniques and Technology MTT 2001, 26 Feb.-2 March 2001, pp. 155-157.

[12] Colin Perkins, *Rtp : audio and video for the internet,* Addison-Wesley, 2003.

[13] F. Papal P.M. Corcoran and A. Zoldi, *User interface technologies for home appliances and networks,* IEEE Transactions on Consumer Electronics, vol. 44, Aug. 1998, pp. 679-685.

[14] Anthony Steed Oliver Otto Robin Wolff, Dave J. Roberts, *A review of telecollaboration technologies with respect to closely coupled collaboration,* International Journal of Computer Applications in Technology, vol. 29, July 2007, pp. 11-26.

[15] V. Schmidt and J.A. How, *Remote participation infrastructure in the european fusion laboratories*, IEEE Transactions on Nuclear Science, vol. 49, Apr. 2002, pp. 532-536.

[16] T. Hase T. Mizuno T. Haraikawa, T. Sakamoto and A. Togashi, *Vnc: a proposal for internet connectivity and interconnectivity of home appliances based on remote display framework,* IEEE Transactions on Consumer Electronics, vol. 44,Aug. 2002, pp. 512-519.

[17] Q. Stafford-Fraser K.R. Wood T. Richardson, T. Richardson and A. Hopper, *Virtual network computing,* IEEE Internet Computing, vol. 2, Jan. 1998, pp. 33-38.

[18] Su Tzong-An, *Distant vcr - a virtual classroom for distance learning,* Proc. First IEEE International Symposium on Information Technologies and Applications in Education ISITAE 2007, 2007, pp. 174-178.

[19] Lu Xiaolin, *Construct collaborative distance learning environment with vnc technology,* International Conference on Semantics, Knowledge and Grid, Nov. 2005, pp. 127-129.

[20] Lu Xiaolin and Lu Xiaolin, *Wsfrb protocol and virtual program computing,* Proc. 8th International Conference on Computer Supported Cooperative Work in De-sign, vol. 1, May 2004, pp. 475-480.

AN FPGA IMPLEMENTATION OF HIGH SPEED FLEXIBLE TURBO CODER

D.srinivas[1], Ch.Rajasekhar[2], G.Govardhan[3]
[1].Dept.of ECE, GMRIT, Rajam. E-Mail ID: srinivasa.dasari@gmail.com
[2].Dept.of ECE ,GIT,Gitam University, Visakhapatnam-530045. E-Mail ID: chukkarajasekhar@gmail.com
[3].Student, Dept.of ECE ,GIT,Gitam University, Visakhapatnam-530045.

Abstract - ***Turbo coding represents a new and very powerful error control technique, which has started to have a significant impact in the late 90s, allowing communication very close to the channel capacity. The powerful error correction capability of turbo codes was recognized and accepted for almost all types of channels leading to increased data rates and improved Quality of Service. Flexible channel coding has received much attention recently, and it is a powerful scheme providing high reliability and high spectral efficiency over rain-fading channels.***

In this work, the turbo coding system model is studied and all the modules are implemented in VHDL. The major components of the turbo coding system are encoder and decoder. The encoder consists of two half-rate recursive systematic convolutional encoders separated by an interleaver. The decoder consists of two iterative decoders, deinterleaver and demux.The proposed high-speed decoding algorithms for solving the latency and complexity are presented in this paper.

I. INTRODUCTION

Wireless technology is fast becoming a trend in present communication systems, as the demand for greater bandwidth allocation is being addressed by fixed wireless broadband access. However, the use of free space, as a medium, introduces many sources of error in the transmission of data across the channel. With accuracy of information being very critical, the use of traditional Forward Error Correction (FEC) methods has become widespread.

FEC provides a significant improvement to the system in terms of reliability of data reception. The basic principle behind error-correcting codes is the application of a mathematical transform onto the message signal such that redundant message information is used to correct any errors that may have been introduced during transmission.

Turbo coding is a forward error correction (FEC) scheme. Iterative decoding is the key feature of turbo codes .Turbo codes consist of concatenation of two convolution codes. Turbo codes give better performance [8] at low SNRs (signal to noise ratio) interestingly, the name Turbo was given to this codes because of the cyclic feedback mechanism (as in Turbo machines) to the decoders in an iterative manner. Turbo codes can be concatenated in series, parallel or in a hybrid manner. Concatenated codes can be classified as parallel concatenated convolution codes (PCCC) or serial concatenated convolutional codes (SCCC).

In this work, a single turbo decoder is presented to decode all modulation schemes in order to need less hardware and less power consumption, and to reduce the receiver cost. Important issues in high speed applications of turbo decoder are decoding delay and computational complexity. To solve the latency and complexity four high-speed decoding algorithms are presented [1].

The presentation of this paper is as follows: Section II provides a detailed description of the blocks comprising the system; Section III gives the description of system model and proposed decoding algorithms, the VHDL simulations and results will be discussed in Section IV; the conclusion and recommendations will follow in Section V.

II. TURBO CODING SYSTEM

2.1 Turbo Encoder

A turbo encoder basically consists of parallel concatenation of two identical recursive systematic convolutional encoders separated by an interleaver. The general turbo encoder block diagram is shown in the figure 2.1.

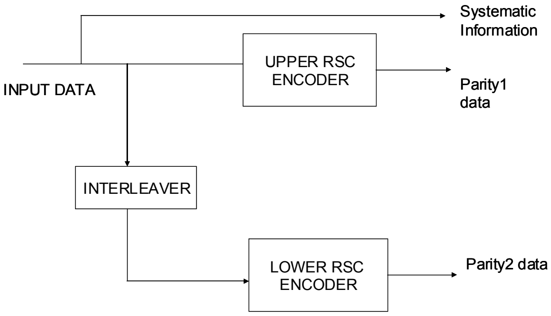


Fig 2.1: Turbo Encoder

In the design of turbo codes, a special kind of convolutional encoder named recursive systematic convolutional (RSC) encoder is used. Systematic means that part of the encoded bits is the same as the input bits.

Recursive means that the registers in the generator has a feedback loop.

The interleaver is used to rearrange the bits in the input sequence such that the two constituent encoders are operating on two input sequences consisting of the same bits, but in different order. The interleaver design is the crucial part in the design of turbo coder. Different interleavers perform differently in the turbo coders.

2.2 Turbo Decoder

The Turbo decoding is performed using a non-optimal Maximum a Posteriori (MAP) algorithm [3]. The Turbo decoder consists of two elementary decoders in a serial concatenation scheme. Since soft decoding performs better than hard decoding, the first decoder provides a weighted soft decision in the form of A Posteriori Probabilities (APPs) to the second decoder. The structure of the turbo decoder is shown in the fig2.2.

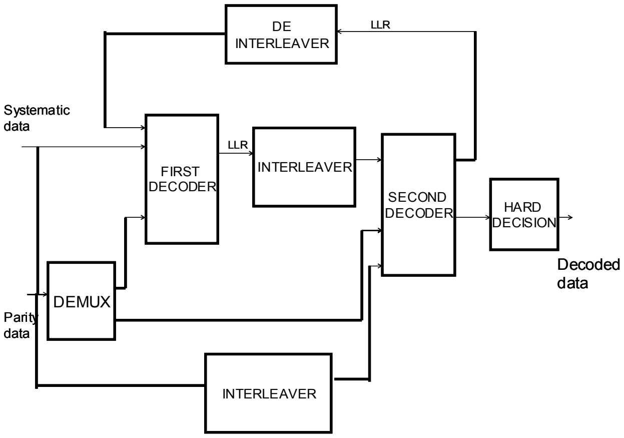


Fig 2.2: Turbo Decoder

The soft information from the second decoder is fed back to the first decoder, after the first iteration is complete. This is called the extrinsic or the a priori information. This information is not available for the first decoder during the first iteration and is therefore initialized to zero. The soft information is exchanged between the two decoders until the desired performance level is achieved.

III.PROPOSED SYSTEM MODEL

3.1 System Model

The proposed high-speed flexible decoding architecture is shown in the Fig 3.1. The proposed high-speed decoder can support both a half-rate turbo decoder BPSK modulation scheme and a two-thirds rate 8-PSK modulation scheme. The phase information is given by the phase sector quantizer (PSQ).

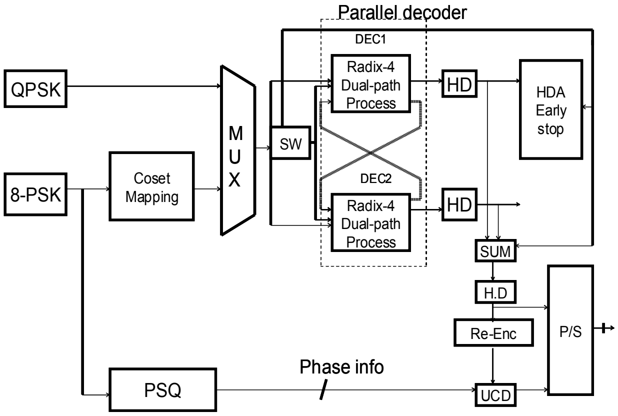


Fig 3.1: System Model

It consists of parallel decoder comprises of two decoders dec1 and dec2.The decoders performs the decoding operation based on the algorithms proposed. The decision of the final decoding information is made by using early stop algorithm

3.2 Proposed High-Speed decoding algorithms

Since convolutional turbo codes are very flexible and are easily adapted to a large number of block sizes and coding rates, they have been adopted in the DVB-RCS standard. However, the applications of turbo codes are limited to low data-rate services because of the decoding speed limitation. Therefore, it is highly desirable to develop a high-speed turbo decoder.

To solve the latency problem of a turbo decoder, four algorithms are proposed: the radix-4 algorithm, the dual-path processing algorithm, the full parallel decoding algorithm, and the early-stop algorithm [1] based on the hard-decision-aided (HAD) scheme. The decoding iteration progresses until a certain stopping condition is satisfied. Then, hard decisions are made based on the reliability measures of the decoded symbol at the last decoding iteration

3.2.1. Radix-4 Algorithm:

In the radix-4 decoding algorithm, the previous state at $t=k–2$ goes forward to the current state at $t=k$, and the reverse state at $t=k+2$ goes backwards to the current one such that the time interval from $t=k–2$ to $t=k$ is merged at time $t=k$. Therefore, we can decode two source data bits at the same time without any performance degradation while reducing the block size buffered in memory. Using the unified approach to state metrics, a 2^{v-1} –state trellis can be iterated from time index *n–k* to *n* by decomposing the trellis into 2^{v-k} sub-trellises, each consisting of *k* iterations of a 2^k –state trellis.

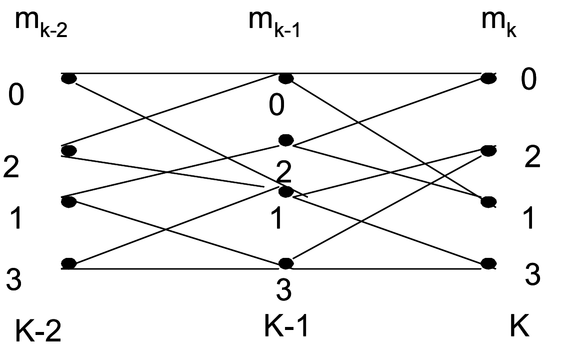


Fig 3.2a: 4-state radix-2 trellis

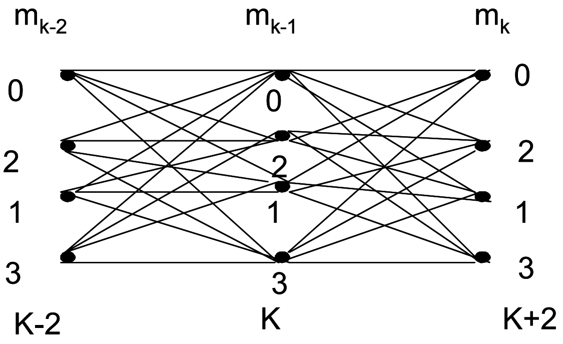


Fig 3.2b: 4-state radix-4 trellis

Each 2^k –state's sub-trellis can be collapsed into an equivalent one-stage radix-2^k trellis by applying *k* levels of look-ahead for the recursive update. Collapsing the trellis does not affect the decoder performance since there is a one-to-one mapping between the collapsed trellis and the radix-2 trellis. An example of decomposition of a 4-state radix-2 into an equivalent radix-4 trellis using one stage of look-ahead is shown in Fig 3.2. Where *v*=4, g_1= (15) octal, g_2= (17) octal with *v* denoting the constraint length.

3.2.2Dual-Path Processing Algorithm:

In a conventional scheme, the decoder must wait for the backward state metric (BSM) or forward state metric (FSM) calculation to be finished before calculating the extrinsic information. The dual-path processing method does not need to wait. The decoder calculates the FSM (left to right), and BSM (right to left), simultaneously. When the FSM and BSM reach the same point, the decoder begins to calculate the extrinsic information. Fig3.3 shows the operation of the dual-path processing.

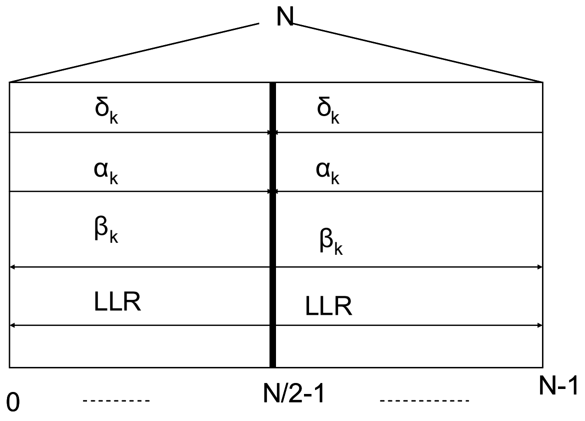


Fig 3.3 Dual path processing algorithm

3.2.3 Parallel Decoding Algorithm:

Unlike the original turbo decoder consisting of two decoders concatenated in a serial fashion, the parallel decoder structure uses two decoders which operate in parallel and update each other simultaneously immediately after each one has completed its decoding. Unlike, to decode the estimated data, we use the sum of the LLR outputs of the parallel decoders to reduce the latency to one half while maintaining the same performance level. Fig 3.4 shows the parallel decoder.

3.2.4. Early Stop Algorithm:

The decoding iteration continues processing until a certain stopping condition is satisfied, then hard decisions are made based on the reliability measures of the decoded symbols at the last decoding iteration. The HDA algorithm is used as an early stop algorithm.

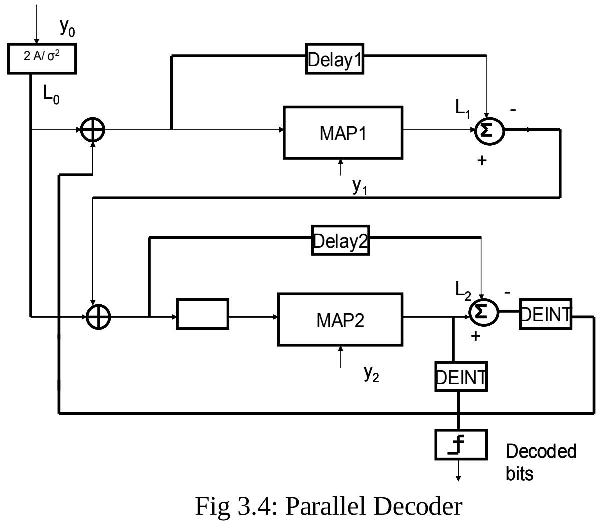
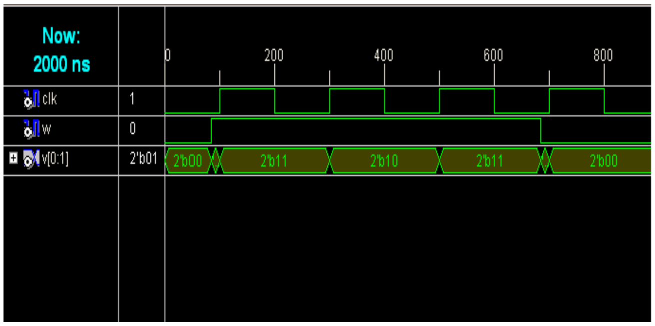


Fig 3.4: Parallel Decoder

IV.SIMULATION RESULTS

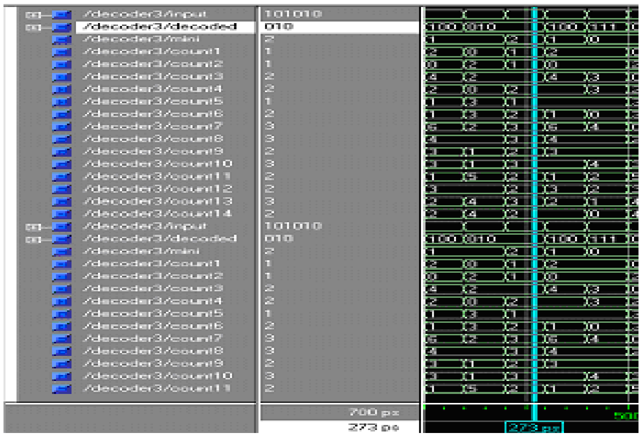
All the modules of the turbo coding system are synthesized and simulated using VHDL. The simulations results for two of the modules are given here.

<u>Recursive Systematic Convolutional Encoder</u>:
Constraint length=3, G (7, 5).



```
Now:
2000 ns
clk       1
w         0
v[0:1]    2'b01
```

<u>Viterbi Decoder:</u>



```
/decoder3/input      101010
/decoder3/decoded    010
/decoder3/mini       2
/decoder3/count1     1
/decoder3/count2     1
/decoder3/count3     2
/decoder3/count4     2
/decoder3/count5     1
/decoder3/count6     2
/decoder3/count7     3
/decoder3/count8     3
/decoder3/count9     2
/decoder3/count10    3
/decoder3/count11    2
/decoder3/count12    2
/decoder3/count13    3
/decoder3/count14    2
/decoder3/input      101010
/decoder3/decoded    010
/decoder3/mini       2
/decoder3/count1     1
/decoder3/count2     1
/decoder3/count3     2
/decoder3/count4     2
/decoder3/count5     1
/decoder3/count6     2
/decoder3/count7     3
/decoder3/count8     3
/decoder3/count9     2
/decoder3/count10    3
/decoder3/count11    2
                     700 ps
                     273 ps
```

V.CONCLUSION

In this work, the turbo coding system model is studied and all the modules are implemented in VHDL. The major components of the turbo coding system are encoder and decoder. To extend the application area of the turbo codes to real-time services, the latency and complexity are to be reduced. The high-speed decoding algorithms proposed will solve the problem of latency and complexity in the decoder with less hardware and less power consumption.

VI.REFERENCES

[1] "High-Speed Adaptive turbo decoding algorithm and its implementation," Duk gun choi, Jin hee jung, Min Hyuk Kim, Ji Won Jung. IEEE 2006.

[2] J.W. Jung et al., "Design and Architecture of Low-Latency High-Speed Turbo Decoder," ETRI Journal, vol. 27, no. 5, Oct. 2005, pp. 525-532.

[3] S.H. Yoon and Y. Bar-Ness, "A Parallel MAP Algorithm for Low Latency Turbo Decoding," Communications Letters, vol. 6, no. 7, Jul. 2002, pp. 288-290.

[4] "Low Latency Algorithms of Iterative Codes for Wireless Broadband Communication Systems," Duk Gun Choi, Jin Hee Jeong, Ji Won Jung. 2005 Asia-Pacific Conference on Communications, Perth, Western Australia, 3 - 5 October 2005.

[5] E.A. Choi, J.W. Jung, N.S. Kim, Y.I. Kim, and D.G. Oh, "A Simplified Decoding Algorithm Using Symbol Transformation for Turbo Pragmatic Trellis-Coded Modulation," ETRI Journal, vol. 27, no. 2, Apr. 2005, pp. 223-226.

[6] "Digital Video Broadcasting Standard for Return Channel via Satellite (DVB-RCS)," ETSI TR 101 790, vol. 2.1, 2003.

[7] "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes," Claude Berrou, Member, IEEE, and Alain Glavieux IEEE transactions on communications, vol. 44, no io, October 1996.

[8] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Code and Decoding: Turbo Codes," Proc.ICC'93, 1993.

Implementation of FPGA based LOW - COST Logic Signal Analyzer (LSA)

S NAGAKISHORE BHAVANAM
Hello: +91–99895 41444, e-mail : satyabhavanam@gmail.com
Guide: Mr. M. MADANA GOPAL
AURORAS TECHNOLOGICAL & RESEARCH INSTITUTE
Parvathapur, Uppal, Hyderabad-39, A.P, INDIA.

Abstract

Logic signal analyzers are very essential instrument for digital circuit or board debugging. The existing market solutions offer several features, but the cost of such instruments is very high and most of the time we don't need that much capable instruments. In this paper a low cost logic signal analyzer is implemented around the Spartan-3 FPGA. The FPGA being capable of offering high frequency data paths in them become suitable for realizing high frequency signal capturing logic.

The paper work includes development of FPGA based logic signal analyzer using VHDL. The logic signal analyzer will be capable of implementing match conditions, counter based triggering, external clocking and internal clocking features. The blocks such as registers, counters, comparators, state machines will be used in realizing these blocks. The captured data will be stored in memory before transferring the data to PC. The UART core will be developed which, will be used for transferring the data to PC.

Modelsim Xilnx edition (MXE) tools will be used for simulation. Xilinx FPGA synthesis tools will be used for synthesizing the design for Spartan FPGAs. The developed application will be tested on Spartan 3E development board. By using HyperTerminal we check our design. By using GUI ,we will show the results with respect to waveforms.

1. Introduction

A logic analyzer is an electronic instrument which displays signals of a digital circuit which are too fast to be observed and presents it to a user who can then precisely observe with greater ease the operation of the digital system under test. It is typically used for capturing data in systems having too many channels to be examined with an oscilloscope. Software running on the logic analyzer can convert the captured data into timing diagrams, protocol decodes, state machine traces, assembly language, or correlate assembly with source-level software.

Presently there are three distinct categories of logic analyzers available on the market:

i. The first is mainframes, which consist of a chassis containing the display, controls, control computer, and multiple slots into which the actual data capturing hardware is installed.
ii. The second category is standalone units which integrate everything into a single package, with options installed at the factory.
iii. The third category is PC-based logic analyzers. The hardware connects to a computer through a USB or LPT connection and then relays the captured signals to the software on the computer. These instruments are less expensive than either mainframes or standalone units although they lack the sophisticated functionality. These devices are typically much smaller, because they do not need displays or hardware input such as dials.

2. Logic analyzer operation

A logic analyzer may be triggered on a complicated sequence of digital events, and then capture a large amount of digital data from the system under test (SUT). The best logic analyzers behave like software debuggers by showing the flow of the computer program and decoding protocols to show messages and violations.

When logic analyzers first came into use, it was common to attach several hundred "clips" to a digital system. Later, specialized connectors came into use. The evolution of logic analyzer probe has led to a common footprint that multiple vendors support, which provides added freedom to end users. Introduced in April, 2002, connector less technology (identified by several vendor specific trade names: Compression Probing; Soft Touch; D-Max) has become popular. These probes provide a durable, reliable mechanical and electrical connection between the probe and the circuit board with less than 0.5pF to 0.7 pF loading per signal. Once the probes are connected, the user programs the analyzer with the names of each signal, and can group several signals into groups for easier manipulation. Next, a capture mode is chosen, either timing mode, where the input signals are sampled at regular intervals based on an internal or external clock source, or state mode, where one or more of the signals are defined as "clocks," and data is taken on the rising or falling edges of these clocks, optionally using other signals to qualify these clocks.

After the mode is chosen, a trigger condition must be set. A trigger condition can range from simple (such as triggering on a rising or falling edge of a single signal), to the very complex (such as configuring the analyzer to decode the higher levels of the TCP/IP stack and triggering on a certain HTTP packet).

At this point, the user sets the analyzer to "run" mode, either triggering once, or repeatedly triggering.

Once the data is captured, it can be displayed several ways, from the simple (showing waveforms or state listings) to the complex (showing decoded Ethernet protocol traffic). The analyzer can also operate in a "compare" mode, where it compares each captured data set to a previously recorded data set, and stopping triggering when this data set is either matched or not. This is useful for long-term empirical testing. Recent analyzers can even be set to email a copy of the test data to the engineer on a successful trigger.

3.Architectural design

The block diagram of logic signal analyser is shown in Figure 3-1.

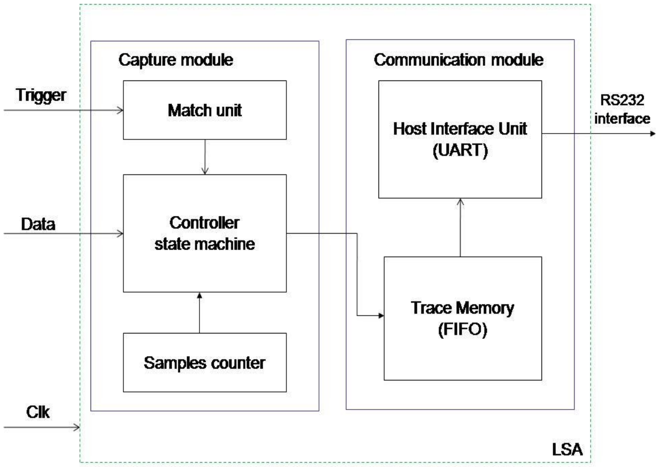


Figure 3-1. Block diagram of Logic Signal Analyser

The block diagram contains 2 basic modules namely capture module and communication module. Capture module is responsible for signal capture. It contains a capture state machine and samples counter. The counter is 8-bit counter. The communication module contains UART and FIFO. The FIFO (Trace Memory) is used to save the data temporarily before sending that to PC through UART.

The state machine for logic signal analyzer is shown in Figure 3-2.

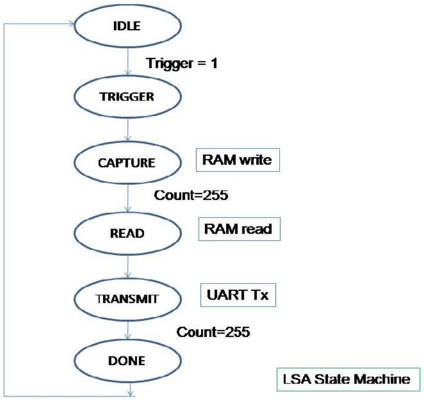


Figure 3-2. LSA State machine

The state machine contains 6 states namely IDLE, TRIGGER, CAPTURE, READ, TRANSMIT and DONE. The state machine stays in IDLE state by default. When trigger input is ‘1’, the state machine enters TRIGGER state. The state machine enters CAPTURE state after that. Memory write takes place in CAPTURE state only. 256 memory writes are done in this state which corresponds to 256 samples being captured.

Once sampling is finished, the state machine enters READ state where RAM read takes place. The read back data is transmitted through UART in TRANSMIT state. Once 256 data transmission finish, the state machine enters DONE state which indicates end of signal capture. The state machine goes back to IDLE state. The block diagram of match unit is shown in Figure 3-3.

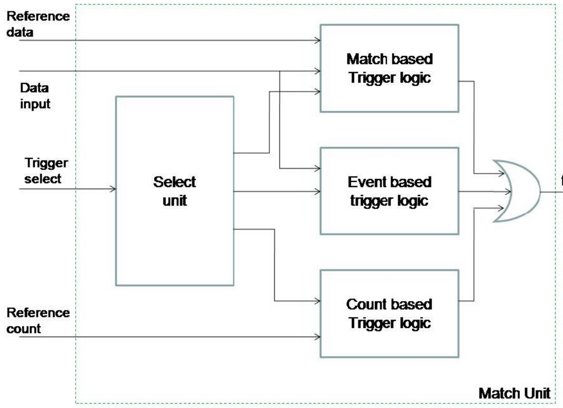


Figure 3-3. Match unit

As shown above, the match unit contains 4 sub blocks namely select unit, match based trigger logic, event based trigger logic and count based trigger logic.

Select unit: The function of select unit is to generate enable signals for other 3 blocks based on the trigger select. 3-bit trigger select is used since there are 8 conditions in total to be enabled.

Match based trigger logic:

This block contains logic to implement the trigger logic based on match condition. It contains 2 data inputs for comparison namely reference data and data input. Reference data is the data with which matching needs to be done and data input is the current sampled data. Various match condition supported are

Din = ref data, Din /= ref data, Din < ref data, Din > ref data, Din <= ref data, Din >= ref data

There are 6 enable signals to this unit to select one of the above match conditions.

Event based trigger logic:

The functionality of this block is to detect an event on the data input. It compares two successive samples to detect the event. To avoid unnecessary glitches pick up, both the current data and previous data latched (or synchronized) versions are used.

Count based trigger logic:

This block contains a counter internally which will be counting the number of samples. The purpose of this block is to enable trigger after a particular number of samples are finished. In this sense, counter based triggering is delayed triggering based on the input sample count. The reference count input decides the number of samples to be discarded (or delayed).

The outputs of these 3 trigger logic modules are logically ORed to generate final trigger signal which would go to the state machine in LSA top.

The block diagram of host interface unit is showninFigure3-4.

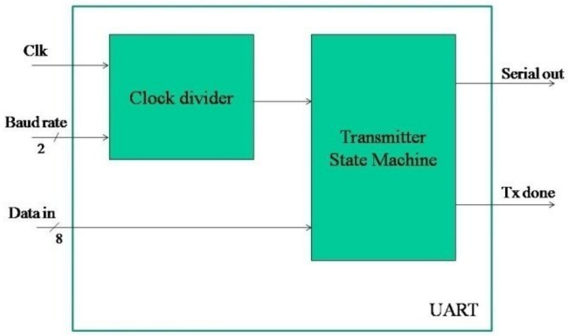


Fig 3-4. Block diagram of host interface unit

The UART contains 2 modules namely clock divider and transmitter state machine. Clock divider generates different clocks corresponding to different baud rates selected. UART transmitter state machine is shown in Figure 3-5.

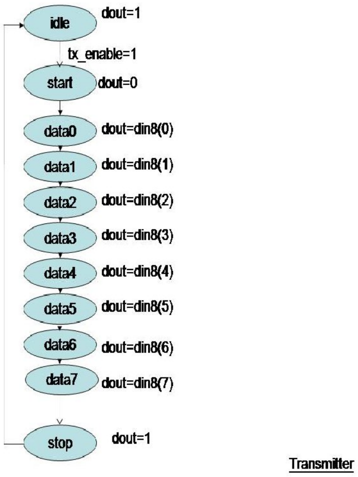


Fig 3-5. UART transmit state machine

The transmitter stays in IDLE state unless transmit enable (tx_enable) is made as '1'. The data transmission starts with tx_enable = 1. As mandated by the protocol, a '0' is transmitted to indicate start of transmission or start bit. This is done in START state. Then data bits 0 to 7 are transmitted in states DATA0 to DATA7. If parity is enabled in configuration register, the data is attached with a parity in PARITY state. Then transmitter enters STOP state and sends a '1'. This indicates the completion of transmission. Then the transmitter enters the IDLE state and waits for next data transmission. An asynchronous FIFO is used as Trace memory in LSA communication module. A FIFO is required since the output data rate of UART is much smaller than the sampling rates supported by capture module.

RTL schematic:

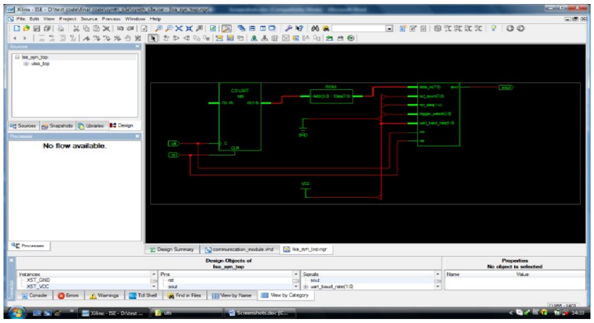


Fig.3.6 RTL schematic for our design

4. Simulation Results

The simulation results for capture module are shown in Figure 4-1.

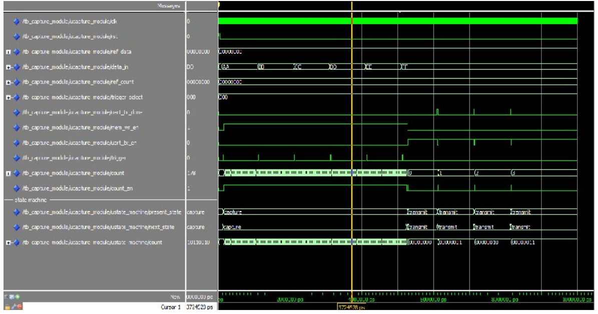
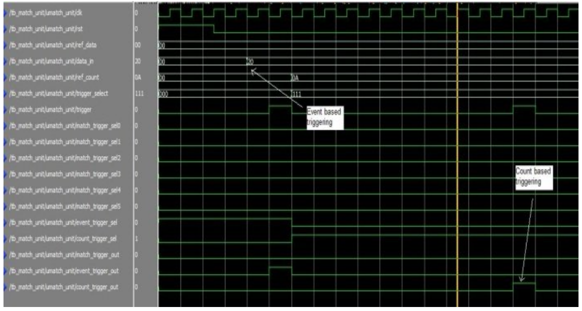


Figure 4-1. Waveforms for capture module

In the waveform data_in refers to the data being captured by LSA. Trigger_select value selects the type of triggering to be used. Ref_data and ref_count are used for different triggering mechanisms of match unit. The signals after divider are state machine signals.

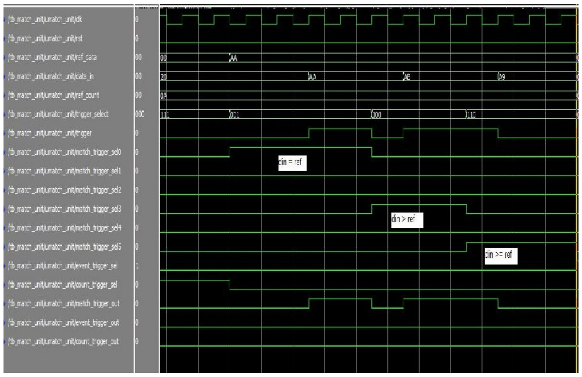
Figure 4-2 shows the simulation waveforms for even based trigger logic and count based trigger logic at the top level.



In this screenshot, initially even based triggering is enabled (trigger select = 000). When the input data (data_in) changes from X"00" to X"20", an event is triggered. We can see the corresponding output on event_trigger_out first and then on trigger output.

Later, count based triggering is enabled with trigger_select = 111. Here a counter is started in count based trigger logic and when it reaches the reference count (ref_count) value of X"0A" a trigger is generated.

Figure 4-3 shows match based trigger logic waveforms from top level.



Here 3 match conditions are exercised, din=ref, din>ref and din>=ref. As we can see first trigger comes when the data_in value reaches the ref_data of X”AA” since the match_trigger_sel0 checks for equality condition.

Then match_trigger_sel3 is enabled which means that din>ref condition is checked. When the data_in value changes to X”AB” which is greater than ref_data of X”AA”, trigger is generated. Third condition set is din>=ref. Since it is already satisfied the trigger continues to be generated. When the data_in changes to X”A9”, trigger becomes 0.

Figure 4-4 shows the simulaton results for communication module.

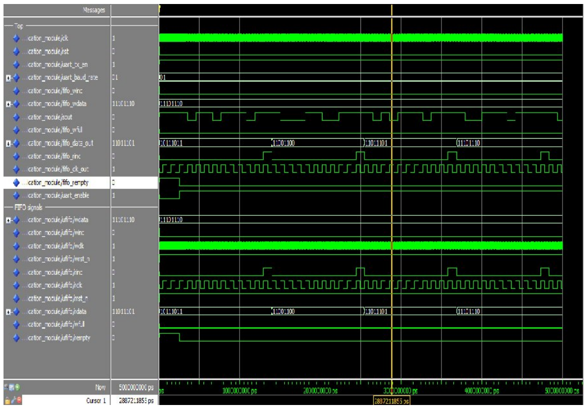


Figure 4-4. Communication module waveforms.Figure 4-4 shows 2 set of signals one for top level communication module waveforms and the other is FIFO signals. For the communication module, data comes in as *fifo_wdata* from capture module. This data is saved in FIFO. This data is read back from FIFO as *fifo_data_out* and is given to UART. Final UART serial output can be seen on *sout*.

Coming to FIFO signals, we can see the *wclk* and *rclk*, clocks used for write and read. As mentioned earlier, the read clock is much slower than write clock. The *wdata* shows the write data and *rdata* is the readback data.

UART (host interface unit) simulation waveforms are shown in Figure 4-5.

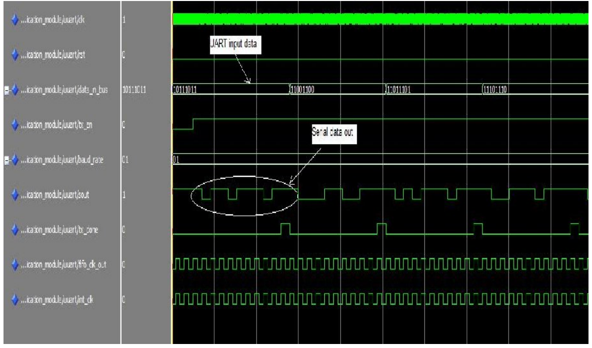


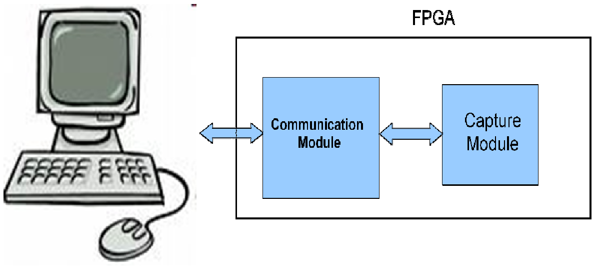
Figure 4-5. UART simulations

Figure 4-5 shows simulation for UART, where *data_in_bus* refers to the parallel data input to be transmitted.

The baud rate is controlled by *baud_rate* input and *clk* is the system clock which is 50MHs currently. The serial output from UART state machine is given on *sout* and corresponding clock can be seen on *fifo_clk_out* which goes to FIFO.

The UART transmission happens only when tx_en signal is ‘1’.these are the simulation results for our design.

Test Setup:



The abow figure can represents the test setup for our design.here our design (VHDL code) is dumped into Spartan 3e FPGA board after that by using requirements we connect our setup to PC (power supply for Spartan 3e, USB connector, RS-232 cable etc..). The requirement figures shown below

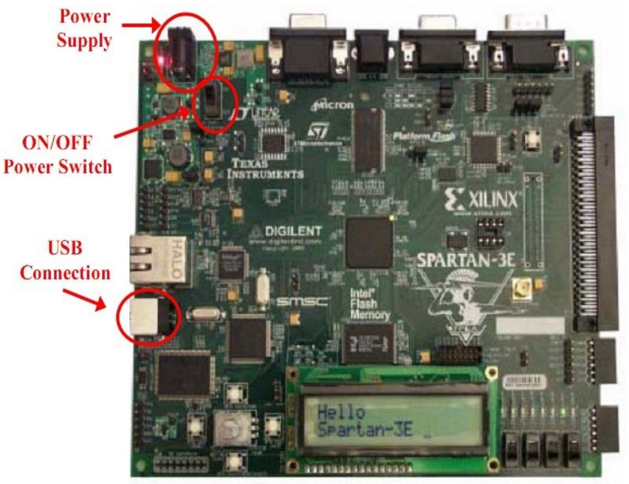


Fig 4.7 spartan 3e FPGA board

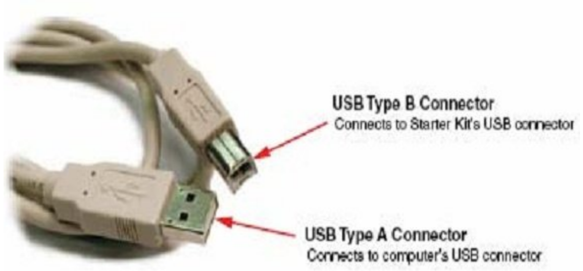
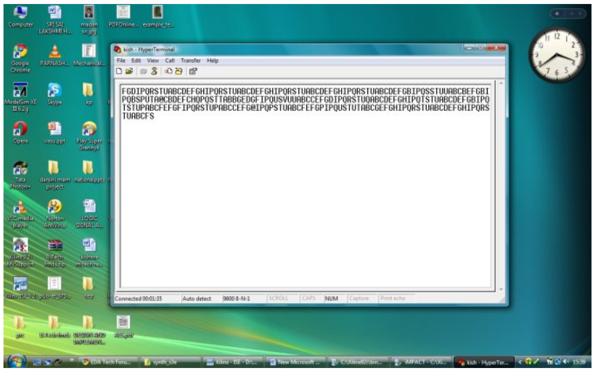

USB Type B Connector
Connects to Starter Kit's USB connector
USB Type A Connector
Connects to computer's USB connector

Fig 4.8 USB connector

HyperTerminal Results:

After test setup we are giving the input data as a hexadecimal to our design the results will be shown by using hyperterminal interms of characters(refer hex to ascii table).the results shown below.



```
FGDIPQRSTUABCDEFGHIPQRSTUABCDEFGHIPQRSTUABCDEFGHIPQRSTUABCDEFGBIPQSSTUUABCBEFGBI
PQBSPUTA@CBDEFCHQPQSTTABBGEDGFIPQUSVUUABCCEFGDIPQRSTUQABCDEFGHIPQTSTUABCDEFGBIPQ
TSTUPABCFEFGFIPQRSTUPABCCEFG@IPQPSTUABCFEFGPIPQUSTUTABCGEFGHIPQRSTUABCDEFGHIPQRS
TUABCFS
```

Fig.4.9 hyperTerminal results

5. Conclusions

We have presented the our low cost logic signal analyzer, a excellent tool for debugging the digital circuits.parallel, independent FPGA-systems. It offers comprehensive trigger functions and can process elaborate trigger sequences. Its most important feature, however, is the autonomous processing of VHDL designs without work required from the user. This has been achieved by means of an instantiation tree structure, which facilitates signal propagation through a VHDL design hierarchy, and a fully parameterized design of the logic analyzer hardware. by using our implementation we can analyse the digital circuits. Our Logic signal analyzer captures and displays many signals at once and analyze their timing relationships.logic analyzer trace the execution of the embedded software and analyze the efficiency of the program execution.

6. References

[1] Integrating logic analyzer functionality into VHDL designs,2008 conference.

[2] Altera Corporation, "*SignalTap II Embedded Logic Analyzer Documentation*".

[3] First Silicon Solutions, "*FPGAView Software*", www.fs2.com/fpgaview.html

[4] P. S. Graham, "*Logical Hardware Debuggers for FPGA-Based Systems*", PhD Dissertation, Bringham Young University.

[5] www.agilent technologies.com

[6] www.wikkipedia/freeencyclopedia.com

[7] www.xilinx .com for spartan-3e features

[8]www.tectronix.com

Multi Technology - Field Programmable Gate Array (MT-FPGA) for Photonic Applications

K.Rambabu,Student,rambabu.karri@gmail.com,+91-9440486801
Under the guidance of Associate Professor M.Madan Gopal
Aurora's Technological & Research Institute, Hyderabad

Abstract

Field Programmable Gate Arrays (FPGA) have revolutionized programmable/reconfigurable digital logic technology.However, one limitation of current FPGAs is that the user is limited to strictly electronic designs. Thus, they are not suitable for applications that are not purely electronic, such as optical communications, photonic information processing systems and other multi-technology applications. While a wide variety of multi-technology devices ranging from micro electromagnetic systems (MEMS) devices, analog devices, photonic information processing devices, telecommunication and digital data processing systems to biological and chemical sensors have been implemented. While these designs are well optimized for a specific application, they do not provide the flexibility associated with reconfigurable/ programmable hardware. As with any ASIC, this approach is very costly, and the turnaround time between design iterations may be several months. The difficulty of using custom designed multi-technology VLSI components is overcome with the introduction of a MT-FPGA with innovative system architecture. The proposed new class of field programmable device will extend the flexibility, rapid prototyping and reusability benefits associated with conventional FPGA technology into photonic and other multi-technology domain.

Index Terms: Application Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs), multi-technology devices, reconfigurable architecture.

1. Introduction

Field Programmable Gate Array (FPGA) is a semiconductor device. It contains programmable logic components, programmable interconnects and programmable I/O blocks.FPGA technology has been developed into a major device technology for implementation of the programmable/reconfigurable digital system with low cost and rapid turnaround time. With multimillion on-chip gates and system clock frequency of several hundred megahertz (MHz), it is possible to implement an entire system of complex digital logic in a state-of-the-art FPGAs.

Unfortunately, the uses of conventional FPGA's have been restricted to digital systems design. Thus, it is not suitable for applications that not only require digital circuits, but other mixed-technology circuit design as well, viz. optical communications, photonic information processing systems and other multi-technology applications. Though a wide variety of multi-technology devices ranging from photonic information processing devices, MEMS devices, telecommunication and digital data processing systems to chemical and biological sensors have been implemented, research in this area has for the most part been limited to systems built with application-specific devices. While ASIC based circuit designs are well optimized for a specific application, they do not provide the flexibility associated with generically reconfigurable/ programmable hardware. This approach is not suitable for either multi-technology test-bed systems or any product development cycle.

The new MT-FPGA extends the FPGA capabilities into the multi-technology realm by integrating different multi-technology blocks in the new FPGA architecture. A multi-technology system designer can test a design idea or concept and verify it in the MT-FPGA hardware without going through the long fabrication process of custom ASIC design. Such a designer can implement incremental changes and iterate in an MT-FPGA design environment within hours instead of months.

2. Overview of the MT-FPGA Architecture

A new and innovative architecture for an FPGA device incorporating mixed-signal and multi technology components has been proposed. The MT-FPGA is a mixed-signal field programmable device that incorporates traditional FPGA technology with different sensor based nodes. The nodes (Multi Technology Blocks) can be used to take input from light, temperature or Chemical sources and convert that to voltages and eventually to information bits. The rest of the MT-FPGA components are strictly digital in nature, programmable in use and closely similar to traditional FPGAs in their setup procedures.

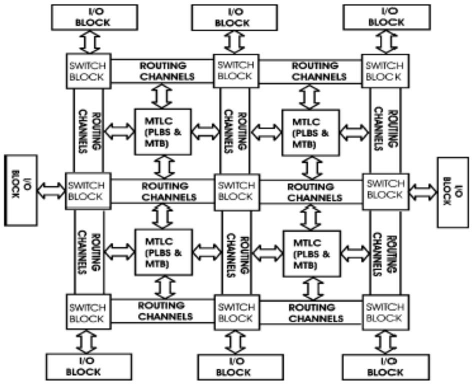


Fig (i) Architectural view of MT-FPGA

MT-FPGA chip contains several MTLCs connected with a complete routing architecture (i.e., connection blocks, switch blocks, input/output (I/O) blocks, segmented and hierarchical routing channels, etc.) It shows the benefits and effects of combined digital/analog components that spanned across several MTLCs and justified the whole MT-FPGA architecture on very strong footings. Where each multi-technology block is embedded within a group of electronic programmable logic blocks (PLBs). Fig. (i) Illustrates the high-level architecture of the MT-FPGA.

To be compatible with the current FPGA design methodology, a symmetrical array of blocks with vertical and horizontal programmable routing channels was adopted in the architecture. Each block in the array consists of a cluster of sub-blocks, viz. a MTB and four programmable logic blocks. This cluster is referred to as a “multi-technology logic cluster” (MTLC).

2.1. Multi Technology Logic Cluster (MTLC)

The MTLC is composed of four programmable logic blocks (also called logic elements), four connection blocks and a multi-technology block. The block diagram of the MTLC depicting the digital components is shown in below Figure (ii).

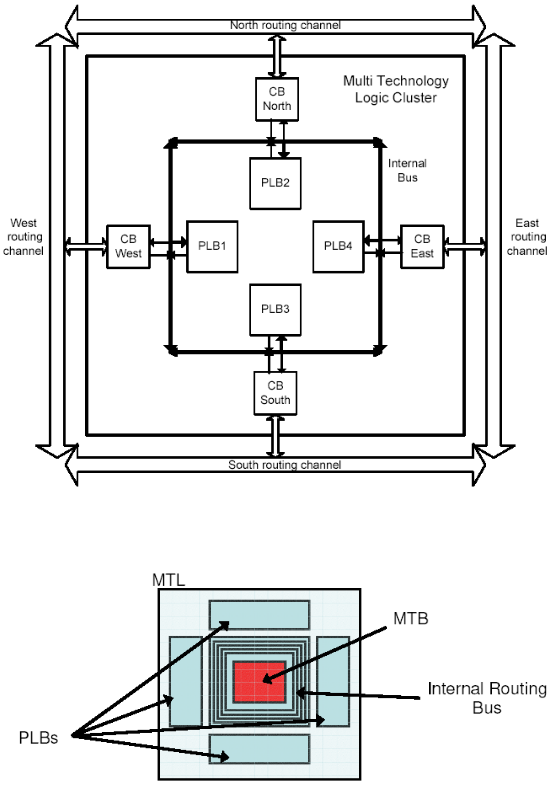


Fig.(ii) Block Diagram of MTLC

The MTLC has an internal bus that is used to connect the different components of the MTLC with each other and with the horizontal and vertical routing channels. The internal bus is 16-bit wide with the following distribution.

- 4-bits are used for output signal distribution of the PLBs.
- 4-bits are used for signals coming from other MTLCs through the routing channels.
- 4-bits are used for communication with the MTB.

- 4-bits are used for clock distribution, reset and configuration signals.

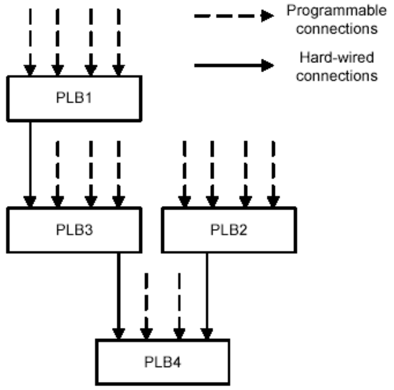
Fig. (ii) Shows the logic diagram of one of the programmable logic blocks used in the MTLC. The MTB may contain one to several multi-technology devices (e.g., photonic, SRAM-based or DRAM-based storage devices, MEMS, chemical/biological sensors, etc.).

2.2 Programmable Logic Block (PLB)

The PLB is composed of a 4-LUT and a number of 8-input multiplexers. The exact number of multiplexers depends on the tree structure that is used in connecting the different PLBs together. The L3-4.2 tree structure used in the MT-FPGA is shown in Figure (iii).

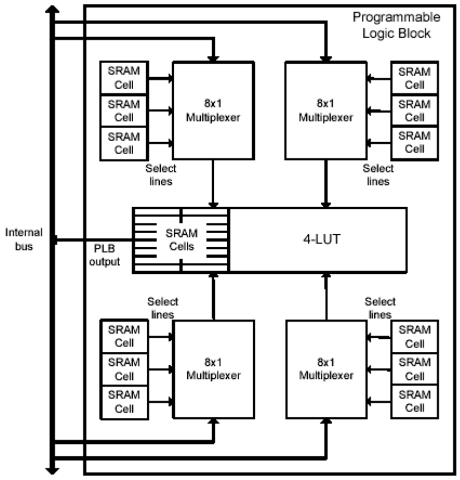
In general, the input of the 4- LUT is connected to an 8:1 MUX (if it is not hardwired). In the MTLC topology, PLB1 and PLB2 are located at the top of the connection tree. Therefore, PLB1 and PLB2 do not have any hardwired-connected inputs and each of their inputs can be programmed to choose from a variety of signals, coming from the same or other MTLC's. In contrast to PLB1 and PLB2, the inputs to PLB3 and PLB4 do not have all programmable connections. The output from PLB1 is hardwired to one of the inputs to PLB3 through the internal bus.

Similarly, the outputs from PLB2 and PLB3 are hardwired to two of PLB4's inputs. This combination of hardwired and programmable links is chosen among the PLBs to conform to the L3-4.2 topology. The goal is to reduce the size/complexity of the interconnection matrix of the present architecture to a point that is sufficient for nearly complete logic utilization. This reduction in interconnection matrix is directly translated into a reduction of area and an increase in speed of the MT-FPGA architecture.



Fig(iii) L3-4.2 Tree Structure in the MTLC

The PLB also has a flip-flop and a de-multiplexer that can be used to select between the normal output of the PLB and the output of the flip-flop. The block diagram of a PLB with four multiplexers is shown in Figure (iv).



Fig(iv) Block Diagram of the PLB

Since the flip-flop and de-multiplexer are not considered during the bit-stream and model generation process. Each of the sixteen entries of the LUT is controlled by an SRAM cell. The output of the LUT for a particular input is determined by the value stored in the SRAM cell corresponding to the input pattern. Thus, The LUT can implement any 4-input function by configuring the SRAM cells appropriately. Similarly, the select lines of the multiplexers are controlled by SRAM cells. The inputs to the LUT are determined by the value stored in the SRAM cells controlling the select lines of the multiplexers. The multiplexer inputs are obtained from the internal bus.

2.3 Connection Block (CB)

The connection block determines the connections between the components of the MTLC and the routing channels through the internal bus of the MTLC. The block diagram of the CB with the digital components is shown in Figure (v).

An MTLC has four connections blocks, one each in the north, east, west and south edges of the MTLC. The connection block has one digital 4-input multiplexer, one digital 4-output de-multiplexer, one analog 2-input multiplexer and one analog 2-output de-multiplexer (not shown in Figure (v)) to facilitate communication between the internal bus of the MTLC and the routing channel. Again, the select lines of the multiplexers and de-multiplexers are controlled by SRAM cells.

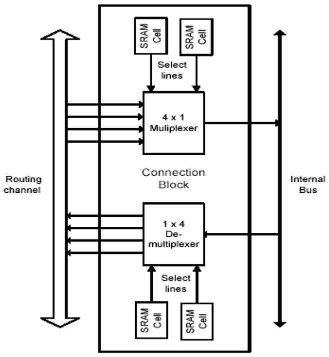


Fig.(v) Block Diagram of Connection Block

2.4 Switch Block (SB)

The switch block is used to provide connections between the various MTLCs and IOBs through the horizontal and vertical routing channels. The block diagram of the digital switch block is shown in Figure (vi).

Each routing channel has a total of six digital routing tracks and two analog routing tracks. The digital switch block is of the disjoint type and provides programmable connections between each track in a particular direction (north, east, west or south) and the corresponding tracks in all the other directions. The programmable connections are implemented using transmission gates whose gate nodes are controlled by SRAM cells. With six digital routing tracks in each direction, each SB consists of thirty six programmable switches.

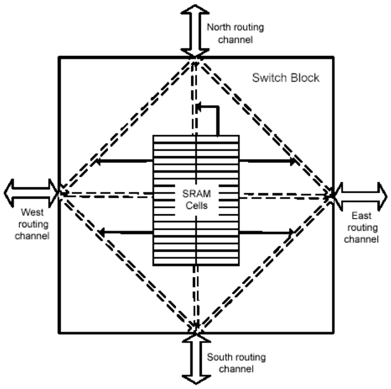


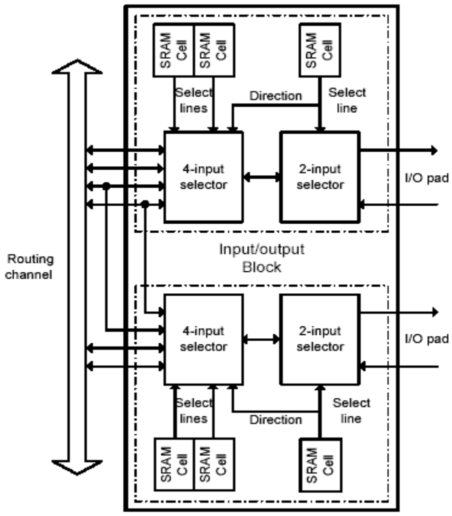
Fig.(vi) Block Diagram of Switch Block (SB)

2.5 Input/output Block

The IOB provides connections between the SBs at the periphery of the FPGA chip and the input and output pads of the chip. A simplified block diagram of the digital part of the IOB used for modeling is shown in Figure (vii).

It can be seen that the IOB is composed of two identical components. Each component provides a connection between an input/output pad and a track in the routing channel. For this purpose, the IOB uses bi-directional 4-input and 2-input path selectors. The path selectors are similar to a multiplexer except that the inputs and/or outputs of the selector are bi-directional. Thus, the IOB can be used to connect two input/output pins to two tracks in the routing channel. The select lines of the selectors and the direction of signal travel through them are controlled by SRAM cells. For modeling purposes, the SRAM cell that controls the 2-input selector is also used to control the direction of signal travel through the 4-input selector, as shown in Figure (vii).

The programmable interconnect structure is responsible for providing routing for all signals that are exchanged among different components within MT-FPGA architecture. Signals from the outside world enter and exit the interconnect framework through I/O blocks.



Fig(vii) Block Diagram of Input/Output Block

Unlike PLBs, MTBs can handle both analog and digital signals as inputs and outputs. It communicates with others using a 4-bit mixed-signal internal bus. Each MTLC block is capable of being configured to suit a target application. Taken individually, each MTLC does not have enough resources to construct an entire system. However, many MTLCs can be combined with a programmable routing architecture into larger practical systems.

2.6 Multi Technology Block (MTB)

The rapid advancing state of photonics technology, optoelectronic systems integrated in a CMOS technology are not only becoming increasingly attractive in low cost communication systems (i.e., local-area networks, fiber-to-the-home, etc.) and optical storage systems (i.e., CD-ROM, DVD and Blue-ray disc) but showing escalating enthusiasm in the mainstream digital computing and switching technology (i.e., optical interconnects at box-to-box, board-to-board or even chip-to-chip levels). In last several years many investigations on integrated optoelectronic components and circuits have highlighted and motivated the merits of optoelectronic approaches in optical data communication and photonic information processing systems.

To exemplify the interconnection and communication among MTBs and PLB clusters located within the same and neighboring MTLCs, we have chosen an optoelectronic block that is embedded in the MTB space of the MT-FPGA. This optoelectronic block is a pixel-scale optical power meter that can be used individually or in an array of sensors within the MT-FPGA's programmable architecture. The optical power meter integrated in the MT-FPGA can be used to sense and quantify the light incident on the detector structure. It is possible to use the MT-FPGA containing "pixel scale optical power meter" MTBs as a general purpose optical gateway where optically encoded signals can be received and converted into digital data and transformed in intelligent ways to compensate for signal losses in the optical paths, help extract useful information or perform useful signal transformations.

2.6.1. Description of Components of the MTB

On the top level, the pixel-scale optical power meter consists of a photo detector, photo receiver, analog-to-digital converter (ADC) module and storage elements. Additionally, buffers and a sample-and-hold circuit are added to maintain signal driving strength and integrity at the nodes between the major components.

Fig. (viii) is a schematic showing how optical light enters into the system and passes through the series of components (mentioned before) until it is converted into a digital signal. The Gain Control and V-bias signals help improve the flexibility of the optical power meter. It is well known that the semiconductor parameters of silicon chips can vary by significant amounts at different points on a wafer even in a single fabrication run. Because of this, each optical power meter may be slightly different from the other and will require a different bias point and have differing gain characteristics. By making the bias and gain control tunable, it is possible to make adjustments for process variations. All that is required is for the chips to be characterized after fabrication, and based on the results, the control signals of any two MTBs are adjusted until the bias point and range are brought in harmony.

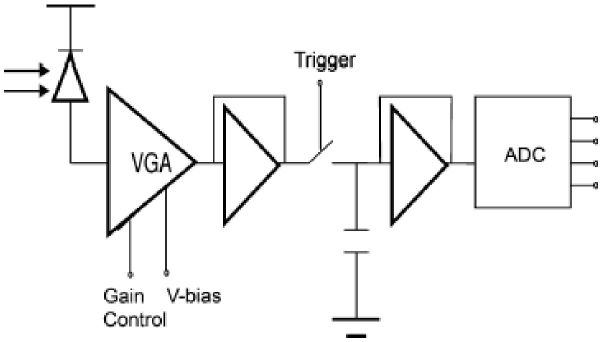


Fig (viii) Top level Diagram of the MTB (Pixel Scale Optical Power Meter)

Furthermore, the MT-FPGA logic could make intelligent decisions to alter the power meter performance based on other factors (i.e., temperature change, differences in light sources, feedback from an auto-calibration routine, signal loss in optical path, etc.).

The purpose of the photo receiver circuit is to provide a flexible amplifier capable of amplifying both small and large amounts of current from the photo detector. This means that a variable gain amplifier (VGA) design with a large range of gain levels is a desirable choice for this design.

The ADC selected for the MTB block design is a flash ADC that has the fastest data conversion capability. The ADC circuit features two different comparator circuit designs, one designed around an NMOS differential amplifier, and another around a PMOS differential amplifier. The PMOS based differential amplifier works well for the lower voltage nodes, while the NMOS handles the higher voltage nodes. To convert a total of 2 -1 comparators outputs to a 4-bit digital output signal, a 16:4 decoder is used. In order for a functional system to be integrated into the MT-FPGA, the four outputs of the ADC were stored into an output register.

2.6.2 Testing of the MTB (Pixel Scale Optical Power Meter)

An important step to evaluate the MTB is to build a specialized optical test setup for measuring the MTB performance and collecting the important data to support the underlying MT-FPGA design concept. Fig. (ix) Represents the line diagram of the optical test bench.

The test MTB of the MT-FPGA chip receives an infrared laser beam (λ = 852 nm) from a modular laser diode controller through a planar-convex lens to prevent beam divergence and then sent into a beam splitter. The beam splitter divides the light into two paths for the purpose of providing an accessible measurement point so that the light can be measured and delivered to the photo detector simultaneously.

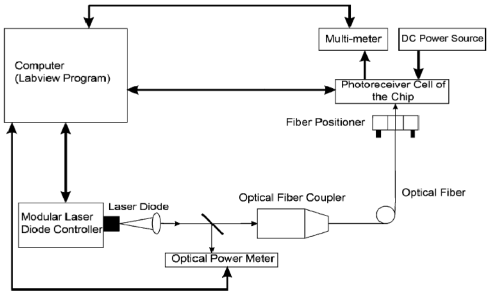


Fig (ix) Experimental setup for testing an optical MTB (Pixel scale power meter)

One of the divided paths from the splitter continues on into a fiber coupler and coupling assembly that allows precise control for positioning the cleaved fiber tip directly over the focal point of the light as it passes through a focal lens. After being coupled into a single mode fiber, the light then travels to a second fiber mounting assembly used to direct the fiber tip onto a precise spot on the exposed photo detector of the fabricated MTB device under test in the MT-FPGA chip. The chip is also mounted on its own specialized assembly that features automatic fine-grained motion control.

In addition to the optical equipment, all electronic equipments (i.e., laser diode' (LD) controller, optical power meter, and multi-meters) and the test chip are interfaced to an electronic host computer. A 'Lab VIEW' program running on the host computer is used for making data acquisition and controlling the electronic equipment. Before doing the actual testing, the test setup needs to be characterized to capture the relationships between the laser driving current and the optical power actually delivered to the photo detector as the chip is being tested.

MT-FPGA not only improves performance but also reduces system area and energy. To illustrate the performance gains possible with the MT-FPGA consider a simple image processing system with an inexpensive digital camera, memory, and an FPGA device for performing basic operations on the image. The image pixels are streamed from memory, N pixels per clock into the FPGA to be processed, where N is determined by the memory port width. This limits the rate of processing to (R x C)/N cycles per image.

Where R and C are the number of rows and columns in the image.

Example:

For a small image, say 128 x128, with a memory port width of 128-bits, N is 32 for 4-bit pixels. Hence, the number of clocks required to find for example the average intensity is (128 x 128)/32 = 512 clocks. However, the MT-FPGA has all pixels available simultaneously and just has to sum them. If pair wise additions are performed each clock then it would require (R x C) clocks to compute the sum, for this example (128 x 128) = 14 clocks. Fig (xiii) highlights this analysis as the resolution of the image increases, the no. of clock cycles required to process a single frame grow rapidly in a traditional FPGA. This scenario could be exacerbated by limited I/O bandwidth of the system. MT-FPGA demonstrates here a considerable performance gain over the traditional FPGA.

In general, the traditional FPGA cannot handle directly any multi-technology inputs (i.e., optically encoded data inputs) as it does not have the facility of programmable and integrated MTBs. Unlike the MT-FPGA, the traditional FPGA is required to use other companion chips (i.e., sensor chip, memory chip, etc.) in conjunction to be able to operate in a multi-technology environment. In the process, however, it creates several bottlenecks in the system performance as we discussed before. To reap the full benefits of conventional FPGA technology in multi-technology domains, users need to rely on MT-FPGA style versatile architecture.

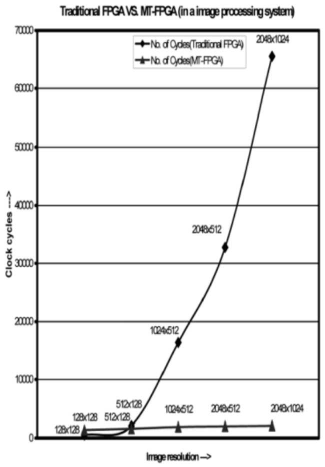


Fig. (x) Performance comparison between a traditional FPGA and an MT-FPGA in an image processing system.

In general, the image and signal processing applications where system size, cost and rapid turnaround of designs are critical factors, the MT-FPGA architecture is far more reasonable choice than a traditional FPGA-based or ASIC based system.

Finally, the MT-FPGA architecture is very much scalable and suitable for high volume productions and offers flexibility and rapid prototyping capabilities in face of increased time-to market concerns. As system requirements often change over time, the cost of making incremental changes to MT-FPGA designs are quite negligible when compared to the large expense of respinning an ASIC.

3. Conclusion

The present work demonstrated a tremendous opportunity to exploit the synergy between traditional FPGA technology and specialized device technologies that are widely used in multi-technology applications. The resulting device is an MT-FPGA that exploits the benefits of both technologies (i.e., generically programmable functionality and benefits of multi-technology domain).In contrast to mainstream FPGA this research has extended the benefits associated with conventional FPGAs to photonics and other multi-technology domain and give rise to the development of a wider class of reconfigurable and embedded integrated systems. To substantiate this novel architectural concept, a test chip has been fabricated through MOSIS foundry service in TSMC 0.35 µm technology.

4. References

[1] Prerna Patel "Design of a Pixel Scale Optical Power Meter Suitable for Incorporation in a Multi-Technology FPGA", MS Thesis, *University of Cincinnati, Cincinnati, Ohio, 2001*

[2] Vishal Sapre " Configuration Bit Stream Generation for the MT-FPGA & Architectural Enhancements for Arithmetic Implementations "MS Thesis, *University of Cincinnati, April 2005*

[3] Prashanth NT Pulipaka ," Modeling and Integration of Connection Block and Multi-Technology Logic Cluster with Enhancement for the Second Generation MT-FPGA" MS Thesis, *University of Cincinnati August 2006*

[4] Raghav Swaminathan " Design of an I/O block and development of a signal flow indicator tool for the second generation MT-FPGA, *University of Cincinnati, August 2006*

[5] "Development of Multi-Technology FPGA Incorporating Photonic Information Processing Block Subsystem", *University of Cincinnati.*

Application of Programming Temporally Integrated Distributed Embedded Systems

Sharmila .B.S , Shwetha.L
7th sem,ECE Kalpataru Institute of Technology, Tiptur, Karnataka
Sharmila241989@gmail.com , Kani.kanasu2@gmail.com
Ph:7204250925

Abstract

The introduction of network time protocols such as NTP (at a coarse granularity) and IEEE 1588 (at a fine granularity) gives a relatively consistent global notion of time that has the potential to significantly change how we design distributed real-time systems. In [4], we present a programming model called PTIDES (Programming Temporally Integrated Distributed Embedded Systems) that uses discrete-event (DE) models as programming specifications for distributed real-time systems and describe an execution model that permits out of order processing of events without sacrificing determinacy and without requiring backtracking. In this paper, we present an interesting networked camera application programmed using PTIDES and show how the execution model in [4] can be used to meet real-time constrains in the system.

INTRODUCTION

The introduction of network time protocols such as NTP (at a coarse granularity) and IEEE 1588 (at a fine granularity) gives a relatively consistent global notion of time that has the potential to significantly change how we design distributed real-time systems. Time synchronization over standard networks, such as provided by NTP [3], can achieve timing precision within ten milliseconds, which is sufficient for many interactive distributed systems, such as computer games, where human-scale time precision is adequate. The recent standardization (IEEE 15881) of high-precision timing synchronization over Ethernet provides much higher timing precision that is essential for many embedded systems. Implementations of IEEE 1588 have demonstrated time synchronization as precise as tens of nanoseconds over networks that cover hundreds of meters, more than adequate for many manufacturing, instrumentation, and vehicular control systems.

In [4], we present a programming model called PTIDES (Programming Temporally Integrated Distributed Embedded Systems) that leverages time synchronization over distributed platforms. PTIDES uses discrete-event models as programming specifications for distributed real-time systems and extends discrete event

models with the capability of mapping certain events to physical time. We use model time to define execution semantics and add constraints that bind certain model time events to physical time. We limit the relationship of model time to physical time to only those circumstances where this relationship is needed. A correct execution will simply obey the ordering constraints implied by model time and meet the constraints on events that are bound to physical time. We then seek execution strategies that can preserve the deterministic behaviors specified in DE models and also provide efficient real-time executions without paying the penalty of totally ordered executions. The key idea is that events only need to be processed in time-stamp order when they are causally related. An execution model that permits out of order processing of events without sacrificing determinacy and without requiring backtracking is described in [4]. The formal foundation is based on the concepts of relevant dependency and relevant order. The results are particularly valuable in time-synchronized distributed systems, since we can take advantage of the globally consistent notion of time as a coordination channel. Based on relevant orders, we can statically analyze whether a given model is deployable on a network of nodes, assuming that we have upper bounds on the network delays.

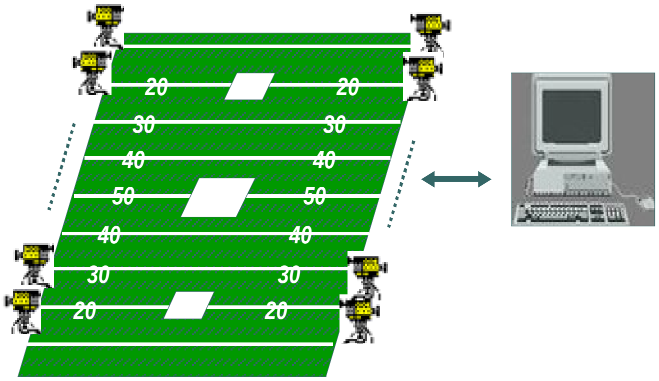


Fig .1. Networked camera application.

In this paper we describe an application of networked cameras and discuss the implementation and execution of the application as a PTIDES model. This paper is organized as follows. Section II introduces the application. Section III shows the implementation of the application. In section IV, we discuss the challenges in executing the system and show how the execution strategy based on the relevant order [4] can be used to execute the system efficiently to meet the real-time constraints.

II. APPLICATION

Consider that we have N cameras distributed over a football field as shown in figure 1. All the cameras have computer-controlled picture and zoom capabilities. Each camera only has a partial view of the field. The images produced by each camera are transferred over the network to the central computer, where the images get processed to produce an entire view of the field or a sequence of views for some interesting moment. On the other hand, a user sitting in front of the central computer may issue commands to the cameras to zoom or change the frequency of taking images. For example, when there is a touchdown, the user may want to control a set of cameras to zoom in or take pictures at a higher frequency.

Suppose that the clocks at all the cameras and the central computer are precisely synchronized, and each camera can generate precisely timed physical events under the control of software, i.e. we can control a camera to take a picture or zoom precisely at some physical time. Zooming takes time κ to set up, and during this period we do not want the camera to take fuzzy pictures. Given that the commands controlling the cameras to zoom or change frequency are transmitted over the network with some delay bounded, the challenges here are how to make sure that all the cameras adjust at the same time and how to coordinate the zoom action and the taking picture action properly on each camera. We discuss our design for this application in the next section.

III. I MPLEMENTATION

Figure.2 shows the implementation of the networked camera application as a PTIDES program. A PIDES program is given as a composition of actors, by which we mean a set of actors and connectors linking their ports. The dashed boxes divide the model into two parts, the top one to be executed on each camera and the bottom one to be executed on the central computer. The parts communicate via signal s1 and s2. We assume that events in these signals are sent over the network as time-stamped values. The Command actor is a software component that wraps interactions with the user input device. When a user input comes in, the Command actor checks with its synchronized clock for the current time, uses the returned time value to time stamp the input message and sends the time stamped message, called an event, to all the cameras. The right part of the model on the central computer processes the images taken at each camera and displays the result.

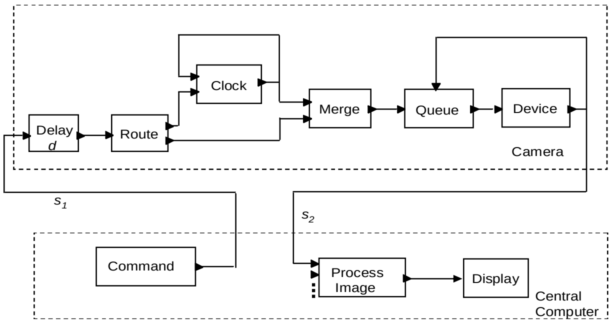


Fig 2. Specification of the networked camera application

The Device actor in the top part of the model shown in figure 2 wraps interactions with the camera driver. At its input port, it receives a potentially infinite sequence of time-stamped values, or events, in chronological order. This actor takes the time stamp of an input event as specification of when an action, determined by the value of the event, happens at physical time. The first output port produces a time stamped value for each input event, where the time stamp is strictly greater than that of the input event.

The second output port produces the time-stamped image and sends it to the central computer.

The Queue actor buffers its input event until an event is received at the trigger port, which is the one at the top of the actor. We assume there is an initial event on the trigger port at the beginning. The feedback loop around the Queue and Device actor ensures that the Device does not get overwhelmed with future requests. It may not be able to buffer those requests, or it may have a finite buffer. The Clock actor produces time-stamped outputs where the time stamp is some integer multiple of a period p. The time stamps are used to control when the camera takes pictures. The period can be different for each clock and can be changed during run time upon receiving an input on the second input port. If there is an event with value v and time stamp t at the second input, the clock actor will change its period from p to p' = p * v and produce an output with time stamp t0+ np' where t0 is the time stamp of the last output and n is the smallest integer so that t0 +np'> t. The feedback loop around the clock actor is used to trigger the next output, and we assume there is an initial event on the first input at the beginning. The Delay actor with a delay parameter d will produce an event with time stamp t + d at its output given an event with time stamp t at its input. Two kinds of user commands are received on each camera, the change frequency command and the adjust zoom command. The Router actor separates these events and sends the change frequency events to the Clock actor and the adjust zoom events to the Merge actor. The Merge actor merges the events on the two input ports in chronological order. It gives priority to the second input port if input events have identical time stamps. That is, we give higher priority to the user to control a camera. In the above discussion, the time stamps are values of model time. Some actors in the model bind model time to physical time. The Command actor binds model time to physical time by producing an event with model time corresponding to the physical time when the user input happens. The Device actor binds model time to physical time by producing some physical action at the real-time corresponding to the model time of each input event. The Device actor also impose real-time constraints to the model. The input events must be made available for the Device actor to process them at a physical time strictly earlier than the time stamp. Otherwise, the component would not be able to produce the physical action at the designated time. We limit the relationship of model time to physical time to only those circumstances where this relationship is needed. For other actors in the model, there is no real-time constraints and model time is used to define execution semantics.

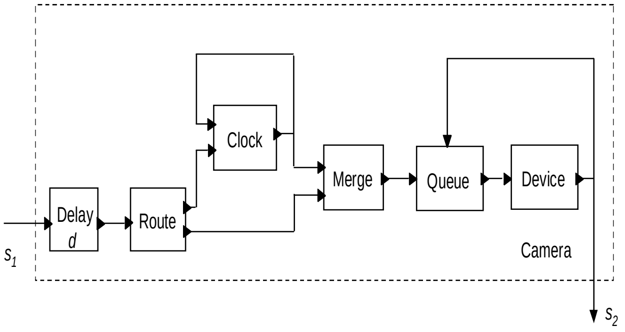


Fig.3 The program on the camera

IV. EXECUTION

How to build a run-time environment to execute the distributed model shown in figure 2 to deliver the correct behavior and meet the real-time constrains in the system is a challenging problem. A brute-force implementation of a conservative distributed DE execution of this model would stall execution in a camera at some time stamp t until an event with time stamp t or larger has been seen on signal s1. Were we to use the Chandy and Misra approach [1], we would insert null events into s1 to minimize the real-time delay of these stalls. However, we have real-time constraints at the Device actors that will not be met if we use this brute-force technique. The so-called "optimistic" techniques for distributed DE execution will also not work in our context. Optimistic approaches perform speculative execution and backtrack if and when the speculation was incorrect [2]. Since we have physical interactions in the system, backtracking is not possible.

An execution model that permits out of order processing of events without sacrificing determinacy and without requiring backtracking is described in [4]. Here, we only give the result after applying dependency analysis to show which events in the system can be processed in a different order than their time stamp order and why this can help to meet real-time constants. For details about dependency analysis, please refer to [4].

Figure 3 shows the program running on each camera and names the ports. The dependencies between the input and output ports of each actor are listed below:

$(p1, p2) = d,$
$\delta(p3, p4) = 0,$
$\delta (p3, p5) = 0$
$\delta(p6, p8) = Pmin,$
$\delta(p7, p8) = 0,$
$\delta(p9, p11) = 0,$
$\delta(p10, p11) = 0$ (1)
$\delta(p12, p14) = 0,$
$\delta(p13, p14) = 0,$
$\delta(p15, p16) =\alpha,$
$\delta(p15, p17) = 0$

where Pmin is the minimum time interval between two consecutive taking picture action on the camera and _ > 0.

Based on the dependencies specified for each actor, we can calculate the relevant dependencies between any pair of input ports. As an example, the relevant dependency d(p1, p9) is d, which means any event with time stamp t at port p9 can be processed when all events at port p1 are known up to time stamp $t - d$. Assume the network delay is bounded by D, at physical time $t - d + D$ or later. Note that although the Delay actor has no real-time properties at all (it simply manipulates model time), its presence loosens the constraints on the execution.

By choosing d properly, i.e. $d > D$, we can deliver e to p15 before physical time reaches t and thus satisfy the real-time constraint on p15.

What we gain from the dependency analysis is that we can specify which events can be processed out of order, and which events have to be processed in order. Please refer to [4] to see how this information

can be used to define a partial order, called the relevant order, on events and how to design execution strategies based on the relevant order.

V. CONCLUSION

We describe the use of DE models as programming specifications for time-synchronized distributed systems. A networked camera application and its implementation as a DE model are studied. The challenges in executing the specification over a distributed platform are discussed, and we then show how an execution model that permits out of order processing of events without sacrificing determinacy and without requiring backtracking can be used to improve the excitability (i.e., to meet real-time constrains) of the application.

VI. ACKNOWLEDGMENTS

This paper describes work that is part of the Ptolemy project, which is supported by the National Science Foundation (NSF award number CCR-00225610), and Chess (the Center for Hybrid and Embedded Software Systems), which receives support from NSF, the State of California Micro Program, and the following companies: Agilent, DGIST, General Motors, Hewlett Packard, Microsoft, National Instruments, and Toyota.

REFERENCES

[1] K. M. Chandy and J. Misra. Distributed simulation: A case study in design and verification of distributed programs. IEEE Trans. On Software Engineering, 5(5), 1979. [
2] D. Jefferson. Virtual time. ACM Trans. Programming Languages and Systems, 7(3):404–425, 1985.
[3] D. Mills. A brief history of ntp time: confessions of an internet timekeeper. ACM Computer Communications Review 33, April 2003.
[4] Y. Zhao, E. A. Lee, and J. Liu. Programming temporally integrated distributed embedded systems. Technical Report UCB/EECS-2006-82,
EECS Department, University of California, Berkeley, May 28 2006. 5. Bouyer, P., Haddad, S., Reynier, P.-A.: Timed unfoldings for networks of timed automata. In: Graf, S., Zhang, W. (eds.) ATVA 2006. LNCS, vol. 4218, pp. 292–306. Springer, Heidelberg (2006)
6. Carlson, J., H°akansson, J., Pettersson, P.: SaveCCM: An analysable component model for real-time systems. In: Proc. of the 2nd Workshop on Formal Aspects of Components Software (FACS 2005). Electronic Notes in Theoretical Computer Science. Elsevier, Amsterdam (2005)
7. Cassez, F., Chatain, T., Jard, C.: Symbolic unfoldings for networks of timed automata. In: Graf, S., Zhang, W. (eds.) ATVA 2006. LNCS, vol. 4218, pp. 307–321. Springer, Heidelberg (2006)
8. David, A., Behrmann, G., Larsen, K.G., Yi,W.: A tool architecture for the next generation of UPPAAL. In: Aichernig, B.K., Maibaum, T.S.E. (eds.) Formal Methods at the Crossroads. From Panacea to Foundational Support. LNCS, vol. 2757, pp. 352–366. Springer, Heidelberg (2003)
9. G¨ossler, G., Sifakis, J.: Composition for component-based modelling. Science of Computer Programming 55(1-3), 161–183 (2005)
10. H°akansson, J., Pettersson, P.: Partial order reduction for verification of real-time components. In: Proc. of 1st InternationalWorkshop on Formal Modeling and Analysis of Timed Systems. LNCS. Springer, Heidelberg (2007)
11. Lugiez, D., Niebert, P., Zennou, S.: A partial order semantics approach to the clock explosion problem of timed automata. Theoretical Computer Science 345(1), 27–59 (2005)

Implementation of FPGA based LOW-COST Logic Signal Analyzer (LSA)

S NAGAKISHORE BHAVANAM, M. MADANA GOPAL

e-mail : satyabhavanam@gmail.com , mekala.madan@gmail.com

Auroras Technological & Research Institute

Uppal, Hyderabad-39, A.P, INDIA.

Abstract

Logic signal analyzers are very essential instrument for digital circuit or board debugging. The existing market solutions offer several features, but the cost of such instruments is very high and most of the time we don't need that much capable instruments. In this paper a low cost logic signal analyzer is implemented around the Spartan-3E FPGA. The FPGA being capable of offering high frequency data paths in them become suitable for realizing high frequency signal capturing logic.

The paper work includes development of FPGA based logic signal analyzer using VHDL. The logic signal analyzer will be capable of implementing match conditions, counter based triggering, external clocking and internal clocking features. The blocks such as registers, counters, comparators, state machines will be used in realizing these blocks. The captured data will be stored in memory before transferring the data to PC. The UART core will be developed which, will be used for transferring the data to PC.

Modelsim Xilnx edition (MXE) tools will be used for simulation. Xilinx FPGA synthesis tools will be used for synthesizing the design for Spartan FPGAs. The developed application will be tested on Spartan 3E development board. By using HyperTerminal we check our design. By using GUI ,we will show the results with respect to waveforms.

1. INTRODUCTION

A logic analyzer is an electronic instrument which displays signals of a digital circuit which are too fast to be observed and presents it to a user who can then precisely observe with greater ease the operation of the digital system under test. It is typically used for capturing data in systems having too many channels to be examined with an oscilloscope. Software running on the logic analyzer can convert the captured data into timing diagrams, protocol decodes, state machine traces, assembly language, or correlate assembly with source-level software.

Presently there are three distinct categories of logic analyzers available on the market:

i. The first is mainframes, which consist of a chassis containing the display, controls, control computer, and multiple slots into which the actual data capturing hardware is installed.

ii. The second category is standalone units which integrate everything into a single package, with options installed at the factory.

iii. The third category is PC-based logic analyzers. The hardware connects to a computer through a USB or LPT connection and then relays the captured signals to the software on the computer. These instruments are less expensive than either mainframes or standalone units although they lack the sophisticated functionality. These devices are typically much smaller, because they do not need displays or hardware input such as dials.

2. LOGIC ANALYZER OPERATION

A logic analyzer may be triggered on a complicated sequence of digital events, and then capture a large amount of digital data from the system under test (SUT). The best logic analyzers behave like software debuggers by showing the flow of the computer program and decoding protocols to show messages and violations. When logic analyzers first came into use, it was common to attach several hundred "clips" to a digital system. Later, specialized connectors came into use. The evolution of logic analyzer probe has led to a common footprint that multiple vendors support, which provides added freedom to end users. Introduced in April, 2002, connectorless technology (identified by several vendor specific trade names: Compression Probing; Soft Touch; D-Max) has become popular. These probes provide a durable, reliable mechanical and electrical connection between the probe and the circuit board with less than 0.5pF to 0.7 pF loading per signal. Once the probes are connected, the user programs the analyzer with the names of each signal, and can group several signals into groups for easier manipulation. Next, a capture mode is chosen, either timing mode, where the input signals are sampled at regular intervals based on an internal or external clock source, or state mode, where one or more of the signals are defined as "clocks," and data is taken on the rising or falling edges of these clocks, optionally using other signals to qualify these clocks.

After the mode is chosen, a trigger condition must be set. A trigger condition can range from simple (such as triggering on a rising or falling edge of a single signal), to the very complex (such as configuring the analyzer to decode the higher levels of the TCP/IP stack and triggering on a certain HTTP packet).

At this point, the user sets the analyzer to "run" mode, either triggering once, or repeatedly triggering.

Once the data is captured, it can be displayed several ways, from the simple (showing waveforms or state listings) to the complex (showing decoded Ethernet protocol traffic). The analyzer can also operate in a "compare" mode, where it compares each captured data set to a previously recorded data set, and stopping triggering when this data set is either matched or not. This is useful for long-term empirical testing. Recent analyzers can even be set to email a copy of the test data to the engineer on a successful trigger.

3. **ARCHITECTURAL DESIGN**

The block diagram of logic signal analyser is shown in Figure 3-1. The block diagram contains 2 basic modules namely capture module and communication module. Capture module is responsible for signal capture. It contains a capture state machine and samples counter. The counter is 8-bit counter. The communication module contains UART and FIFO. The FIFO (Trace Memory) is used to save the data temporarily before sending that to PC through UART.

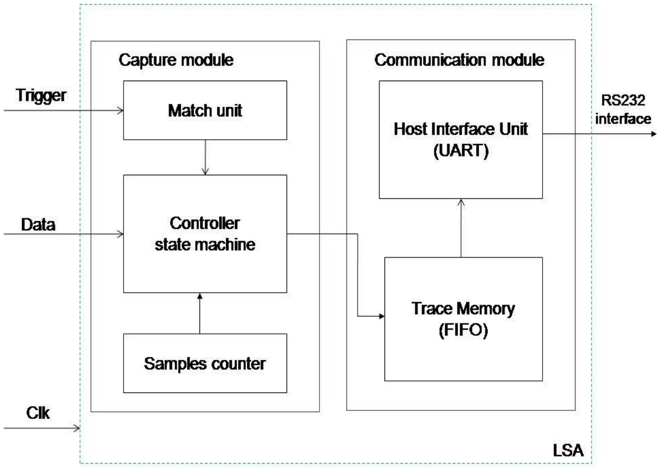


Figure 3-1. Block diagram of Logic Signal Analyser

The state machine for logic signal analyzer is shown in Figure 3-2.

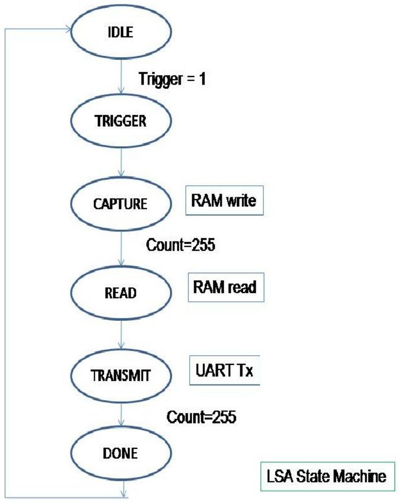


Figure 3-2. LSA State machine

The state machine contains 6 states namely IDLE, TRIGGER, CAPTURE, READ, TRANSMIT and DONE. The state machine stays in IDLE state by default. When trigger input is ‘1’, the state machine enters TRIGGER state. The state machine enters CAPTURE state after that. Memory write takes place in CAPTURE state only. 256 memory writes are done in this state which corresponds to 256 samples being captured.

Once sampling is finished, the state machine enters READ state where RAM read takes place. The read back data is transmitted through UART in TRANSMIT state. Once 256 data transmission finish, the state machine enters DONE state which indicates end of signal capture. The state machine goes back to IDLE state.The block diagram of match unit is shown in Figure 3-3.

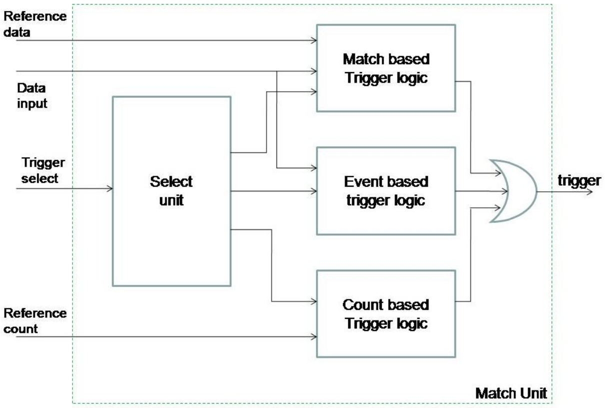


Figure 3-3. Match unit

As shown above, the match unit contains 4 sub blocks namely select unit, match based trigger logic, event based trigger logic and count based trigger logic.

Select unit: The function of select unit is to generate enable signals for other 3 blocks based on the trigger select. 3-bit trigger select is used since there are 8 conditions in total to be enabled.

Match based trigger logic:

This block contains logic to implement the trigger logic based on match condition. It contains 2 data inputs for comparison namely reference data and data input. Reference data is the data with which matching needs to be done and data input is the current sampled data. Various match condition supported are

Din = ref data, Din /= ref data, Din < ref data, Din > ref data, Din <= ref data, Din >= ref data

There are 6 enable signals to this unit to select one of the above match conditions.

Event based trigger logic:

The functionality of this block is to detect an event on the data input. It compares two successive samples to detect the event. To avoid unnecessary glitches pick up, both the current data and previous data latched (or synchronized) versions are used.

Count based trigger logic:

This block contains a counter internally which will be counting the number of samples. The purpose of this block is to enable trigger after a particular number of samples are finished. In this sense, counter based triggering is delayed triggering based on the input sample count. The reference count input decides the number of samples to be discarded (or delayed).

The outputs of these 3 trigger logic modules are logically ORed to generate final trigger signal which would go to the state machine in LSA top.

The block diagram of host interface unit is showninFigure3-4.

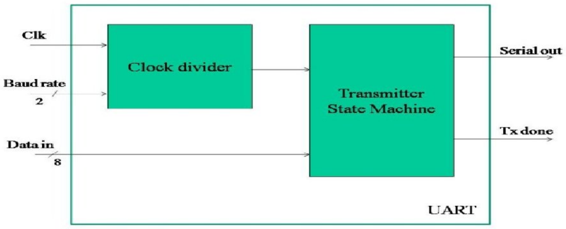


Fig 3-4. Block diagram of host interface unit

The UART contains 2 modules namely clock divider and transmitter state machine. Clock divider generates different clocks corresponding to different baud rates selected. UART transmitter state machine is shown in Figure 3-5.

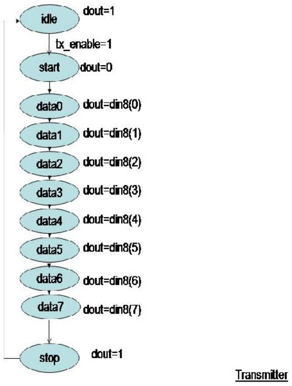


Fig 3-5. UART transmit state machine

The transmitter stays in IDLE state unless transmit enable (tx_enable) is made as ‘1’. The data transmission starts with tx_enable = 1. As mandated by the protocol, a ‘0’ is transmitted to indicate start of transmission or start bit. This is done in START state. Then data bits 0 to 7 are transmitted in states DATA0 to DATA7. If parity is enabled in configuration register, the data is attached with a parity in PARITY state. Then transmitter enters STOP state and sends a ‘1’. This indicates the completion of transmission. Then the transmitter enters the IDLE state and waits for next data transmission. An asynchronous FIFO is used as Trace memory in LSA communication module. A FIFO is required since the output data rate of UART is much smaller than the sampling rates supported by capture module.

RTL schematic:

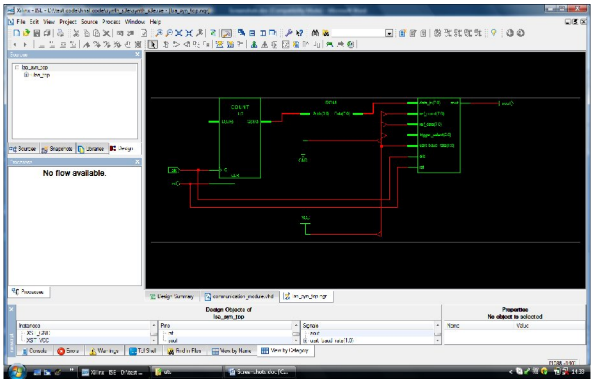


Fig.3.6 RTL schematic for our design

4. SIMULATION RESULTS

The simulation results for capture module are shown in Figure 4-1.

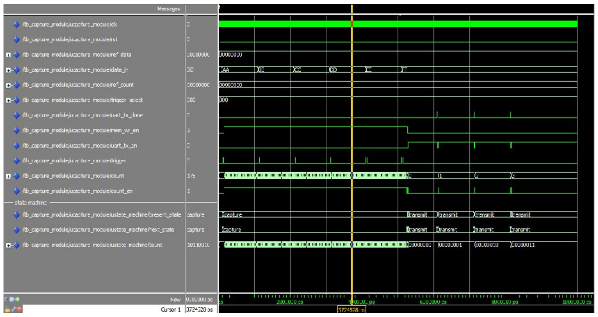
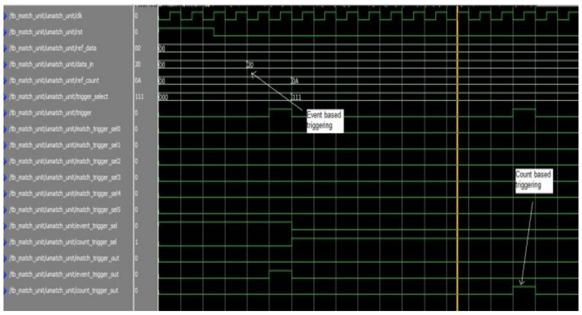


Figure 4-1. Waveforms for capture module

In the waveform data_in refers to the data being captured by LSA. Trigger_select value selects the type of triggering to be used. Ref_data and ref_count are used for different triggering mechanisms of match unit. The signals after divider are state machine signals.

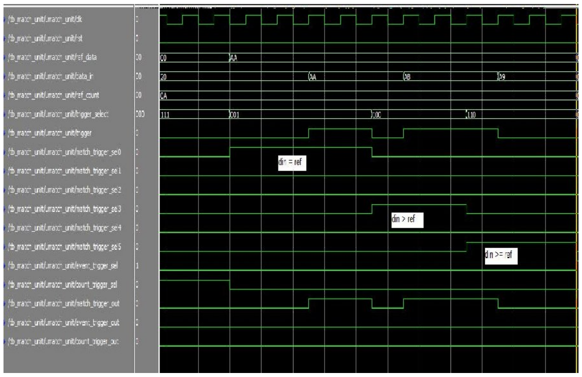
Figure 4-2 shows the simulation waveforms for even based trigger logic and count based trigger logic at the top level.



In this screenshot, initially even based triggering is enabled (trigger select = 000). When the input data (data_in) changes from X”00” to X”20”, an event is triggered. We can see the corresponding output on event_trigger_out first and then on trigger output.

Later, count based triggering is enabled with trigger_select = 111. Here a counter is started in count based trigger logic and when it reaches the reference count (ref_count) value of X”0A” a trigger is generated.

Figure 4-3 shows match based trigger logic waveforms from top level.



Here 3 match conditions are exercised, din=ref, din>ref and din>=ref. As we can see first trigger comes when the data_in value reaches the ref_data of X”AA” since the match_trigger_sel0 checks for equality condition.

Then match_trigger_sel3 is enabled which means that din>ref condition is checked. When the data_in value changes to X”AB” which is greater than ref_data of X”AA”, trigger is generated. Third condition set is din>=ref. Since it is already satisfied the trigger continues to be generated. When the data_in changes to X”A9”, trigger becomes 0.

Figure 4-4 shows the simulaton results for communication module.

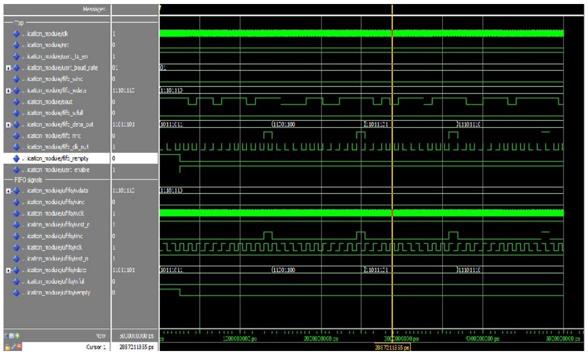


Figure 4-4. Communication module waveforms

.Figure 4-4 shows 2 set of signals one for top level communication module waveforms and the other is FIFO signals. For the communication module, data comes in as *fifo_wdata* from capture module. This data is saved in FIFO. This data is read back from FIFO as *fifo_data_out* and is given to UART. Final UART serial output can be seen on *sout*.

Coming to FIFO signals, we can see the *wclk* and *rclk*, clocks used for write and read. As mentioned earlier, the read clock is much slower than write clock. The *wdata* shows the write data and *rdata* is the readback data.

UART (host interface unit) simulation waveforms are shown in Figure 4-5.

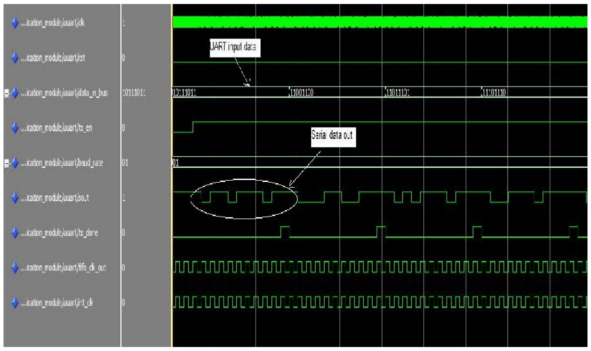

...ication_module/uuart/clk
...ication_module/uuart/rst
...ication_module/uuart/data_in_bus
...ication_module/uuart/tx_en
...ication_module/uuart/baud_rate
...ication_module/uuart/sout
...ication_module/uuart/tx_done
...ication_module/uuart/fifo_clk_out
...ication_module/uuart/int_clk
10111011
11001110
11011101
11101110
01
UART input data
Serial data out

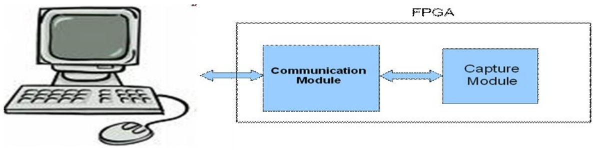
Figure 4-5. UART simulations

Figure 4-5 shows simulation for UART, where *data_in_bus* refers to the parallel data input to be transmitted.

The baud rate is controlled by *baud_rate* input and *clk* is the system clock which is 50MHs currently. The serial output from UART state machine is given on *sout* and corresponding clock can be seen on *fifo_clk_out* which goes to FIFO.

The UART transmission happens only when tx_en signal is ‘1’.these are the simulation results for our design.

Test setup:



The abow figure can represents the test setup for our design.here our design (VHDL code) is dumped into Spartan 3e FPGA board after that by using requirements we connect our setup to PC (power supply for Spartan 3e, USB connector, RS-232 cable etc..). The requirement figures shown below

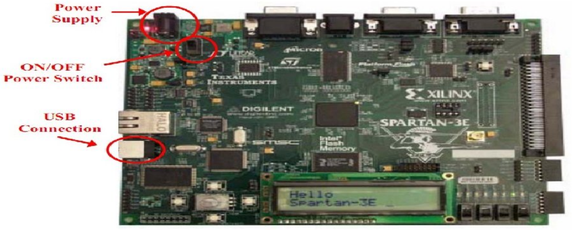


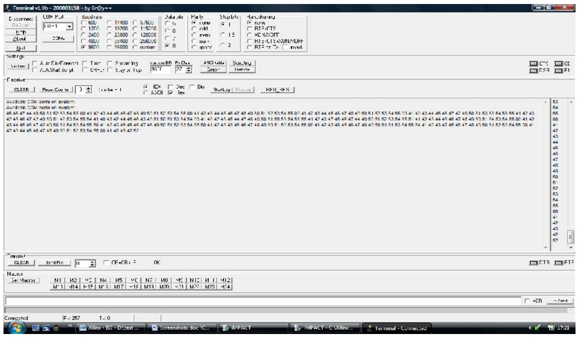
Fig 4.7 spartan 3e FPGA board

USB Type B Connector
Connects to Starter Kit's USB connector
USB Type A Connector
Connects to computer's USB connector

Fig4.8 USB connector ,4.9 USB-to-RS232 converter

HyperTerminal Results:

After test setup we are giving the input data as a hexadecimal to our design the results will be shown by using hyperterminal interms of characters(refer hex to ascii table).the results shown below.

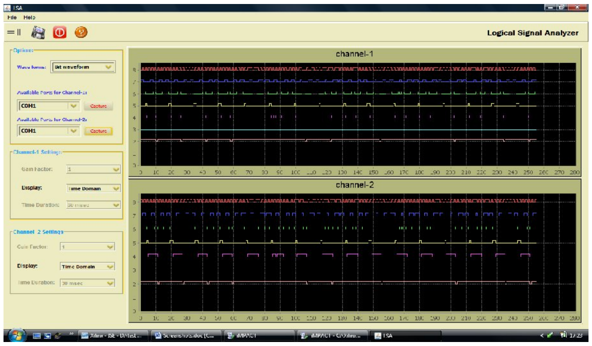

Terminal v1.9b - 20080315B - by Br@y++

Available COM ports on system:
Available COM ports on system:

Fig.4.9 hyperTerminal results

GUI results:

What we are obeserved the results from the hyperterminal we will see that results with respect to bit waveform by using java based GUI setup file for LSA.The below figure can represents the GUI results for LSA



Wave forms: Bit waveform

Available Ports for Channel-1: COM1

Available Ports for Channel-2: COM1

Channel-1 Settings

Gain Factor: 1

Display: Time Domain

Time Duration: 30 msec

Channel 2 Settings

Gain Factor: 1

Display: Time Domain

Time Duration: 30 msec

Fig. 4.10 GUI results for LSA

5. CONCLUSIONS

A low cost Logic Signal Analyzer is designed targeted to Xilinx FPGA. The design contains 16-channels and samples at a maximum frequency of 25Msps. The design is functionally verified using ModelSim simulator. Hardware verification is done using a test board with PSoC microcontroller. The design occupies 20% of Spartan 3E FPGA XC3S500E. A Java based GUI is used to plot and verify the results on PC.

6. FUTURE SCOPE

The design can be implemented as ASIC in future to get better speeds. In ASIC form, the ADC portion also can be included on chip to get speeds over 100Msps.Alternatively the design can also be implemented on high end FPGAs like Virtex4 or Virtex5 for better speeds.

6. REFERENCES

[1] Integrating logic analyzer functionality into VHDL designs,2008 conference.
[2] Altera Corporation, "*SignalTap II Embedded Logic Analyzer Documentation*".
[3] First Silicon Solutions, "*FPGAView Software*", www.fs2.com/fpgaview.html
[4] P. S. Graham, "*Logical Hardware Debuggers for FPGA-Based Systems*", PhD Dissertation, Bringham Young University.
[5] www.agilent technologies.com
[6] www.wikkipedia/freeencyclopedia.com
[7] www.xilinx .com/ise
[8]www.tectronix.com
[9] www.java/netbeans.com
[10] www. Xilinx.com/s3eboards
[11] Apply Error Vector Measurements in Communications Design, ken Voelker, Microwaves & RF, December 1995.
[12] Vector signal analysis of digital baseband and if signals within an FPGA", autotestcon, 2005. Ieee
[13] Testing of digital circuitry using xilinx chipscope logic analyzer", cas 2005 proceedings, orest oltu, petru lucian milea.
[14] Remote Logic Analyzer Implemented onFPGA", Proceedings of the Argentine School of Micro Nano electronics, Technology and Applications 2008,Luisa García, Alejandra González, Henry Moreno

A New High-Speed Architecture for Reed-Solomon Decoder

C. Rammohan
mailme_rammohan28@reddiffmail.com
SVPCET, puttur, AP
Mobile no:+91-9505056746

G.sunil, M.Tech
Asst Prof
SVPCET, puttur, AP

Introduction:

Abstract—This paper proposes a new VLSI architecture for decoding Reed-Solomon codes with a modified Berlekamp Massey algorithm. By employing t-folded architecture, we achieve the highest throughput and the resource utilization efficiency without degrading performance on critical path delay. More interestingly, on the basis of the proposed architecture, further complexity benefit can be realized by sharing hardware units among sub-blocks, which is usually neglected in previous research. Two algorithms using this sharing technique are given and demonstrated to reduce the hardware complexity dramatically. Compared to the current commercial IP core, the proposed architectures are more advantageous in a certain content of the characteristics.

Scope of the Project:

Nearly all the decoding schemes reported have been focused on the improvement of a separate sub block, and failed to consider the possible sharing of hardware resource among sub-blocks. Another drawback of the algorithms available today is that both the ability to support continuous decoding (inputting data with no gap between code blocks) and the resource utilization efficiency are neglected when an algorithm is evaluated. This project will show how the hardware complexity can be substantially reduced by merging into key equation solver (KES) block some computation of other sub-blocks. Surprisingly, this goal can be achieved without degrading the performance on critical path delay, throughput or regularity.

Literature survey:

Non-binary BCH codes defined over GF(qm) of length n = qm – 1 are called Reed–Solomon codes [4] and are the best-performing BCH codes. Hence, the non-binary BCH code of Example 3.4 is the (15, 11, 5) Reed–Solomon code. Reed–Solomon codes also differ from other BCH codes in that their minimum Hamming distance is equal to the designed minimum distance. An important upper bound on the minimum Hamming distance in error-correction is the Singleton Bound, defined as

$$d \leq n - k + 1.$$

Any code with a minimum Hamming distance d = n – k + 1 is known as maximum distance separable (MDS) and has optimal minimum Hamming distance. Therefore, Reed–Solomon codes can correct

$$t = \left\lfloor \frac{d-1}{2} \right\rfloor = \left\lfloor \frac{n-k+1-1}{2} \right\rfloor = \left\lfloor \frac{n-k}{2} \right\rfloor.$$

Rearranging equation gives a message length of k = n – 2t. In summary, Reed–Solomon codes defined over GF(qm) have the following parameters :

Codeword length n = qm − 1
Message length k = n − 2t
Minimum Hamming distance d = n − k + 1

The error-correcting ability of any Reed–Solomon code is determined by n − k, the measure of redundancy in the block. If the locations of the errored symbols are not known in advance, then a Reed–Solomon code can correct up to (n − k) / 2 erroneous symbols, i.e., it can correct half as many errors as there are redundant symbols added to the block. Sometimes error locations are known in advance (e.g., “side information” in demodulator signal-to-noise ratios)—these are called erasures. A Reed–Solomon code (like any MDS code) is able to correct twice as many erasures as errors, and any combination of errors and erasures can be corrected as long as the relation

$$2E + S \leq n - k$$

is satisfied, where E is the number of errors and S is the number of erasures in the block.

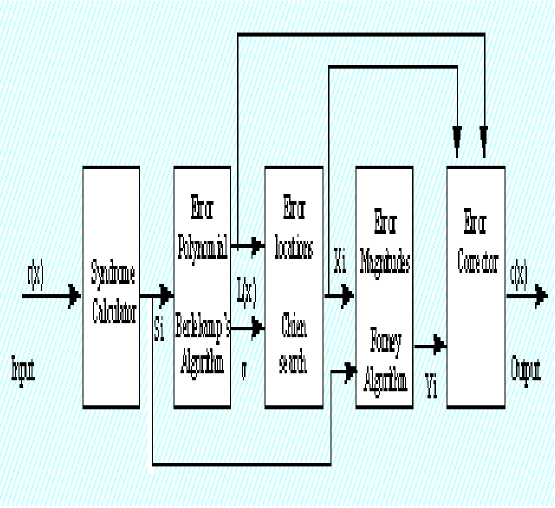
The properties of Reed–Solomon codes make them especially well-suited to applications where errors occur in bursts. This is because it does not matter to the code how many bits in a symbol are in error—if multiple bits in a symbol are corrupted it only counts as a single error. Conversely, if a data stream is not characterized by error bursts or drop-outs but by random single bit errors, a Reed–Solomon code is usually a poor choice.

Decoding Reed–Solomon Codes:

To decode binary codes only the locations of the errors in the received word are required, since the value at these location can be flipped, that is a ‘1’ becomes a ‘0’ and vice versa. However, for a non-binary code an error value can be many different values and a

RS Decoder Block Diagram:

A general architecture for decoding Reed-Solomon codes is shown in the following diagram.



$\Lambda(x)$ and an error-magnitude polynomial $\Omega(x)$. These two polynomials are related to

Syndrome Calculation & Error Detection
each other by the key equation, given as a non-binary block code: an error-locating algorithm and an error-evaluation algorithm.

For decoding we will make use of two polynomials: an error-locating polynomial

$$\Lambda(x)[1 + S(x)] \equiv \Omega(x) \bmod x^{2t+1}$$

where S(x) is the syndrome polynomial
secondary process is needed to evaluate the error value. So, two algorithms are required to decode

Algorithms:

A typical RS decoder consists of three parts: syndrome computation (SC) block, key equation solver (KES) block and Chien search and error evaluator (CSEE) block . Let N and t be the code length and error-correction capacity respectively. The received polynomial can be denoted as R(x)=RN-1xN-1+...+R1x+R0. Given the fact that the code symbols are fed into RS decoder one by one in most of the current applications, a serial calculation based on Horner's rule is particularly suitable for implementing SC block, which takes N clock cycles. A common method used in CSEE block is exhaustive searching-all possible locations from RN-1 to R0, and the latency is N clock cycles. By contrast, the delay of KES block varies significantly from an algorithm to another. Therefore, KES block usually plays a central role in RS decoder

Conventional Algorithms :

The coefficients of error locator polynomial σ(x) and error value polynomial Δ(x) in riBM algorithm are updated as

$$\begin{cases} \sigma_j(k+1) = \gamma(k)\cdot\sigma_j(k) - \delta(k)\cdot B_{j-1}(k), & j=0,1\ldots t \\ \Delta_j(k+1) = \gamma(k)\cdot\Delta_{j+1}(k) - \delta(k)\cdot\theta_j(k), & j=0,1\ldots 2t\text{-}1 \end{cases} \quad (1)$$

As to the implementation, the 6t+2 multipliers in KES block allow all the coefficients of these polynomials to be updated in one clock cycle. UiBM is a completely-serial version of riBM. It is true that this algorithm demands far less multipliers than riBM algorithm. However, the enormous latency, as much as 6t2, makes it very difficult to support high throughput or continuous decoding.

Modified Algorithm:

To express (1) in the same form, we substitute σj (k) and Bj (k) for σt −j (k) and Bt −j (k) respectively. And then

$$\sigma_j(k+1) = \gamma(k)\cdot\sigma_j(k) - \delta(k)\cdot B_{j+1}(k), \quad j = 0, 1 \ldots t \quad (2)$$

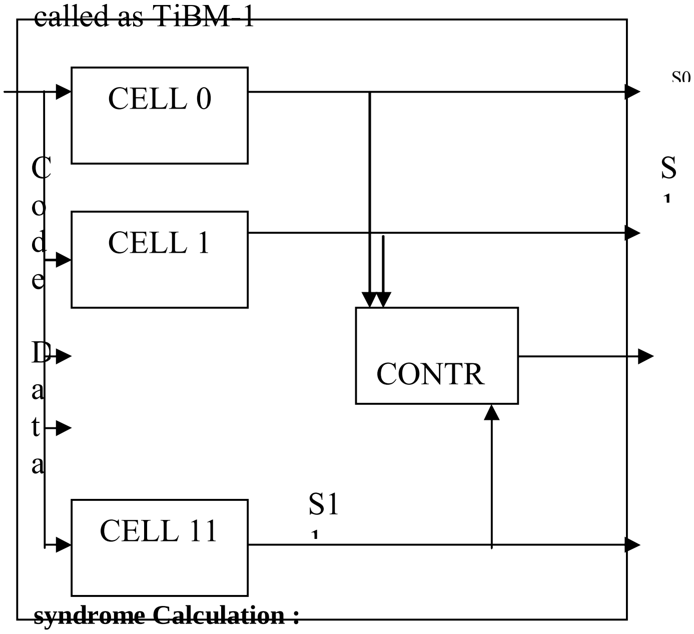
Define ΔH, ΔL, θH, θL as follows:

$$\begin{cases} \Delta H_j(k) = \Delta_j(k), & \theta H_j(k) = \theta_j(k) \\ \Delta L_j(k) = \Delta_{t+j}(k), & \theta L_j(k) = \theta_{t+j}(k) \end{cases}, \quad j = 0, 1 \ldots t \quad (3)$$

Thus, the coefficients of polynomials are updated as

$$\begin{cases} \Delta H_j(k+1) = \gamma(k)\cdot\Delta H_{j+1}(k) - \delta(k)\cdot\theta H_j(k) \\ \Delta L_j(k+1) = \gamma(k)\cdot\Delta L_{j+1}(k) - \delta(k)\cdot\theta L_j(k) \end{cases}, j = 0, \ldots t \quad (4)$$

This means that the coefficients σj(k+1), ΔHj(k+1) and ΔLj(k+1) can be computed in one clock cycle. In this way, it would take t+1 clock cycles-which is called a round of iteration-to obtain the polynomials σ(k+1,x), ΔH(k+1, x) and ΔL(k+1, x) from σ(k, x), ΔH(k, x) and ΔL(k, x) respectively. Thus, parameter δ, L and γ are needed to be kept unchanged during a round of iteration, and be updated after every round of iteration is finished. The modified algorithm is called as TiBM-1



syndrome Calculation :
Block Diagram:

Syndrome calculation & Error Detection block contain mainly

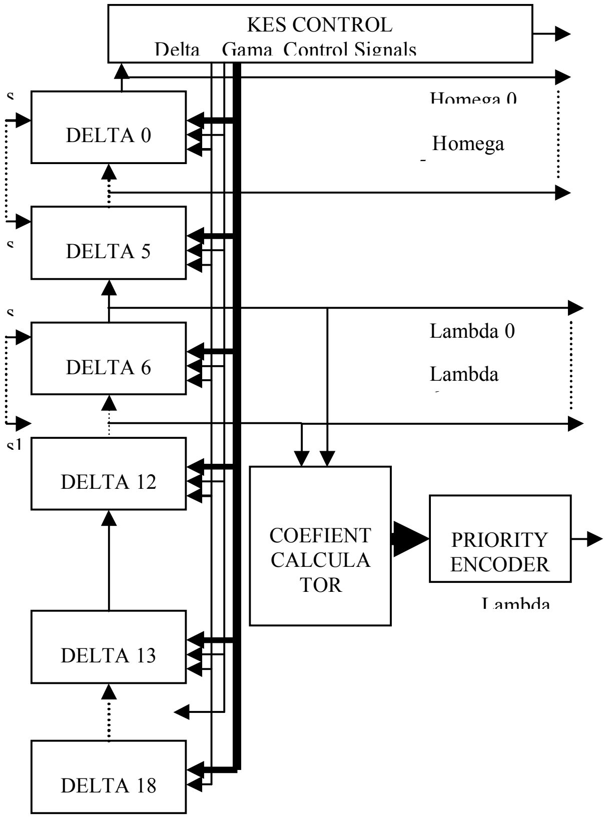
1. Cell
2. Control FSM

Each Cell consist of
Adder (GF)
Multiplier (GF)
D-Flip Flop

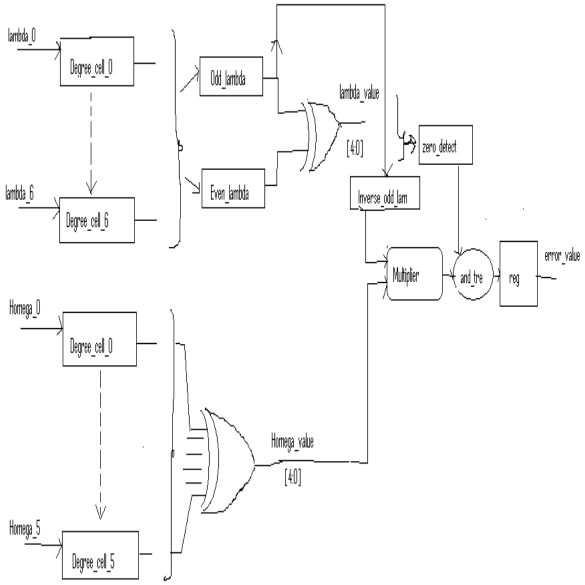
Adder will add previous value with Alpha which is constant. Multiplier will adder output with code data input. This output will give to D-Flip Flop for providing delay.

Syndrome calculator output will be zero if no error occurs. Any value indicates in syndrome calculator will give to Key equation solver for further calculation.

BLOCK DIAGRAM :
CSEE BLOCK :



Delta cell



Advantages:

As to the implementation, the 6t+2 multipliers in KES block allow all the coefficients of these polynomials to be updated in one clock cycle. As a result, riBM algorithm requires only 2t clock cycle to solve the key equation. However, the short delay of KES block is hardly conducive to an overall performance except for the decoding latency when applied to various practical applications. Take DVB1 standard (N=204, t=8) for example. The KES block, the most resource-consuming part, has to stay in idle for N−2t=188 clock cycles during the process of every code block, which means a fairly low resource utilization efficiency

Basically, UiBM is a completely-serial version of riBM. It is true that this algorithm demands far less multipliers than riBM algorithm. However, the enormous latency, as much as 6t2, makes it very difficult to support high throughput or continuous decoding. In the case of IEEE802.16d standards (N=255, t=8), due to 6t2>N, code blocks can not be continually fed into the decoder, and the throughput pales in comparison with riBM algorithm. It is clear that the problems discussed above should be ascribed to the large gap between KES block latency and code length N. Consequently, any techniques to render the delay of KES block close to, but less than, N by means of a trade-off between latency and hardware complexity would result in architectures with high throughput and high resource utilization efficiency, as well as, the competence for continuous decoding. That is exactly the starting point for developing our methods.

Applications :

Reed-Solomon codes are block-based error correcting codes with a wide range of applications in digital communications and storage. Reed-Solomon codes are used to correct errors in many systems including:

- ❖ Storage devices (including tape, Compact Disk, DVD, barcodes, etc)
- ❖ Wireless or mobile communications (including cellular telephones, microwave links, etc)
- ❖ Satellite communications
- ❖ Digital television / DVB
- ❖ High-speed modems such as ADSL, xDSL, etc.

Simulation Results :

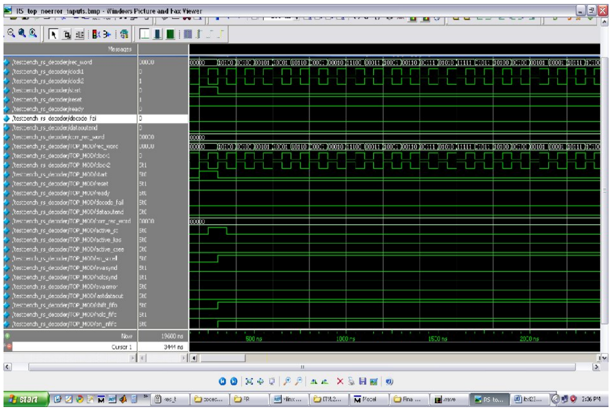


Fig shows bit error location in the received codeword in RS decoder

Synthesis Results:

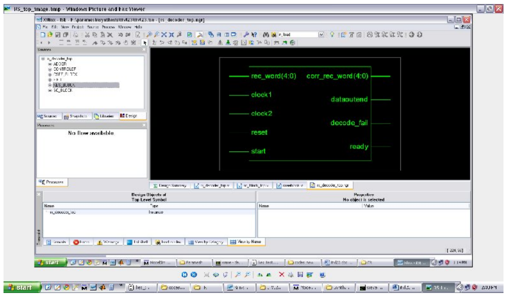


Fig shows RTL schematic of RS decoder

```
RTL Top Level Output File Name     : rs_decoder_top.ngr
Top Level Output File Name         : rs_decoder_top
Output Format                      : NGC
Optimization Goal                  : Speed
Keep Hierarchy                     : NO

Design Statistics
# IOs                              : 17

Cell Usage :
# BELS                             : 1619
#      GND                         : 1
#      INV                         : 8
#      LUT2                        : 237
#      LUT3                        : 160
#      LUT3_D                      : 1
#      LUT3_L                      : 19
#      LUT4                        : 846
#      LUT4_D                      : 2
#      LUT4_L                      : 187
#      MUXF5                       : 156
#      MUXF6                       : 1
#      VCC                         : 1
# FlipFlops/Latches                : 519
#      FD                          : 2
#      FDC                         : 12
#      FDE                         : 401
#      FDR                         : 97
#      FDRE                        : 7
# Clock Buffers                    : 2
#      BUFGP                       : 2
# IO Buffers                       : 15
#      IBUF                        : 7
#      OBUF                        : 8
=========================================

Device utilization summary:
---------------------------

Selected Device : 3s400ft256-5

 Number of Slices:                781  out of   3584
21%
 Number of Slice Flip Flops:      519  out of   7168
7%
 Number of 4 input LUTs:         1460  out of   7168
20%
 Number of IOs:                    17
 Number of bonded IOBs:            17  out of    173     9%
 Number of GCLKs:                   2  out of      8    25%
TIMING REPORT

---------------
Speed Grade: -5

   Minimum period: 7.378ns (Maximum Frequency:
135.538MHz)
   Minimum input arrival time before clock: 6.411ns
   Maximum output required time after clock: 8.295ns
   Maximum combinational path delay: No path found
```

Conclusion:

In conclusion, with three processing elements for solving Key Equation, TiBM-1 strikes better balance between area and latency than UiBM or riBM, and hence gains complexity benefit without degrading performance on overall throughput or critical path delay. Both TiBM-2 and TiBM-3 further reduce the hardware complexity at the expense of the increase in latency of KES block.

The former algorithm can save 2t finite field multiplexers, through merging part of syndrome-loading process into KES block. In the latter one, the sharing of hardware unit between KES and CESS allows the number of finite field multiplier used for shortened code to be cut from 2t to 2. Thus, our algorithms are highly recommended as long as the KES latency is less than code length N, since they achieve better area-delay balance in that case, in comparison with the conventional algorithms.

References:

1. J Lee H. “High-speed VLSI Architecture for Parallel Reed-Solomon Decoder”,
IEEE Trans on Very Large Scale (VLSI) Integer Syst, 2003, 11(2): pp. 288-294.
2. A New High-Speed Architecture for Reed-Solomon Decoder, 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing
3. Wikkipedia.com
4. Non-Binary Error Control Coding for Wireless Communication and Data Storage
by Ronaldo Antonio Corosavo

VHDL MODELING OF HDLC TRANSMITTER AND RECEIVER

M.R. Srinivas [1] *P. Sunil Kumar* [3]
[1] *Asst. Professor Dept of E.C.E SKTRMCE Kondair*
srinivas.reiki@gmail.com
[2] *Project associate SKTRMCE Kondair*
sunil-suni@hotmail.com

ABSTRACT

This paper implements the HDLC transmitter and Receiver. High-Level data-link control (HDLC) is a synchronous bit-oriented protocol developed for use in system network architecture (SNA) environments. HDLC can transfer data simplex, half duplex or full duplex and can support a variety of link types and topologies. HDLC can be used on point-to-point or multipoint networks over both circuit and packet switched networks.

Keywords*: HDLC, System network architecture, synchronous bit-oriented protocol, packet switched networks.*

I. INTRODUCTION

The primary goal of network architecture is to give users of a network the tools necessary for setting up the network and performing data flow control. A network architecture outlines the way in which a data communications network is arranged or structured and generally includes the concepts of levels or layers within the architecture. Each layer within the network consists of specific protocols or rules for communicating that perform a given set of functions.

Protocols are arrangements between people or processes. A data-link protocol is a set of rules implementing and governing an orderly exchange of data between layer two devices, such as line control units and front-end processors.

The main functionalities of Data Link Layer of High Level Data Link controller include Packet framing and deframing, addition of error check bits and error detection of received data.

The HDLC Transmitter and HDLC Receiver are implemented individually. The information to be transmitted is the input to the Transmitter. The Transmitter then performs packet framing which includes adding the destination address, CRC data and the delimiting flags. These frames are then transmitted by the transmitter. On successful reception, the receiver deframes the packet and checks for the authenticity of the information.The Logic of the application for HDLC Transmitter and receiver are implemented by following the steps below,

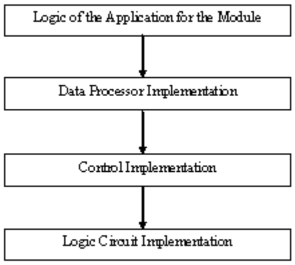

Logic of the Application for the Module
Data Processor Implementation
Control Implementation
Logic Circuit Implementation

Figure 1: Implementation steps

The Transmitter State Machine Controller or the Transmitter Control Unit is the heart of the HDLC Transmitter. For this module the Data Processor implementation is made through ASM Chart. ASM (Algorithmic State Machine) Chart is the Hardware internal operation of an application with respect to time. Then for this module the Control Implementation is made through State Diagram which emphasizes the various states and their operation. Finally, using the ASM Chart and State Diagram the Logic Circuit is implemented. The procedure is applied for developing the Logic circuit for HDLC Receiver State Machine Controller or Receiver Control Unit which is the heart of the HDLC Receiver

II. HIGH - LEVEL DATA LINK CONTROL PROTOCOL

In 1975, the International Organization for Standardization (ISO) defined several sets of sub standards that, when combined, are called High-level Data Link Control (HDLC). It falls under the ISO standards ISO 3309 and ISO 4335. It is a data link control protocol, and falls within layer 2, the Data Link Layer, of the Open Systems Interface (OSI) model. HDLC is a superset of SDLC.

Synchronous data-link control (SDLC) is a synchronous bit-oriented protocol developed in 1970s by IBM for use in system network architecture (SNA) environments. After developing SDLC, IBM submitted it to ANSI and ISO for acceptance as U.S. and international standards, respectively. ANSI modified it to become ADCCP (Advanced Data Communication Control Procedure), and ISO modified it to become HDLC.

HDLC can transfer data simplex, half duplex, or full duplex and can support a variety of link types and topologies. It can be used on point-to-point or multipoint networks and switched or non-switched networks. HDLC is a bit-oriented protocol (BOP) where there is a single control field within a message frame that performs essentially all the data-link control functions. The bit patterns are standard in HDLC and therefore the information exchanged follows the same pattern. This minimizes the chance of any errors.

HDLC Stations

HDLC specifies the following three types of stations for data link control. They are Primary Station, Secondary station and combined station.

Primary Station

The Primary Station is responsible for controlling all other secondary stations for a network that uses the HDLC protocol. It also takes care of the error control aspect and organizes the data flow on the links. The primary station's frames are called commands.

Secondary Station

The Secondary Station is controlled by the primary station and is activated when the primary station sends a request. It has no ability, or direct responsibility for controlling the link. The secondary station's frames are called responses. It can only send response frames when requested by the primary station.

Combined Station

A Combined Station is a combination of a primary and secondary station. On the link, all combined stations are able to send and receive commands and responses without any permission from any other stations on the link. Each combined station is in full control of itself, and does not rely on any other stations on the link. No other stations can control any combined station.

HDLC Configurations

HDLC also defines three types of configurations for the three types of stations. They are Unbalanced Configuration, Balanced Configuration and Symmetrical Configuration**.**

Unbalanced Configuration

In an unbalanced configuration there is one primary station and the remaining stations are all secondary stations that are controlled by the primary station. Unbalanced configuration supports many types of operations like half duplex, full duplex, point to point and multi-point configuration.

Unbalanced configuration is depicted below in Figure 2

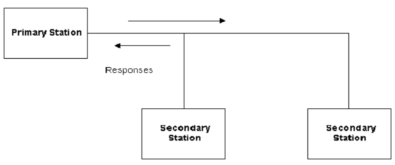


Figure 2: An Unbalanced Configuration

Balanced Configuration

In a balanced configuration there are of two or more combined stations. Each of the stations has equal and complimentary responsibility compared to each other. Balanced configurations can supports specific configurations like half duplex and full duplex operations or point to point network configurations.

Balanced configuration is depicted below in Figure 3

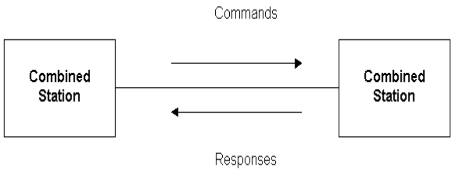


Figure 3: Balanced Configuration

Symmetrical Configuration

The Symmetrical Configuration is a combination of balanced and unbalanced configurations and is rarely used in current day technology. In this configuration, each station has a primary and secondary status. Each station is logically considered as two stations.

2.5 HDLC Modes

HDLC supports several modes of operation for the configurations discussed above. They are categorized as operational modes and non operational modes.

Operational Modes

HDLC offers three different modes of operation. These three modes of operations are:

- Normal Response Mode (NRM)
- Asynchronous Response Mode (ARM)
- Asynchronous Balanced Mode (ABM)

Normal Response Mode

Normal Response Mode is an unbalanced configuration in which the primary station usually requests information from the secondary station and only then the secondary station initiates the data transfer. The response is transferred after the primary station authorizes the transaction. This procedure is followed for each and every frame that is being transferred. After the last frame is sent or transferred then the procedure sets back itself to zero where the entire gamut of taking permission is repeated for the next set of frames. Normal Response Mode is used only with the unbalanced configurations.

Asynchronous Response Mode

Asynchronous Response Mode is an unbalanced configuration in which secondary terminals may transmit without permission from the primary terminal. However, the primary terminal still retains responsibility for line initialization, error recovery, and logical disconnect.

Asynchronous Balanced Mode

Asynchronous Balanced Mode is a balanced configuration in which either station may initiate the transmission.

Non-Operational Modes

HDLC also defines three non-operational modes. These three non-operational modes are:

- Normal Disconnected Mode (NDM)
- Asynchronous Disconnected Mode (ADM)
- Initialization Mode (IM)

Disconnected mode is the mode in which a secondary station is as before it is initialized by the primary, or when it is explicitly disconnected. In this mode, the secondary responds to almost every frame other than a mode set command with a "Disconnected mode" response. The purpose of this mode is to allow the primary to reliably detect a secondary being powered off or otherwise reset.

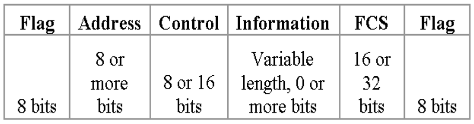
The two disconnected modes (NDM and ADM) differ from the operational modes in that the secondary station is logically disconnected from the link (note the secondary station is not physically disconnected from the link). The IM mode is different from the operations modes in that the secondary station's data link control program is in need of regeneration or it is in need of an exchange of parameters to be used in an operational mode.

III. DESIGN METHODOLOGY

HDLC Frame Structure

Each frame in HDLC may contain up to six fields, as shown in Figure 2.4

A beginning flag field, an address field, a control field, an information field, a frame cheek sequence (FCS) field, and an ending flag field. In multiple-frame transmissions, the ending flag of one frame can serve as the beginning flag of the next frame.



Flag	Address	Control	Information	FCS	Flag
8 bits	8 or more bits	8 or 16 bits	Variable length, 0 or more bits	16 or 32 bits	8 bits

Figure 4: HDLC Frame Structure

The fields and their use in different fame types are discussed below:

Flag field: The flag field of an HDLC frame is an 8-bit sequence with the bit pattern 01111110 that identified both the beginning and the end of a frame and serves as a synchronization pattern for the receiver.

Address field: The second field of an HDLC frame contains the address of the secondary station. If a primary station created the frame, it contains a “To” address. If a secondary creates the frame, it contains a “From” address. In many cases the address field is typically just a single byte, but an Extended Address [EA] bit may be used allowing for multi-byte addresses. A one residing in the LSB bit indicates [the end of the field] that the length of the address field will be 8 bits long. A zero in this bit location [now the first byte of a multi-byte field] indicates the continuation of the field [adding 8 additional bits]. The first [MSB] bit in the Address field indicates if the frame is a unicast or multicast message. A zero in the MSB bit location indicates a unicast message; the remaining bits indicate the destination node address. A one in the MSB bit location indicates multicast message, the remaining bits indicate the group address. The address field structure is shown below in Figure 5

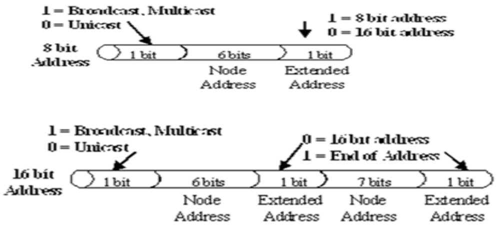


Figure 5: Address Field

Control field: The control field is an eight – bit field that identifies the type of frame being transmitted. It is used for flow and error control. The interpretation of bits in this field depends on the frame type. HDLC uses the control field(C) to determine how to control the communications process. This field contains the commands, responses and sequences numbers used to maintain the data flow accountability of the link, defines the functions of the frame and initiates the logic to control the movement of traffic between sending and receiving stations. The three control field formats are:

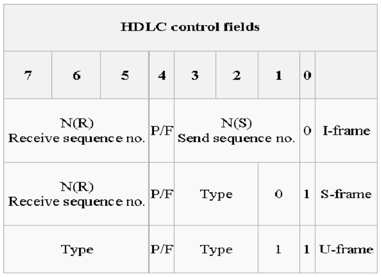
- **Information Frame:** This frame is used to transmit end-user data between two devices.
- **Supervisory Frame:** The control field performs control functions such as acknowledgment of frames, requests for re-transmission, and requests for temporary suspension of frames being transmitted. Its use depends on the operational mode being used.
- **Unnumbered Frame:** This control field format is also used for control purposes. It is used to perform link initialization, link disconnection and other link control functions.

Information field: The information field contains the user’s data from the network layer or management information. Its length can vary from one network to another.

FCS field: The fame check sequence (FCS) is the HDLC error detection field. It can contain either a 2-or 4-byte ITU-T CRC.

HDLC Protocol Operation

HDLC operation consists of the exchange of I, S, U-frames between two stations. The control field determines the type of frames as shown in Figure 6



HDLC control fields								
7	6	5	4	3	2	1	0	
N(R) Receive sequence no.			P/F	N(S) Send sequence no.			0	I-frame
N(R) Receive sequence no.			P/F	Type		0	1	S-frame
Type			P/F	Type		1	1	U-frame

Figure 6: Types of Frames

Poll/Final is a single bit with two names. It is called Poll when set by the primary station to obtain a response from a secondary station, and Final when set by the secondary station to indicate a response or the end of transmission. In all other cases, the bit is clear. The bit is used as a token that is passed back and forth between the stations. Only one token should exist at a time. The secondary only sends a Final when it has received a Poll from the primary. The primary only sends a Poll when it has received a Final back from the secondary, or after a timeout indicating that the bit has been lost.

The various commands and responses defined for the above frame types are:

Information Transfer Format Command and Response:

The functions of the information command and response is to transfer sequentially numbered frames, each containing an information field, across the data link.

Supervisory Format Commands and Responses:

Supervisory(S) commands and responses are used to perform numbered supervisory functions such as acknowledgment, polling, temporary suspension of information transfer, or error recovery. Frames with the S format control field cannot contain an information field. A primary station may use the S format command frame with the P bit set to 1 to request a response from a secondary station regarding its status.

Unnumbered Format Commands and Responses:

The unnumbered format commands and responses are used to extend the number of data link control functions. The unnumbered format frames have 5 modifier bits which allow for up to 32 additional commands and 32 additional response functions.

The operation of HDLC involves three phases:

First, one side or another initialize the data link so that frames may be exchanged in an orderly fashion.

After initialization, the two sides exchange user data and control information to exercise flow control and error control.

Finally, one of the two sides signals the termination of the operation.

Initialization:

Initialization may be requested by either side by issuing one of the six set-mode commands. This command serves three purposes:

1. It signals the other side that initialization is requested.
2. It specifies which of three modes (NRM, ABM, and ARM) is requested.
3. It specifies the whether 3 or 7 bit sequence numbers are to be used.

If the other side accepts this request, then HDLC module on that end transmits an unnumbered acknowledged (UA) frame back to the initiating side. If the request is rejected, then a disconnected mode (DM) frame is sent.

Data transfer:

When the initialization has been requested and accepted, then a logical connection is established. Both sides may begin to send user data in I-frames, starting with sequence number 0. The N(S) and N(R) fields of the I-frame are sequence numbers that support flow control and error control .An HDLC module sending a sequence of I-frames will number them sequentially. S-frames are also used for flow control and error control. The RR frame is used to acknowledge the last I-frame received by indicating the next I-frame expected. RNR acknowledges an I-frame, as with RR, but also asks the peer entity to suspend transmission of I-frames.

REJ initiates the go-back-n ARQ, which indicates the last I-frame received has been rejected. SREJ is used to request transmission of just a single frame.**Disconnect:** Either HDLC module can initiate a Disconnect, either on its own or at the request of its higher layer user. DLC issues a Disconnect by sending a (DISC) frame. The other side must accept disconnect by replying with an unnumbered acknowledgement (UA).

IV. BLOCK DIAGRAM

The implementation of HDLC Data link layer is divided into two parts.

- Transmitter
- Receiver.

The HDLC transmitter and receiver design is done by dividing the top order module into various sub modules.

TRANSMITTER IMPLEMENTATION:

The figure shows the block diagram of HDLC Transmitter.

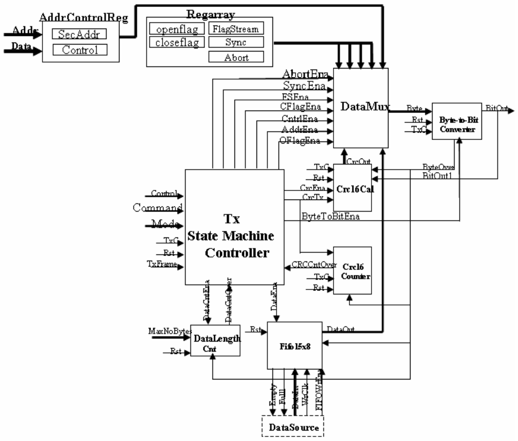


Figure 7: Transmitter Architecture

Architecture of HDLC Transmitter consists of 9 modules

AddrControlRegister:

Address control register stores the data in the registers depending on the given addr and outputs them to their respective outputs.

Regarray:

It is a Register array. Host writes the Secondary station addresses and Packet specific Control information into Register array through Data port by placing the respective address locations of Registers on address bus.

DataMux:

Data MUX works as a multiplexer to select any one data line out of 16 inputs and allow it to be transmitted out to Byte output based on that particular SelEna input signal.

Crc16Cal:

The 16 bit LFSR for the CRC is constructed using the CRC-CCITT generator polynomial g(D)= D^{16} + D^{12} + D^5 +1. For this case, the 16 bits of LFSR shall be initially loaded with the Zeros when Rst is Low. When CRCEna is asserted high the data is shifted in. After the last bit has entered the LFSR, CRCEna is made low and CRCTx signal is asserted high so that the register's contents shall be transmitted, from right to left (i.e., starting with position 15, then position 14, etc.). CRCOver is made high when all 16 bits LFSR are transmitted.

Crc16Counter:

CRC Counter is used to maintain the count for number of bytes the CRC is transmitted. When CRCTx is asserted high, CRC Counter starts incrementing for every byte transmission. CRCCntOver is made high when entire 16 bits of CRC are transmitted out.

FIFO15x8:

FIFO is memory device used to store data temporarily with a protocol First In and First Out bases. This is a Asynchronous FIFO means Write and Read operations are performed using the different clocks (Speed may be same/different) which are with different phases (The phase difference is not zero). FIFO asserted the status signals according to the status of the FIFO.

DataLengthCnt:

Data Length Counter is used to maintain the count for number of bytes the Data is transmitted. When DataLengthEna signal is asserted high, Data Length Counter starts incrementing for every byte transmission. DataCntOver is made high when entire Text Data is transmitted out.

Byte To Bit Converter:

Byte to Bit Converter is used to convert byte data received at one clock cycle into 8 bits i.e., 8 clock cycles to transmit 8 bits of data. When ByteToBitEna is enabled then the converted byte is given as output from BitOut1.After tranmission of a Byte, then ByteOver is enabled.

Tx State Machine Controller:

The heart of the HDLC transmitter is Transmitter control state machine, it controls each and every block of the transmitter. Hence it is also called Transmitter Control Unit.

ASM CHART OF TRANSMITTER STATE MACHINE CONTROLLER

The flow chart of the design is given above which gives a graphical representation of the design process of transmitter that is carried out. The data to be transmitted will be sent to the RegArray of the transmitter device and depending on the status of the control signal read or write operation is performed is performed and then the data is forwarded to the CRC generator where the FCS is generated and also to the frame builder where the frame is build and then the parallel data is converted to serial for transmission and send over the network

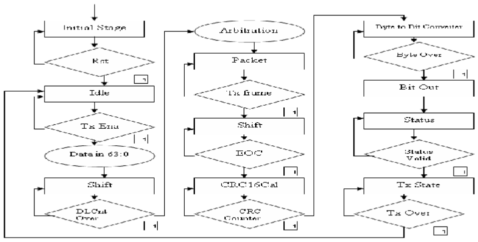


Figure 8: Transmitter ASM Chart

RECEIVER IMPLEMENTATION:

The figure shows the block diagram of HDLC Receiver.

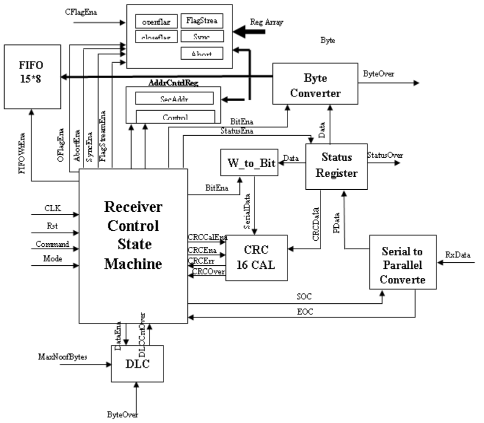


Figure 9: Receiver Architecture

Architecture of HDLC Receiver consists of 10 modules

AddrControlReg:

Address control register stores the data in the registers depending on the given particular address registers.

RegArray:

Host Receive the Secondary station addresses and Packet specific Control information into Register array through Data port.

CRC16CAL:

The 16 bit LFSR for the CRC is constructed using the CRC-CCITT generator polynomial g(D)= $D^{16} + D^{12} + D^{5} +1$. For this case, the 16 bits of LFSR shall be initially loaded with the Zeros when Rst is Low. When CRCEna is asserted high the data is shifted in. After the last bit has entered the LFSR, CRCEna is made low and CRCCalEna signal is asserted high so that the register's contents i.e., the CRCData coming form the Transmitter and the CRC generated in the receiver is compared. During this comparison when the CRCErr signal asserts high then it again asks the transmitter for the same packet to transmit, if CRCOver signal asserts to high, it declares that there is no error in the CRC.

FIFO:

FIFO is memory device used to store data temporarily with a protocol First In and First Out bases. This is a Asynchronous FIFO means Write operation is performed here.

DLC:

Data Length Counter is used to maintain the count for number of bytes the Data is received and stored in the FIFO. When DataLengthEna signal is asserted high, Data Length Counter starts incrementing for every byte transmission. DataCntOver is made high when entire Text Data is transmitted out.

Serial TO Parallel Converter:

Serial to Parallel converter is used to convert the serial data which is coming form the transmitter is stored in the inter buffers of this module and the parallel data is send out to this module.

Status Register:

Status Register divides the parallel data into the CRCData and the data. This data field contains, RegArray data and the AddrCntrl reg data and the actual data.

W_to_Bit:

W_to_Bit is the word to bit converter .Status register out put is given in the form of Word data. This word data is converted into serial bits and this serial data is given to CRC calculation.

Byte Converter:

Byte converter is used to convert the data which is coming form the Status register (in the form of word-32 bits) is converted into the byte.

Receiver Control State Machine:

The heart of the HDLC receiver is Receiver control state machine, it controls each and every block of the Receiver. Hence it is also called Receiver Control Unit.

ASM CHART OF RECEIVER STATE MACHINE CONTROL

The data is transmitted over the protocol network and at the receiving side the data is reconverted into parallel form and this data is fed into the status register and send to the frame reader where it is decomposed into different fields. The destination address field is fed to RegArray to check if the frame is addressed to this particular receiver or not. If the address matches then the FCS is fed to CRC checker to check for its correctness and if the frame is proved to be received without any error then the frame is processed else the frame is discarded.

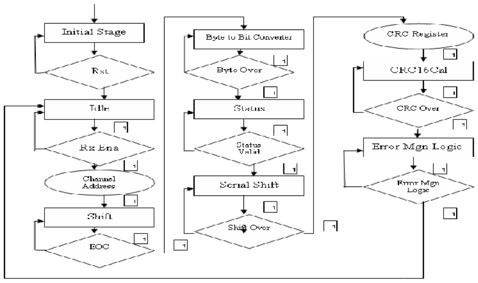


Figure 10: Receiver ASM Chart

V. RESULT OBSERVATIONS

Transmitter

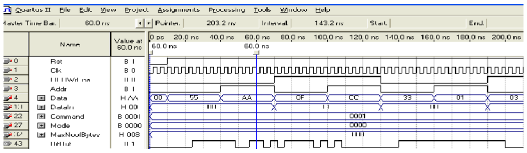


Figure 11: Result of Tranamitter

Flow Status	Successful - Fri Oct 03 15:19:11 2008
Quartus II Version	8.0 Build 215 05/09/2008 SJ Web Edition
Revision Name	TopModule
Top-level Entity Name	TopModule
Family	CycloneII
Device	EP2C5F256C6
Timing Models	Final
Met timing requirements	Yes
Total logic elements	292/ 4,608 (6%)
Total pins	39/ 158 (25 %)
Total virtual pins	0
Total memory bits	0 / 920,448 (0 %)
DSP block 9-bit elements	0 / 48 (0 %)
Total PLLs	0 / 6 (0 %)
Total DLLs	0 / 2 (0 %)

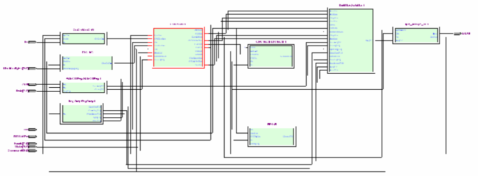


Figure 12: RTL View of Transmitter

The above Figure 5.2 shows the RTL View of the HDLC Transmitter Top Order Module. It represents the logic diagram of HDLC Transmitter. The User Data is stored in FIFO. The Transmitter State Machine Controller enables the transmission of Frames. By using Command and Mode registers various states can be achieved. Max.No.of Bytes specify the length of the frame. When Addr is '0' then the Data is Secondary Station Address otherwise it is the Control field data. DataIn is the Information.It is transmitted when FIFOWrEna is high. When Rst is high then the frame is transmitted serially through the Bitout.

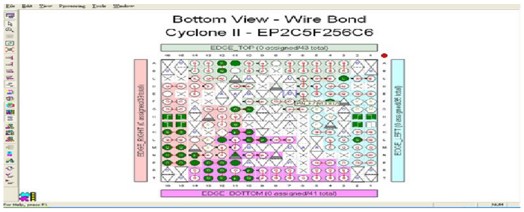


Figure 13: Chip View of Transmitter

The Figure 5.5 shows the Chip view of HDLC Transmitter Top Order Module. The internal structure of Altera devices and incrementally edit logic element (LE) and i/o cell configuration after place and route has been performed. Quartus II version 8.0 allows designers to add or remove terms in an LE's "look up table sum equation" to create or delete connections between LEs.

Receiver

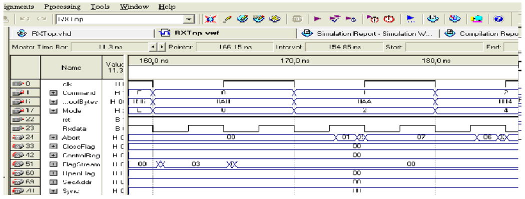


Figure 14: Result of Receiver

Flow Status Successful -	Fri Oct 03 15:29:19 2008
Quartus II Version	8.0 Build 215 05/09/2008 SJ Web Edition
Revision Name	RxTop
Top-level Entity Name	RxTop
Family	CycloneII
Device	EP2C5F256C6
Timing Models	Final
Met timing requirements	Yes
Total logic elements	304/ 4,608 (7%)
Total pins	77/ 158 (49 %)
Total virtual pins	0
Total memory bits	0 / 920,448 (0 %)
DSP block 9-bit elements	0 / 48 (0 %)
Total PLLs	0 / 6 (0 %)
Total DLLs	0 / 2 (0 %)

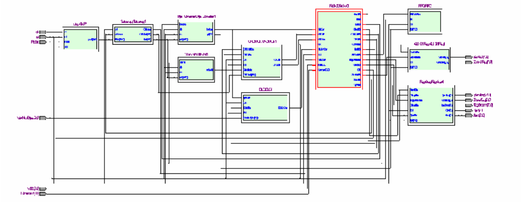


Figure 15: RTL view of Receiver

The above Figure 5.6 shows the RTL View of the HDLC Receiver Top Order Module. It represents the logic diagram of HDLC Receiver. The User Data is stored in FIFO. Clk Input synchronizes all the operations of Rx Top Module. Rst Input Initializes the Top module. RxData Input is the Data coming form the Transmitter .Max.NoofBytes Input gives Max No of Bytes. ModeInput gives the Mode Data. CommandInput gives the Command. The Receiver State Machine Controller enables the reception of the Frame.

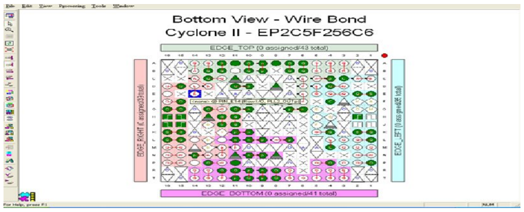


Figure 16: Chip view of Receiver

The Figure 5.9 shows the Chip view of HDLC Receiver Top Order Module. The internal structure of Altera devices and incrementally edit logic element (LE) and i/o cell configuration after place and route has been performed. Quartus II version 8.0 allows designers to add or remove terms in an LE's "look up table sum equation" to create or delete connections between LEs.

VI. CONCLUSION

In this paper a synchronization protocol for multicast network is proposed. The design is modeled with different architecture for point to point communication with simultaneous communication. Following observations were made for the implemented design, The HDLC protocol has been designed, verified functionally in the VHDL simulator, synthesized by the Quartus II and Place and Route of the design is also done. The functional simulation has been successfully carried out with the results matching the expected ones. The design has been synthesized using FPGA technology from Altera. This design has targeted the device EP2C5F256C6. Total equivalent gate count for the design: 40%. The maximum frequency with which the implementation can operate is 68.885MHz

VII. REFERENCES

Allaire, F.C.J., M. Tarbouchi, G. Labonte and G. Fusina, 2009. "FPGA implementation of genetic algorithm for UAV real-time path planning," Journal of Intelligent and Robotic Systems: Theory and Applications, 54(1-3): 495-510.

Deepthi, P.P., D.S. John and P.S. Sathidevi, 2009. "Design and analysis of a highly secure stream cipher based on linear feedback shift register," Computers and Electrical Engineering, 35(2): 235-243.

Hachicha, K., D. Faura, O. Romain and P. Garda, 2009. "Accelerating the multiple reference frames compensation in the H.264 video coder," Journal of Real-Time Image Processing, 4(1): 55-65.

Peiravi, A., 2009. "Reliability improvement of the analog computer of a naval navigation system by derating and accelerated life testing," Journal of Applied Sciences (JAS), 9(1): 173-177.

Ying, S.C. and X. Zhang, 2008. "New HDLC protocol controller based on the FPGA," Journal of Sichuan University (Engineering Science Edition), 40(3): 116-120.

Gao, Zhen-bin and Jian-Fei Liu, 2005. "FPGA implementation of a multi-channel HDLC protocol transceiver", In Proceedings of the 2005 International Conference on Communications, Circuits and Systems, 2: 1300-1302.

A Fast VLSI Design of SMS4 Cipher Based on Twisted BDD S-Box Architecture

T.Sivaramakrishna[1],
SVPCET, Puttur,
AP, India.
Siva37ram@gmail.com

P.P.Nagarajarao, M-tech[2].
Asst.Professor
Dept of ECE, SVPCET,
Puttur, AP, India
Nagraj9s@yahoo.com

Abstract: **SMS4 is a 128-bit block cipher used in the WAPI standard for protecting data packets in WLAN. In this paper, various S-box circuit architectures were evaluated firstly and the twisted BDD with m=4 was proved as the fastest one. A fast SMS4 cipher VLSI implementation was completed based on the twisted BDD S-box architecture, and achieved over 200MHz and 100MHz maximal frequency on SMIC 0.18µm and Chartered 0.35µm CMOS technology respectively.**

Keywords-SMS4 cipher; S-box; twisted BDD Architecture V LSI design

Introduction

In 2006, the Office of State Commercial Cipher Administration of China (OSCCA) released the specification of the SMS4 block cipher [1], which was employed in the Wide Authentication and Privacy Infrastructure (WAPI) standard to provide the data confidentiality in wireless networks.

To date, several studies have been performed on the SMS4 cipher, such as differential power analysis [2],differential fault analysis [3][4], the algebraic structure [5], and hardware implementations [6]. However, no research has been reported on speed evaluation and optimization of the SMS4 cipher circuit design.

In this paper, we evaluated various hardware architectures of the S-box and completed a fast SMS4 VLSI design with the twisted binary decision diagram (BDD) S-box architecture presented in [7]. The simulation results indicate that, our design achieves a 200MHz clock frequency on SMIC 0.18µm technology, and 103MHz on Chartered0.35µm technology.

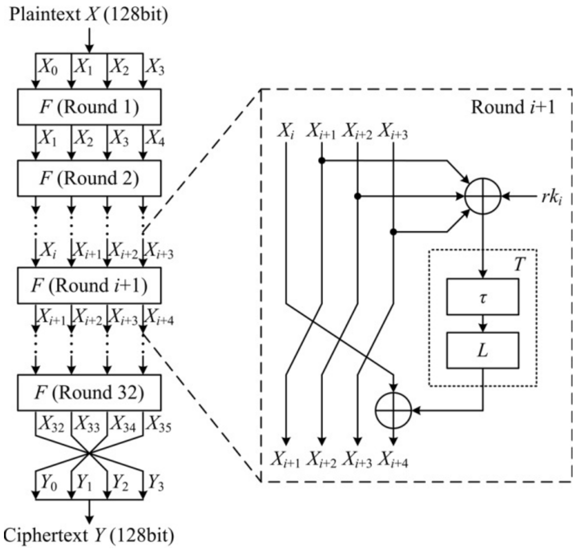


Fig.1

The rest of this paper is organized as follows. Section II describes the SMS4 block cipher. In section III, some SMS4 S-box circuit architectures are introduced briefly and our evaluation results are presented. A fast full SMS4 VLSI implementation and its simulation results are described in section IV. Finally, conclusions are reported in section V.

II.SMS4 BLOCK CIPHER

SMS4 is a 32-round iterative algorithm, and both the data block and the key size are fixed to 128 bits [1]. The encryption flow of the SMS4 cipher is showed in Fig. 1

A. Encryption Algorithm

Let X = $(X_0,X_1,X_2,X_3) \in (GF(2^{32}))^4$ be the plaintext and Y = $(Y_0,Y_1,Y_2,Y_3) \in (GF(2^{32}))^4$ be the cipher text. Let denoted by $rk_i \in GF(2^{32})$ the round keys and by $(X_i,X_{i+1},X_{i+2},X_{i+3})$ the (i+1)th round inputs, $i\in\{0,1,...,31\}$. Then the SMS4 Scheme can be written as

$$= \ (\ , \quad , \quad , \quad , \quad) \qquad (1)$$

$$= \quad \oplus \ (\quad \oplus \quad \oplus \quad \oplus \quad)$$

and

$(Y_0,Y_1,Y_2,Y_3)=(X_{35},X_{34},X_{33},X_{32})$ (2)

Where $i\in\{0,1,...,31\}$, F is the round function and T is the composite transformation.

The transformation T: $GF(2^{32}) \rightarrow GF(2^{32})$ is composed of the nonlinear transformation τ and the linear transformation L:

$$T(.) = L(\tau(.)) \qquad (3)$$

The transformation τ includes four 8-bit non linear S-boxes in parallel. Let denoted by A = $(a_0,a_1,a_2,a_3) \in (GF(2^8))^4$

The input of τ and by B = $(b_0,b_1,b_2,b_3) \in (GF(2^8))^4$ the output. Then τ can be defined as

$$=(\ , \ , \ , \ ) = (\ ) \qquad (4)$$

= (Sbox(a0),Sbox(a1),Sbox(a2)Sbox(a3))

where Sbox(.) is the S-box byte substitution which will be described in detail later.

The output of τ, B, is also the input of the linear transformation L. Let denoted by C $\in GF(2^{32})$ the output of L. Then L can be defined as

$$=(\ ) = \quad \oplus \quad <<< 2 \oplus (\quad <<< 10 \oplus \ <<< 18 \oplus \ <<< 24 \qquad (5)$$

where <<< i denotes a 32-bit cyclic left shift by i positions.

SMS4 is a symmetric key cipher, in which both the sender and the receiver use a single key for encryption and decryption. The decryption procedure of SMS4 can be done in the same way as the encryption procedure by reversing the order of the round keys.

B. Key Schedule

The key schedule process of SMS4 cipher has the same structure as that in the encryption process except for L function. Let MK = $(MK_0,MK_1,MK_2,MK_3) \in (GF(2^{32}))^4$ denote the cipher key, $rk_i \in GF(2^{32})$, $i\in\{0,1,...,31\}$ denote the round keys, and $K_i \in GF(2^{32})$, $i\in\{0,1,...,35\}$. Then key schedule algorithm is defined as

$$(\ , \ , \ , \ ) = (\quad \oplus \quad , \quad \oplus \quad , \quad \oplus \quad , \quad \oplus \quad) \qquad (6)$$

and

$$=$$

$$= \quad \oplus \ '(\quad \oplus \quad \oplus \quad \oplus \quad) \qquad (7)$$

where FK_i, $i\in\{0,1,2,3\}$ are system parameters, CK_i, $i\in\{0,1,2,3\}$ are key constants, and T' is a transformation similar to T in the encryption process. The only difference between T and T' is the linear transformation. Instead of L, the following transformation L' is used in T':

$$(\ ) = \quad \oplus \quad <<< 13 \oplus \quad <<< 23 \qquad (8)$$

The system parameters FK_i are defined in hexadecimal as

$$= 0 \quad 3 \quad 1 \quad = 06, \ 56 \quad 3350$$

$$= 0 \quad 677 \quad 9197, \ = 0 \quad 27022 \qquad (9)$$

The key constants $CK_i = (ck_{i,0}, ck_{i,1}, ck_{i,2}, ck_{i,3}) \in (GF(2^8))^4$ can be computed as follows:

$$Ck_{i,j} = (4 \times i + j) \times 7 \ (\text{mod } 256) \qquad (10)$$

where $i\in\{0,1,...,31\}$, and $j\in\{0,1,2,3\}$.

C. S-box

The S-box lookup table of the SMS4 cipher is shown as Table I [1], and the algebraic structure of the S-box can be described as [5]

$$(\ ) = (\quad . + \quad). \quad + \qquad (11)$$

Where *I* (.) is the patched inversion over *GF* (2^8), the matrices $\mathbf{A_1}$, $\mathbf{A_2} \in GL(8, 2)$, and the vectors $C_1, C_2 \in GF(2)^8$. The cyclic matrices and the row vectors in (11) are as follows:

TABLE I. SMS4 S-BOX LOOKUP TABLE

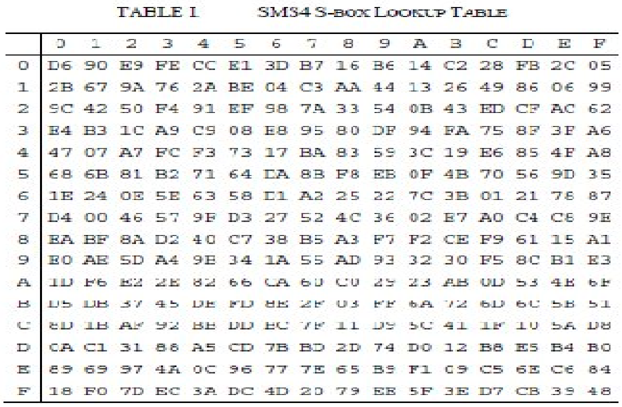


TABLE I SMS4 S-BOX LOOKUP TABLE

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	D6	90	E9	FE	CC	E1	3D	B7	16	B6	14	C2	28	FB	2C	05
1	2B	67	9A	76	2A	BE	04	C3	AA	44	13	26	49	86	06	99
2	9C	42	50	F4	91	EF	98	7A	33	54	0B	43	ED	CF	AC	62
3	E4	B3	1C	A9	C9	08	E8	95	80	DF	94	FA	75	8F	3F	A6
4	47	07	A7	FC	F3	73	17	BA	83	59	3C	19	E6	85	4F	A8
5	68	6B	81	B2	71	64	DA	8B	F8	EB	0F	4B	70	56	9D	35
6	1E	24	0E	5E	63	58	D1	A2	25	22	7C	3B	01	21	78	87
7	D4	00	46	57	9F	D3	27	52	4C	36	02	E7	A0	C4	C8	9E
8	EA	BF	8A	D2	40	C7	38	B5	A3	F7	F2	CE	F9	61	15	A1
9	E0	AE	5D	A4	9B	34	1A	55	AD	93	32	30	F5	8C	B1	E3
A	1D	F6	E2	2E	82	66	CA	60	C0	29	23	AB	0D	53	4E	6F
B	D5	DB	37	45	DE	FD	8E	2F	03	FF	6A	72	6D	6C	5B	51
C	8D	1B	AF	92	BB	DD	BC	7F	11	D9	5C	41	1F	10	5A	D8
D	CA	C1	31	88	A5	CD	7B	BD	2D	74	D0	12	B8	E5	B4	B0
E	89	69	97	4A	0C	96	77	7E	65	B9	F1	09	C5	6E	C6	84
F	18	F0	7D	EC	3A	DC	4D	20	79	EE	5F	3E	D7	CB	39	48

$$A_1 = A_2 = \begin{bmatrix} 1&1&1&0&0&1&0&1 \\ 1&1&1&1&0&0&1&0 \\ 0&1&1&1&1&0&0&1 \\ 1&0&1&1&1&1&0&0 \\ 0&1&0&1&1&1&1&0 \\ 0&0&1&0&1&1&1&1 \\ 1&0&0&1&0&1&1&1 \\ 1&1&0&0&1&0&1&1 \end{bmatrix}$$

$$= \quad = (1,1,0,0,1,0,1,1)$$

The irreducible polynomial is

$$(\quad =+ \quad + \quad + \quad + \quad + \quad + \quad +1)$$

III. S-BOX CIRCUIT ARCHITECTURES EVALUATION

As the S-box operation is the most critical part in the SMS4 encryption and key schedule process, we evaluated various S-box circuit architectures firstly.

A. Approaches

Because the SMS4 cipher was released not long ago, few studies have been reported on the S-box circuit architectures. But there are several approaches on the AES S-box circuit design [7][8], which is very similar to the SMS4 S-box.

In the AES S-box implementations, the most intuitive approach is to use the lookup table method, where the S-box circuit is synthesized from the complete S-box mapping table directly using EDA tools. The S-box circuit can be also obtained from its truth table using some logic architectures, such as sum of products (SOP), a BDD and a twisted BDD. In addition, compact S-box circuits can be designed based on mathematical operations over composite fields.

In the various S-box circuit architectures, the twisted BDD was reported as the fastest one [7]. In this method, the m levels on the output side in each BDD was replaced by a 2m:1 selector which was comprise of a select-signal decoder and a data selection part to reduce the delay, as shown in Fig. 2. With different m value, the delay of each BDD is different. When m=0, there is no selector replacement described above.

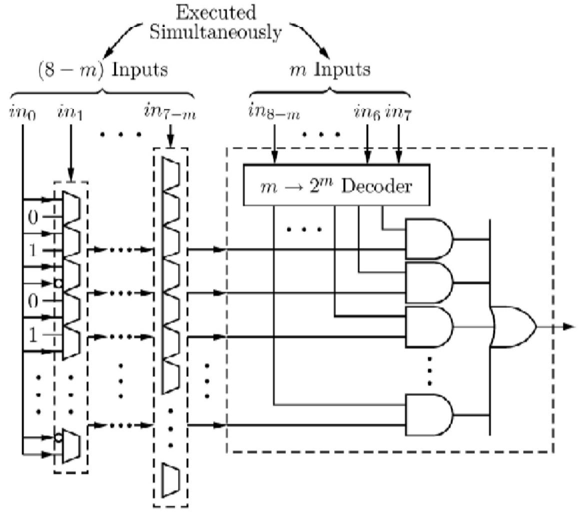


Figure 2.BDD structure in the twisted BDD architecture

B. Evaluation Results

We completed various S-box circuit architecture designs including coding, logic synthesis and physical design with the same constraint on SMIC 0.18μm and Chartered 0.35μm CMOS technology respectively. The BDDs were constructed using the CUDD package [9].

The area and delay of the S-box designs are shown in Table II and Table III. From the implementation results, we can find that the twisted BDD with *m*=4 is the fastest Architecture and it is more than 26% and 30% faster than the directly synthesized lookup table method on SMIC 0.18μm and Chartered 0.35μm technology, respectively. Although the absolute values of the delay and area also depend on the technology, EDA tools and synthesis constraints, the comparisons of the performances between the various architectures are almost the same.

IV. FAST SMS4 VLSI IMPLEMENTATION

According to the evaluation results in section III, we selected the twisted BDD with *m*=4 as the S-box architecture in our SMS4 VLSI design to obtain a higher speed.

A.BDDs Construction

We calculated out the sizes of the BDDs with all possible input orders using the CUDD package firstly. The size ranges of the BDDs are shown in Table IV. From the results, we can find that the input order of the BDD does not much affect the overall size of the BDD, which is similar to the BDDs in the AES implementation [7].

In addition, an exhaustive search for the smallest or fastest twisted BDD architecture will consume much time, because the total number of the possible input orders combinations of the 8 BDDs in the twisted BDD architecture is 8!×7!=203,212,800

TABLE II COMPARISON OF VARIOUS SMS4 S-BOX ARCHITECTURES ON SMIC 0.18μm CMOS TECHNOLOGY

Architecture	**Area(gates)**	**Delay(ns)**
Lookup table	1640.67	1.9409
Composite field	1321.33	6.7263
SOP	1690.33	1.6785
BDD	1795.67	1.7628
Twisted BDD(m=0)	1975.00	1.6765
Twisted BDD(m=3)	2092.33	1.6180
Twisted BDD(m=4)	2097.33	1.5326
Twisted BDD(m=5)	2206.33	1.6087

TABLE III COMPARISON OF VARIOUS SMS4 S-BOX ARCHITECTURES ON CHARTERED 0.35μm CMOS TECHNOLOGY

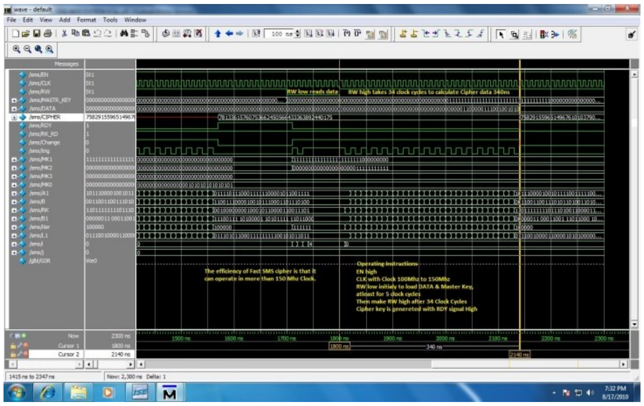
Architecture	**Area(gates)**	**Delay(ns)**
Lookup table	1595.33	3.6649
Composite field	1068.00	12.9101
SOP	1657.67	3.0659
BDD	1781.33	3.2620
Twisted BDD(m=0)	1826.67	3.1353
Twisted BDD(m=3)	1799.67	2.9788
Twisted BDD(m=4)	1788.33	2.8057
Twisted BDD(m=5)	1941.00	2.8516

TABLE IV BDD SIZE RANGES

S-box Output Bit No	0	1	2	3	4	5	6	7
Minimal BDD size	69	69	71	71	70	69	69	72
Minimal BDD size	78	79	79	78	78	78	79	79

B.Implementation Results

Simulation Results

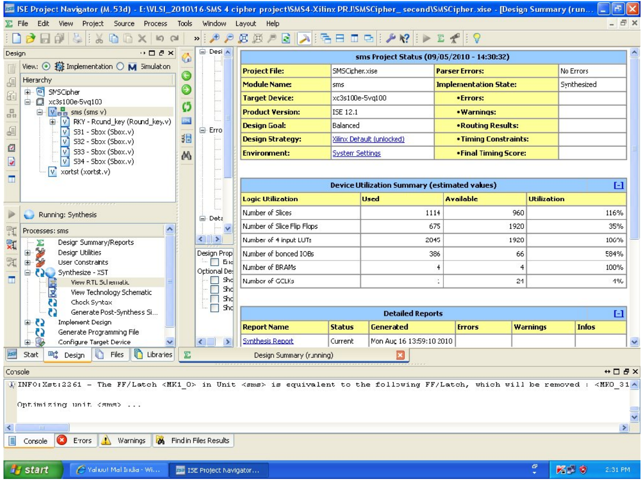

RW low reads data

RW high takes 34 clock cycles to calculate Cipher data 340ns

The efficiency of Fast SMS cipher is that it can operate in more than 150 Mhz Clock.

Operating Instructions
EN high
CLK with Clock 100Mhz to 150Mhz
RW low initialy to load DATA & Master Key, atleast for 5 clock cycles
Then make RW high after 34 Clock Cycles
Cipher key is genereted with RDY signal High

Synthesis Results:


ISE Project Navigator (M.53d) - E:\VLSI_2010\16-SMS 4 cipher project\SMS4-Xilinx PRJ\SMSCipher_ second\SMSCipher.xise - [Design Summary (run...

sms Project Status (09/05/2010 - 14:30:32)			
Project File:	SMSCipher.xise	**Parser Errors:**	No Errors
Module Name:	sms	**Implementation State:**	Synthesized
Target Device:	xc3s100e-5vq100	•**Errors:**	
Product Version:	ISE 12.1	•**Warnings:**	
Design Goal:	Balanced	•**Routing Results:**	
Design Strategy:	Xilinx Default (unlocked)	•**Timing Constraints:**	
Environment:	System Settings	•**Final Timing Score:**	

Device Utilization Summary (estimated values)

Logic Utilization	**Used**	**Available**	**Utilization**
Number of Slices	1114	960	116%
Number of Slice Flip Flops	675	1920	35%
Number of 4 input LUTs	2045	1920	106%
Number of bonced IOBs	386	66	584%
Number of BRAMs	4	4	100%
Number of GCLKs	1	24	4%

Detailed Reports

Report Name	**Status**	**Generated**	**Errors**	**Warnings**	**Infos**
Synthesis Report	Current	Mon Aug 16 13:59:10 2010			

Console

```
INFO:Xst:2261 - The FF/Latch <MK1_0> in Unit <sms> is equivalent to the following FF/Latch, which will be removed : <MKO_31

Optimizing unit <sms> ...
```

TABLE V SMS4 VLSI IMPLIMENTATION RESULTS

Technology	SSMIC 0.18μm	Chartered0.35 μm
Area(k-gates)	31.9850	27.4933
Delay(ns)	4.9927	9.6964
Max.Frequency(MHz)	200.29	103.13
Throughput	801.17	412.52

On SMIC 0.18μm technology and Chartered 0.35μm technology, we completed the RTL coding, logic synthesis and physical design of the full SMS4 cipher VLSI design with the twisted BDD S-box architecture with *m*=4. The implementation results are shown in Table V, including the area, the delay, the maximal frequency and the throughput in the cipher block chaining (CBC) mode. In the SMS4 VLSI implementation, we achieved a frequency more than 200MHz and a throughput more than

800Mbps on SMIC 0.18µm technology, and a frequency more than 100MHz and a throughput more than 400Mbps on Chartered 0.35µm technology.

C. Discussion

Based on twisted BDD architecture, we improved the operation speed of the S-box circuit and also the full SMS4 cipher VLSI design. On the other hand, from the implementation results, we can find that much of the critical path delay is used by other operations other than S-box, including XORs, multiplexors and setup time required by the technology. So, future works on improving the SMS4 circuit speed can be focused on the fast circuit architecture design of Other parts in SMS4 cipher.

V. CONCLUSIONS

In this paper, we evaluated several circuit architectures of the S-box in SMS4 cipher and completed a fast full SMS4 cipher VLSI implementation based on the twisted BDD Sbox architecture with *m*=4. According to the experiment results, our SMS4 circuit can run at speeds over 200MHz on SMIC 0.18µm technology and over 100MHz on Chartered 0.35µm technology, and achieves over 800Mbps and 400Mbps throughputs in the CBC mode respectively. The design presented in this paper is suitable for the application fields that require a high operation speed and throughput.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (No. 60606005).

REFERENCES

[1] Office of State Commercial Cipher Administration of China, “SMS4 cipher for WLAN products (in Chinese),” 2006. [Online].Available: http://www.oscca.gov.cn/UpFile/200621016423197990.pdf

[2] X. Bai, L. Guo, and T. Li, “Differential power analysis attack on SMS4 block cipher,” in *Proceedings of 4th IEEE International Conference on Circuits and Systems for Communications, ICCSC 2008*, Shanghai, China, May 2008, pp. 613–617.

[3] L. Zhang and W. Wu, “Differential fault analysis on SMS4 (in Chinese),” *Chinese Journal of Computers*, vol. 29, no. 9, pp. 1596– 1602, 2006.

[4] W. Li and D. Gu, “An improved method of differential fault analysis on the SMS4 cryptosystem,” in *Proceedings of 1st International Symposium on Data, Privacy, and E-Commerce, ISDPE 2007*, Chengdu, China, Nov. 2007, pp. 175–180.

[5] F. Liu, W. Ji, L. Hu, J. Ding, S. Lv, A. Pyshkin, and R.-P. Weinmann, “Analysis of the SMS4 block cipher,” in *Information Security and Privacy, 12th Australasian Conference, ACISP 2007, Proceedings, LNCS 4586*, Townsville, Australia, Jul. 2007, pp. 158–170.

[6] Y. Jin, H. Shen, and R. You, Implementation of SMS4 block cipher on FPGA,” in *Proceedings of 1st International Conference on Communications and Networking in China, ChinaCom’06*, Beijing, China, Oct. 2006, pp. 1–4.

[7] S. Morioka and A. Satoh, “A 10-Gbps full-AES crypto design with a twisted BDD S-box architecture,” *IEEE Trans. VLSI Syst.*, vol. 12, no. 7, pp. 686–691, Jul. 2004.

[8] U. Mayer, C. Oelsner, and T. Köhler, “Evaluation of different Rijndael implementations for high end servers,” in *Proceedings of 2002 IEEE International Symposium on Circuits and Systems, ISCAS 2002*, vol. 2, Scottsdale, AZ, USA, May 2002, pp. 348–351.

[9] F. Somenzi, “CUDD: CU decision diagram package release 2.4.1,” 2005. [Online]. Available: http://vlsi.colorado.edu/~fabio/CUDD

Dynamic Circuits For CMOS And Bicmos Low Power VLSI Design

Basavaraj.M.K,Ajith.G.Paranjpeand Pradeep.M.

basavaraj.basumk@gmail.com ,pradeepm032@gmail.com

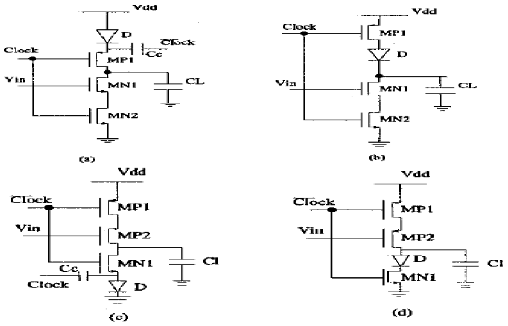
ABSTRACT

We present new types of circuits for dynamic logic gates in CMOS and BiCMOS technologies which use a diode to control precharge and predischarge to decrease the output voltage swing. This technique improves both delay and power dissipation compared to conventional dynamic CMOS and BiCMOS logic gates. Analysis indicates an improvement in delay and power of 50% and 26%, respectively for dynamic CMOS and 10% and 26%, respectively for dynamic BiCMOS using Vdd = 1.5V.

1. INTRODUCTION

The trend of VLSI digital circuit design is high speed and low power. The low power problem has been given considerable attention not only for saving power for battery applications but also for the thermal management of high performance systems. One way to approach the problem is to scale down device dimensions and reducethe supply voltage, thus decreasing power dissipation and maintaining switching speeds. However, there are economic and physical limitations to device scaling. Once the technology is chosen the requirements for minimum power and delay conflict, and one has to sacrifice power for delay or vice versa. Circuit and system designers can use optimization methods to minimize the delay-power product, but the result is not significant. This is true for both CMOS and BiCMOS circuits. We present a new dynamic circuit approach to reduce both delay and power. The idea is based on, in general, the reduction of the logic voltage swing which can reduce the change in charge of the load capacitance therefore decreasing the dynamic power dissipation. In addition, the switching time is decreased since the signals do not switch from rail to rail and the transistors are driven at full strength (i.e., maximum gate to source voltage). This idea can be implemented in dynamic gates since in the dynamic gate the switching occurs in one direction, either from high to low in precharge circuits or from low to high in predischarge circuits. The new dynamic CMOS gate is presented in the next section and the new dynamic BiCMOS gate is presented in Section 3. Simulation results and comparisons to conventional designs are presented n Section 4. Tradeoffs in diode placement within the CMOS gate are analyzed in Section ***5,*** and the paper is concluded in Section 6. **All** of the simulations presented in this paper are performed with HSPICE using MOSIS 1.Opm device models.

Figure 1. New dynamic CMOS logic gate: (a) and (b) precharge gates, and (c) and (d) predischarge gates.



The new dynamic CMOS logic gate is shown in Fig. 1. There are two possible circuit configurations depending on where the diode is placed: **(a)** the diode is placed above the precharge transistor MP1 (Fig. l(a)), and (b) the diode is placed below MP1 (Fig. l(b)). These two circuits have different characteristics which are presented below. The analysis in this paper refers to the precharge circuit, and the analysis of the predischarge circuit is similar. In both of the circuits in Figs. l(a) and (b) the diode limits the precharge voltage of the load capacitance to ***Vdd*** - &, where & is the diode forward voltage where the diode current is severely constrained. phase of the circuit in Fig. l(a) (clock is low) the load capacitance is charged by a current flowing through the diode and MPl. In the conventional CMOS gate the current is substantial until the load voltage is near ***Vdd;*** however, in our circuit the current is constrained by the diode when the load voltage reaches a level of ***Vdd*** - r/i. The current is not zero and will charge the load all the way to ***Vdd*** if left for a very long time, but the current is very small and can be neglected for even low speed (1MHz) applications. In the conventional dynamic CMOS precharge gate the charge injected by the clock onto the gate to source capacitance *C,,* causes the output voltage to be slightly large than ***Vdd.*** In the the circuit of Fig. l(a) however, there is charge injected on both the gate to drain capacitance ***Cgd*** and gate to source capacitance ***Cga.*** This causes the voltage overshoot to be larger than the conventional circuit. The additional charge injection can be compensated using an additional capacitor *C,* as shown in Fig. l(a). In general MN1 in this circuit can be any logic network. *V„* is the input from the previous gate which we assume has the dual structure (Fig. l(c)). The logic network must be connected in the same way as the NP domino logic [l]. *V„* predischarges to x above ground. Thus, *V„* is switched from to ***Vdd.*** In the precharge stage the clock is low, ***vo,t*** is charged to ***Vdd*** - V; and *V„* is predischarged to K by the previous gate. When the clock is high it goes into the evaluation stage and MN1 waits for *V„* to switch from V; to ***V&.*** MN1 operates in the subthreshold region if *V„* stays at % (assuming the threshold voltage is about **0.7** to 0.8 volts which is reasonable in most MOS processes); therefore, there is only a small subthreshold current which slowly discharges the load capacitance. Since the circuit is dynamic the node is refreshed every clock cycle and the subthreshold current can be neglected. When *V„* switches from low (%) to high ***(Vdd)*** MN1 is on and the load capacitance is discharged from ***Vdd*** - x to ground.

3. DYNAMIC BICMOS GATE

When the power supply voltage is scaled down in the BiCMOS digital design for low power applications the speed degradation is severe since the logic voltage swing is not rail to rail. Recently circuits have been presented to solve this problem [2, **3, 41** and their progress is significant. Given in [5] is a circuit for low voltage dynamic BiCMOS. The main feature of this circuit is the use of an NPN bipolar transistor in the precharge circuit to make the output full swing. Our new dynamic BiCMOS design uses an NPN bipolar transistor in the precharge circuit to improve the switching speed. We use a different control mechanism to cut the DC current which makes our circuit simpler than the previous circuit. In addition, we use a diode to limit the precharge voltage as in our CMOS case to boost both speed and power. The design of the dynamic BiCMOS gate is shown in Fig. 2. This is a PMOS driven bipolar pull down circuit which implements non-inverting logic like the design in **Figure 2. New dynamic BiCMOS ogic gate.** 151. We have a different controlling circuit to omplete the same goal as theirs, and in addition, we save two transistors which is better for both delay and power. Consider an input transition from high to low for the circuit in Fig. 2. When in the precharge stage clock (assume the clock signal is rail to rail) is high and MNl is turned on. In the same time MP1 is turned off whether ***Vout***i s high *(V&-* Vi) or low ***(0.lV).***W hen ***Vout***is high MN2 is on, MP3 is off and the high clock passes through making the gate of MP1 high ***(Vdd).*** When ***Vout*** is low MN2 is off and MP3 is on driving the gate of MP1 high. ***vb*** is pulled down to ground and forces the BJT Q into the cut-off region. The precharge PMOS MP4 is turnedon in the same time (since ***clock*** is low) and charges theload capacitance *CL* from ***0.1V*** (saturated voltage of Q) to ***Vdd*** minus the voltage drop of the diode. Becauseall other gates in the system have the same structure as this, the gate of MP2 driven by *V„* (which represents the logic PMOS network) will be precharged to ***V&*** -V;. When the clock goes low, since ***Voutl*** is high, ***V,*** switches low and forces MP1 on. In addition MN1 and MP4 are off therefore, the circuit is in the evaluation stage and waiting for the input signal *V„* to switch. Note that if *V„* remains at ***Vdd*** - K, MP2 is in the subthreshold region. Its subthreshold current is able to turn on Q after approximately 5011s. Thus, care must be used in specifying the minimum clock rate of the circuit. If *V„* switches from ***Vdd*** - % to 0.1, there is a charging path from ***Vdd,*** MP1, MP2 to ***Q. vb*** switches high and forces Q into the forward active region. The collector current of ***Q*** discharges the load capacitance from ***Vdd-K*** to 0.1 which is the saturated voltage of ***Q.*** MP3 and MN2 are thereafter used to cut off the static current from ***Vdd,*** MP1, MP2, ***Q*** to ground. Since the low ***Kutl*** forces MN2 off and ***V,*** is isolated from clock, the low ***V„t*** forces MP3 on and ***v,*** is pulled up to ***Vdd*** which forces MP1 off.

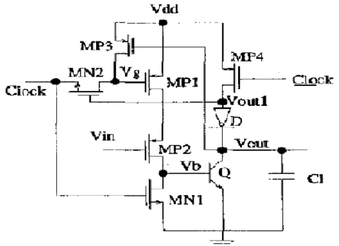


Figure 2. New dynamic BiCMOS logic gate.

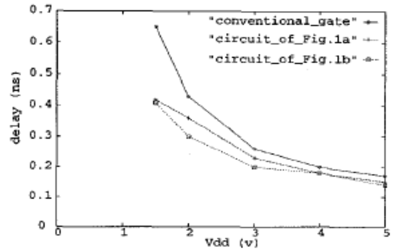


Figure 3. **Delay versus V_{dd} for our dynamic gate and the conventional dynamic gate with**

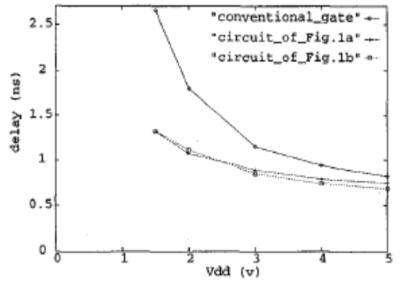
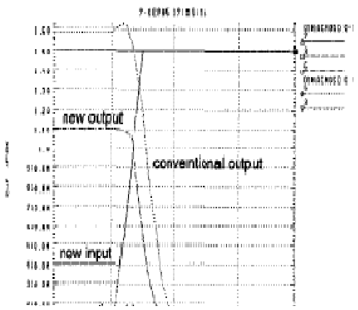


Figure 4. **Delay versus V_{dd} for our dynamic gate and the conventional dynamic gate with $C_l = 500ff$.**



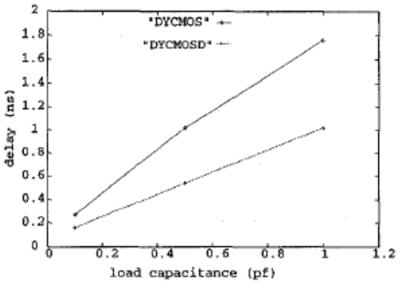


Figure 6. Delay versus load capacitance for our dynamic gate (DYCMOSD) and the conventional dynamic gate (DYCMOS) with $V_{dd} = 1.5V$.

4. SIMULATION AND PERFORMANCE COMPARISON

4.1. Delay

The delay reduction over the conventional gate comes from the small output swing. In the case of the precharge CMOS circuit we can make a delay reduction estimation. Assuming the output falls linearly the delay D in the conventional case is the time it takes for the output to fall from *Vdd* to *V d d / 2* . If the output is from *v d d* - K to 0 and we assume the output f d s linearly and with the same rate as does the conventional circuit, the delay of the new circuit is *o (l - 2 K / v d d)* . Based on this
estimation the percent delay reduction is *2%* **x** *100.* If we use *Vdd* = 1.5V and = 0.4, the delay reduction is about 50%. Shown in Figs. **3** and **4** are the SPICE simulation results of delay vs. *V d d* for two load capacitances (Ci=lOOff and 9=500ff). It can be seen that when ***Ci*** is large the simulation results are similar to the above estimation, but when *Cr* is small the improvement is less than the estimation because the estimation ignores the intrinsic delay of the gates. Waveforms from SPICE simulations of the two circuits are shown in Fig. 5. The delay comparison between our circuit and the conventional circuit for different load capacitances is shown in Fig. 6. In the precharge BiCMOS circuit the output falling delay is composed of two parts of which the first part is the time to charge the base and diffusion capacitance to switch the bipolar device on, which is independent of the load capacitance. The second part is the time to discharge the load capacitor. Our precharge-limit circuits do not affect the first part of the delay. The second part of the delay is reduced in the same way as the CMOS case by the reduction of the output swing. The simula-tion results of the two circuits are shown in Fig. **7.** The delay reduction is approximately 10% for typical load capacitances. The delay comparison between our circuit and the circuit in *[5]* for different load capacitances is shown in Fig. 8.

4.2. Power Dissipation

The dynamic power dissipation of a gate is proportional to *Vdd* times the voltage swing of the output capacitance *(v i d* in conventional CMOS). Therefore the power savings against the conventional case is *x / T / d d* . If Vi is **0.4V** and *Vdd* is *1.5V,* then the savings is **26%.** However we
have to consider the static power in this case since we assume the input(s) to the gate are not rail to rail. For the precharge gate the input(s) may have the low logic value VI which forces the transistor(s) to operate in the subthreshold region and there is a subthreshold leakage current flow to ground. If this leakage current is large enough to discharge the load capacitance significantly in a clock period we have to put a PFET in the circuit to compensate for the leakage, and the static power dissipation in this case is comparable with the dynamic power savings. We found that a gate driving about the same size gate(s) does not encounter significant output voltage drop. For example, when the input voltage is about **0.4V** the leakage current is on the order of several hundreds of nA for an NFET of size 16X1um2. This current takes 16 microseconds to decrease the output voltage by 0.4V when the load capacitance is on the order of 20-3Off. This means in the normal case ,the static power caused by the non-full swing is not significant if the .transition rate is not too slow. In the above example, if we assume the clock period is 16 nanoseconds the dynamic power reduction in our circuit is greater than the increase in static power for switching probabilities larger than 0.1%. Thus, the savings in dynamic power is dominant.

*5.*DIODE PLACEMENT IN THE MOS CIRCUIT

The placement of the diode(Fig. l(a) or (b)) is chosen based on diode realization and the supply voltage. There are two possible realizations of the diode in the CMOS process. Diffusion and floating well can be used to form a p-n junction, but the smallest well to substrate capacitance is comparable to typical gate capacitance which increases *Ci* of the circuit in Fig. l(b) and **degrades** the **performance.** In **this *case* it is** better to **use** the circuit in Fig. l(a) because MP1 isolates the diode capacitance from the load. The circuit in Fig. l(a) has unreasonable precharge time when Vdd is scaled down to less than 2V (Fig. 9). The increased precharge time is due to the body effect and the reduced V,, which is caused by the voltage drop across the diode. Therefore, the circuit in Fig. l(a) can only be used in the applications which have the Vdd not less than 2v. Another realization of the diode is to overlap the P and N diffusion in the CMOS process [SI; therefore the parasitic capacitance is small compared to gate capacitance. In this case the diode is below the precharge transistor MP1, and the precharge time is comparable to the conventional gate even when Vdd is scaled down to 1.5V (Fig. 9).

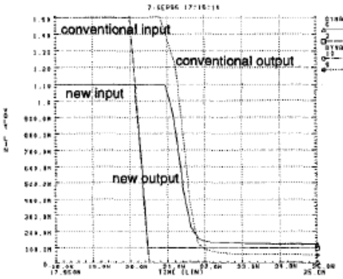


Figure 7. Waveforms obtained from SPICE simulations of dynamic BiCMOS gates.

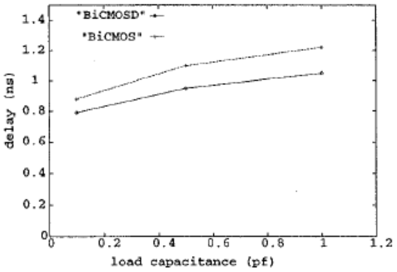


Figure 8. Delay versus load capacitance for our circuit (BiCMOSD) and the circuit in [5] (BiCMOS) with $V_{dd} = 1.5V$.

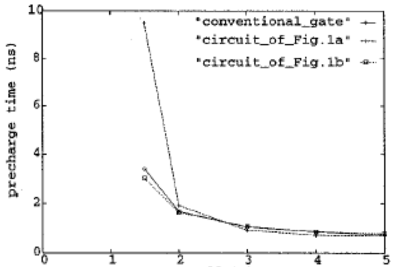


Figure 9. Precharge time versus V_{dd} for our dynamic gate and the conventional dynamic gate.

6. CONCLUSION AND FUTURE WORK

We have proposed new dynamic CMOS and BiCMOS gates for low voltage supply applications which can significantly improve both speed and power using the diode to limit the voltage swing. The improvement in delay (power) is as much as 50% (26%) for CMOS and 10% (26%) for BiCMOS with Vdd = 1.5V. More investigations are needed to verify these results experimentally and explore CAD tools to simplify their design.

REFERENCES

Neil H. E. Weste and K. Eshraghian, *Principles* of *CMOS* VLSI *Design: A Systems Perspective*, Addison-Wesley, 2nd ed., 1993.

M. Hiraki, et al., "A 1.5-V Full-Swing BiCMOS Logic Circuit," IEEE J. Solid-state Circ., vol. 27, pp. 1568-1574, Nov. 1992.

R.L. Geiger, P.E. Allen and N.R. Strader, VLSI *Design Techniques* for *Analog and Digital Circuits*, McGraw Hill, 1990.

Biodata:

Basavaraj.M.K.
7th sem ECE
KIT, Tiptur.
9743393101

Pradeep.M.
7th sem ECE
KIT, Tiptur.
9538125136

Ajith.G.Paranjith
7th sem ECE
KIT, Tiptur.
9743179458

H.264/AVC CABAC DECODER DESIGN USING VHDL

Ch. Anil kumar[1], G.Babu[2],

[1, 2] ECE Dept. Vaagdevi College of Engineering, Jawaharlal Nehru Technological University, Hyderabad.

anil.chidra@gmail.com, +91-8801808600[1], babugundlapally@gmail.com, +91-9652935193[2],

Abstract: The H.264/AVC is the most recent standard of video compression/decompression for future broadband network. This standard was developed through the Joint Video Team (JVT) from the ITU-T Video Coding Experts Group and the ISO/IEC MPEG standardization committee. In this project H.264 decoder functional block such as Context based Binary arithmetic coding (CABAC) is designed using VHDL. CABAC includes three basic building blocks of context modeling, binary arithmetic coding and Inverse binarization. Here the compressed bit-stream from NAL unit is expanded by CABAC module to generate various syntax elements. Here the basic arithmetic decoding circuit units are designed to share efficiently by all syntax elements.

Keywords---*MPEG-2, H.264, MPEG-4 Part 10, AVC, digital video codec standard, Lossy compression, lossy transform codecs, lossy predictive codecs*

I. INTRODUCTION

H.264/AVC is the latest video compression standard developed by ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) for next-generation multimedia coding applications. H.264 adopts many brand new technologies such as variable block size motion estimation with multiple reference frames, de-blocking filtering, B-frame coding, context-adaptive entropy coding, picture adaptive frame field (PAFF) coding, macroblock (MB) adaptive frame field coding, and 8 × 8 transform in the newly developed H.264 high-profile (HP) specification for high definition television (HDTV) applications.

Compared to the previous MPEG standards, H.264 provides over two times higher compression ratio under the same video coding quality. However, the computational complexity of H.264 video coding is much higher than those of the previous MPEG standards.

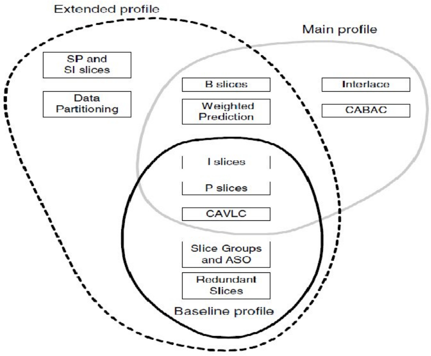


Fig1: Baseline, Main and Extended profiles

There are two techniques adopted in H.264 entropy coding. One is context-based adaptive variable length coding (CAVLC) for baseline profile (BP). The other is context-based adaptive binary arithmetic coding (CABAC) for main profile (MP) and HP. Compared to the CAVLC, adopting CABAC can save about 9–14% bit rates at the cost of much higher computational complexity.

The further paper is organized as follows: the next section describes the problem statement. Section III, explains about cabac decoder and the different parts of CABAC Decoder. Section IV, provides simulation and synthesis results of CABAC Decoder. Section V, we conclude.

II. PROBLEM STATEMENT

Nowadays, a large number of consumer products such as digital cameras, Personal Digital Assistants, video telephony, portable DVD player as well as storage, broadcast and streaming of standard definition TV are common practice. All those applications demand efficient management of the large amount of video data. This motivated a large body of research in industry as well as in academia to develop advanced video coding technology. H.264/AVC video coding standard developed by the ITU-T/ISO/IEC is the latest international standard developed by ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group.

The new standard provides gains in compression efficiency of up to 50% over a wide range of bit rates and video resolutions compared with the former standards. Averagely 30~40 cycles are needed to decode a single bin on DSP. That means for such typical 4M bit stream, averagely about 1.5× 100× 30=4500 cycles are needed simply to implement the arithmetic decoding task for one MB, while the cycles for other controls are not counted in. This speed is unacceptable for real time applications, where 30 frames of D1 resolution are required to be decoded within 1s with 100MHz clock, i.e. a MB has to be decoded within at most 2000 cycles. So hardware acceleration is necessary for a commercially viable H.264/AVC based video application, especially with increase in image size and quality settings in the future. To speed up the decoding process, multiplication-free logic is used to calculate subinterval range.

III OVERVIEW OF CABAC DECODER

At first, it needs to re-initialize both the context tables in the beginning of each slice and probability model from bit-streams. Then, the syntax elements (SEs) of each MB could be decoded one by one. The “Decode Core” in Fig. 3 consists of three arithmetic decoding processes defined in H.264 video coding standard. According to our observation, the behaviors between *DecodeDecision* and *DecodeTerminate* are almost the same except that *DecodeTerminate* applies the fixed context index 276 that will only be used without being updated. Therefore, we refer the combination of *DecodeDecision* and *DecodeTerminate* to the decision decoding engine and use a bypass decoding engine to stand for *DecodeBypass* in the rest of this letter. Only one of them works at a time to generate a “bin”

value. The “bin” is the basic decoding unit of the SEs in CABAC. After a bin is decoded, the binarization stage checks if the successive decoded bin is in bin string. If it is not, the decoder will keep on decoding the next bin. Otherwise, it will prepare for decoding the first bin of the next SE.

The MBAFF coding tool, which is provided in H.264 MP/HP, provides good coding efficiency when a scene consists of both stationary and significant motion regions The MBAFF coding tool was reported to reduce 28–33% of bit rate than the frame-based coding. In decision decoding engine, it is required to refer syntax information from the left MB and top MB to select the context index. To support the MBAFF coding tool, it becomes more complicated to obtain the essential syntax information from the neighboring MBs. It has to refer to the top or the bottom MB in the neighboring MB pairs according to the related MBs coded in frame or field mode, which greatly increases the complexity for hardware realization.

For getting better compression efficiency for HD videos, an 8 × 8 transform is adopted in H.264 HP. That is, the CABAC decoding architecture should be able to support both the 4 × 4 and 8 × 8 transform blocks, which complicates the hardware complexity in increasing both the hardware cost for 8 × 8 blocks and the decoding latency. Therefore, we have to consider the hardware sharing of CABAC decoders when operating on 4×4 and 8×8 blocks for reducing the hardware cost.

Motion Estimation

Motion estimation of a macroblock involves finding a 16 × 16-sample region in a reference frame that closely matches the current macroblock. The reference frame is a previously encoded frame from the sequence and may be before or after the current frame in display order. An area in the reference frame centred on the current macroblock position (the search area) is searched and the 16 × 16 region within the search area that minimizes a matching criterion is chosen as the 'best match'.

Motion Compensation

The selected 'best' matching region in the reference frame is subtracted from the current macroblock to produce a residual macroblock (luminance and chrominance) that is encoded and transmitted together with a motion vector describing the position of the best matching region (relative to the current macroblock position).Within the encoder, the residual is encoded and decoded and added to the matching region to form a reconstructed macroblock which is stored as a reference for further motion-compensated prediction. It is necessary to use a decoded residual to reconstruct the macroblock in order to ensure that encoder and decoder use an identical reference frame for motion compensation.

Transform Coding

The purpose of the transform stage in an image or video CODEC is to convert image or motion-compensated residual data into another domain (the transform domain). The choice of transform depends on a number of criteria:

1. Data in the transform domain should be decorrelated (separated into components with minimal inter-dependence) and compact (most of the energy in the transformed data should be concentrated into a small number of values).
2. The transform should be reversible.
3. The transform should be computationally tractable (low memory requirement, achievable using limited-precision arithmetic, low number of arithmetic operations, etc.).

Quantisation

A quantiser maps a signal with a range of values *X* to a quantised signal with a reduced range of values *Y*. It should be possible to represent the quantised signal with fewer bits than the original since the range of possible values is smaller. A *scalar quantiser* maps one sample of the input signal to one quantised output value and a *vector quantiser* maps a group of input samples (a 'vector') to a group of quantised
values.

Arithmetic Coding

The variable length coding schemes share the fundamental disadvantage that assigning a codeword containing an integral number of bits to each symbol is sub-optimal, since the optimal number of bits for a symbol depends on the information content and is usually a fractional number. Compression efficiency of variable length codes is particularly poor for symbols with probabilities greater than 0.5 as the best that can be achieved is to represent these symbols with a single-bit code. Arithmetic coding provides a practical alternative to Huffman coding that can more closely approach theoretical maximum compression ratios. An arithmetic encoder converts a sequence of data symbols into a single fractional number and can approach the optimal fractional number of bits required to represent each symbol.

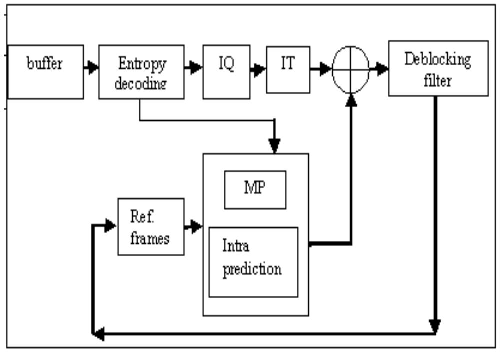


Fig2. Block diagram of CABAC Decoder

Flow chart for cabac decoder

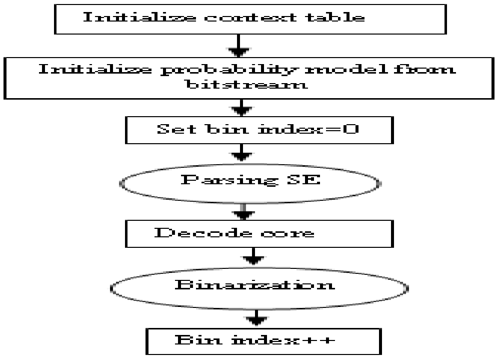


Fig3: decoding flow of cabac decoder

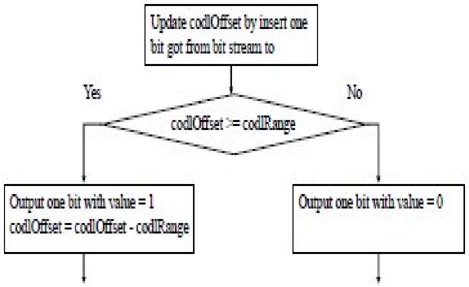

Update codIOffset by insert one bit got from bit stream to
Yes
No
codIOffset >= codIRange
Output one bit with value = 1
codIOffset = codIOffset - codIRange
Output one bit with value = 0

Fig4. decode bypass process

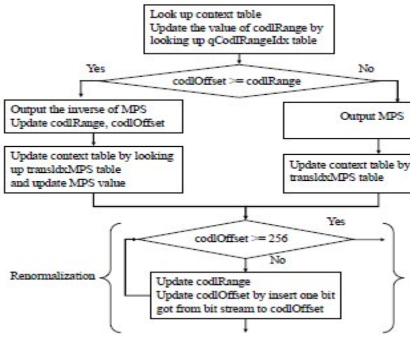

Look up context table
Update the value of codlRange by looking up qCodlRangeIdx table
Yes
No
codlOffset >= codlRange
Output the inverse of MPS
Update codlRange, codlOffset
Output MPS
Update context table by looking up transldxMPS table and update MPS value
Update context table by looking up transldxMPS table
Yes
codlOffset >= 256
No
Renormalization
Update codlRange
Update codlOffset by insert one bit got from bit stream to codlOffset

Fig5. Normal Decode process

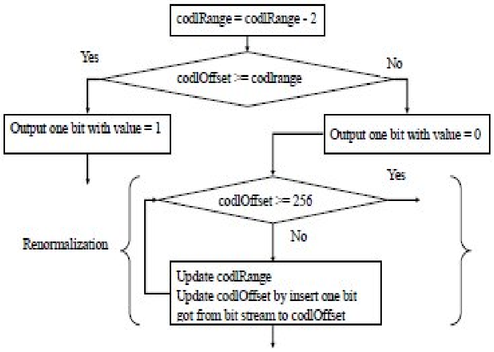
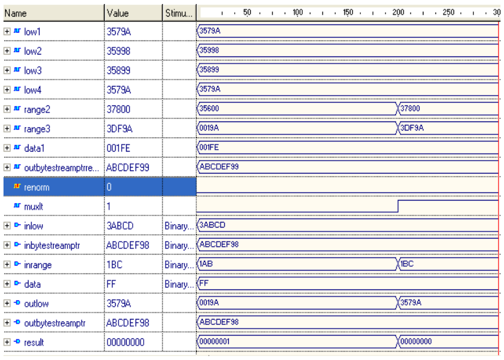

codIRange = codIRange - 2
Yes
No
codIOffset >= codIrange
Output one bit with value = 1
Output one bit with value = 0
Yes
codIOffset >= 256
No
Renormalization
Update codIRange
Update codIOffset by insert one bit got from bit stream to codIOffset

Fig6. Terminal decoding process

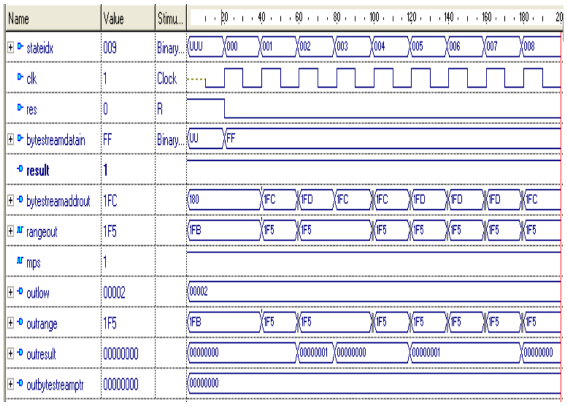
IV Simulation and Synthesis Results

Simuation Results:



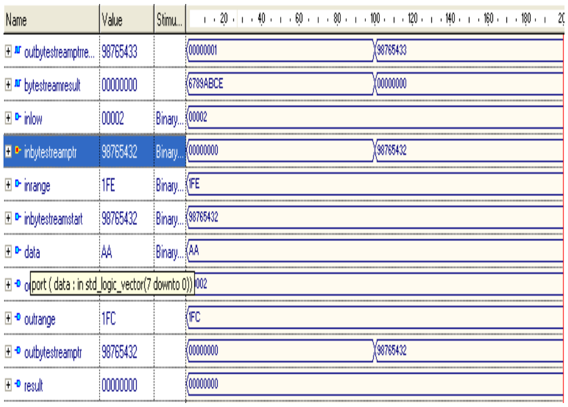
Name	Value	Stimu...
low1	3579A	
low2	35998	
low3	35899	
low4	3579A	
range2	37800	
range3	3DF9A	
data1	001FE	
outbytestreamptrre...	ABCDEF99	
renorm	0	
muxlt	1	
inlow	3ABCD	Binary...
inbytestreamptr	ABCDEF98	Binary...
inrange	1BC	Binary...
data	FF	Binary...
outlow	3579A	
outbytestreamptr	ABCDEF98	
result	00000000	

Fig7. Bypass decode



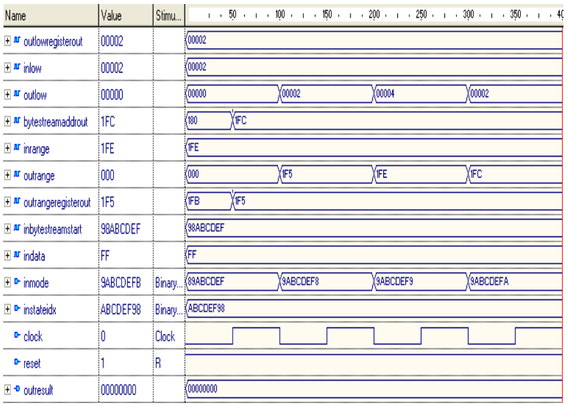
Name	Value	Stimu...
stateidx	009	Binary...
clk	1	Clock
res	0	R
bytestreamdatain	FF	Binary...
result	1	
bytestreamaddrout	1FC	
rangeout	1F5	
mps	1	
outlow	00002	
outrange	1F5	
outresult	00000000	
outbytestreamptr	00000000	

Fig8. Getcabac decoding



Name	Value	Stimu...
outbytestreamptrre...	98765433	
bytestreamresult	00000000	
inlow	00002	Binary...
inbytestreamptr	98765432	Binary...
inrange	1FE	Binary...
inbytestreamstart	98765432	Binary...
data	AA	Binary...
o port (data : in std_logic_vector(7 downto 0))		
outrange	1FC	
outbytestreamptr	98765432	
result	00000000	

Fig9. Terminate decode



Name	Value	Stimu...
outlowregisterout	00002	
inlow	00002	
outlow	00000	
bytestreamaddrout	1FC	
inrange	1FE	
outrange	000	
outrangeregisterout	1F5	
inbytestreamstart	98ABCDEF	
indata	FF	
inmode	9ABCDEFB	Binary...
instateidx	ABCDEF98	Binary...
clock	0	Clock
reset	1	R
outresult	00000000	

Fig10**.** Cabac decoder

Synthesis Results:

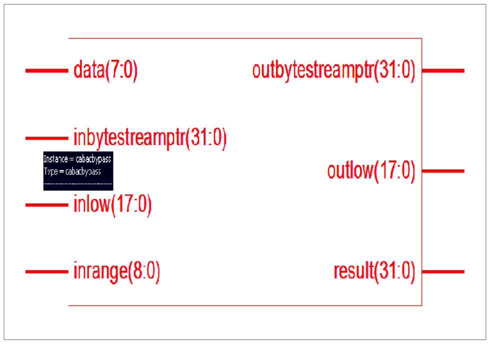


Fig11 Bypass Decode

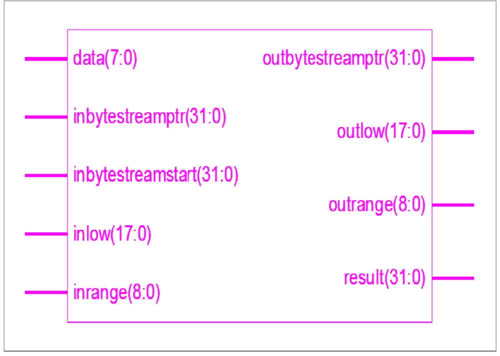


Fig12. Termiate decode

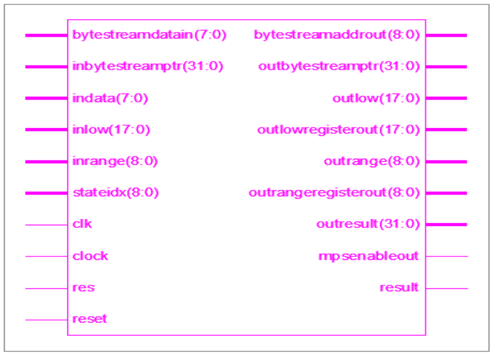


Fig13. Getcabac

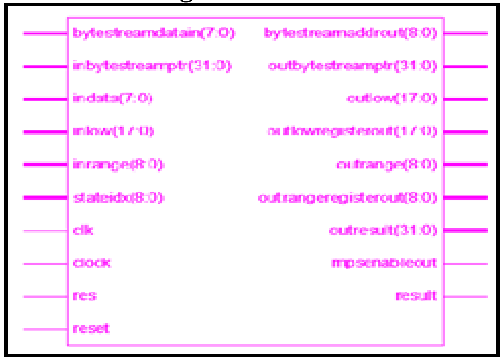


Fig14.Cabac decoder

V.CONCLUSION

In this project H.264 decoder functional blocks such as Context based Binary arithmetic coding (CABAC), Inverse Quantization and Inverse Discrete Cosine Transform are designed using VHDL to increase the speed of decoding operation. Since CABAC decoding is a highly time consuming process, CPU or DSP is not being the appropriate choice for real-time CABAC decoding applications. This project work shows that the hardware design of CABAC Decoder is possible for a commercially viable H.264/AVC based video application, especially with increase in image size and quality settings in the future.

Future Developments

In this project work, CABAC decoder is designed using VHDL to increase the speed of decoding operation. Since CABAC is a key technology adopted in H.264/AVC standard, it offers a 16% bit-rate reduction when compared to baseline entropy coder while increasing access frequency from 25% to 30%.So CABAC decoding is a highly time consuming process. Multiple decoding engines and shared memory between the modules can be implemented in future to increase the decoding speed especially to suite for high bit rate applications such as HDTV, High Definition DVD, Broadcast and Streaming, Digital Television. So Much space is left for real-time applications of higher video quality and larger image resolutions in the future.

REFERENCES:

[1] Joint Video Team (JVT) of ISO/IEC MPEG&ITU-T VCEG, ISO/IEC 14496-10, 2003.
[2] ITU-T Recommendations for H.264 Conformance
[3] Bitstream Files, Available:
http://ftp3.itu.ch/avarch/jvtsite/ draft_conformance/
Joint Video Team (JVT) Reference Software JM10.2.
[4] I. Richardson, “Video CODEC Design”, John Wiley & Sons, 2002.
[5]. D. Marpe, G Blättermann and T Wiegand, “Adaptive Codes for H.26L”, ITU-T SG16/6 document VCEG-L13, Eibsee, Germany, January 2001
[6] vcodex.com/h.264mal
[7] http://www.ittiam.com/pages/products/h264-dec.htm

CAD TOOL DEVELOPMENT FOR MAXIMUM SPEED ESTIMATION AND ITS ACHIEVEMENT USING UNFOLDING

Sumeera Afreen,
M.Tech Student, Dept. of Electronics and Communication,
Sir M.VIT Bangalore-562157, India.
afreensumera@gmail.com

***Abstract*— This Paper presents determination of maximum speed of execution (iteration bound) of recursive DSP algorithms by using two algorithms Longest Path Matrix (LPM) and Minimum Cycle Mean (MCM) algorithm. Comparison of performance analysis is done between LPM & MCM. Cases are considered where the iteration bound cannot be attained without unfolding and an algorithm is used to solve the problem. An algorithm is used to convert Multirate circuit to single-rate circuit and results are compared with theoretical values.**

Exclusively to help the designers to design efficient DSP systems onto the chip, a set of programs are written and executed successfully using MATLAB software for LPM and MCM ot find iteration bound of recursive circuits, to unfold the circuit and to convert Multirate circuit to single-rate circuit. These set of programs form a tool called as VLSI CAD Tool which helps the designers to quickly find iteration bound of recursive circuits and efficiently design DSP systems on chip.

***Keywords*-LPM, MCM, Units of Time (u.t.), DFG, DSP, Unfolding, SRDFG, MRDFG.**

I. INTRODUCTION

The field of DSP has always been driven by the advances in DSP applications and in scaled VLSI technologies. Therefore at any given time DSP applications impose several challenges on the implementation of DSP system onto the chip in terms of speed of execution, amount of hardware circuitry and resources used and amount of power consumption so as to maximize the performance and to meet the specifications of the customer.

DSP systems are designed using DSP algorithms. Because of the recursive nature of DSP algorithms there exists an inherent fundamental lower bound on the ***iteration period*** of the algorithms referred to as the ***iteration bound***. This bound is fundamental to an algorithm and is independent of the implementation architecture. In other words it is impossible to achieve an iteration period less than the iteration bound even when infinite processors are available to execute the recursive algorithm. Determination of iteration bound of the recursive circuits is an important problem because it discourages the designer to attempt to design architecture with an iteration period less than the iteration bound. Many DSP Algorithms such as recursive and adaptive digital Filters contain feedback loops, which impose an inherent fundamental lower bound on the achievable iteration or sample period. This bound is referred to as the iteration period bound, or simply the iteration bound. The iteration bound is a characteristic of the representation of an algorithm in the form of a data flow graph (DFG).

A DFG can be classified as nonrecursive or recursive. A nonrecursive DFG contains no loops, while a recursive DFG contains atleast one loop. For example an FIR filter is non recursive, while the DFG in 1 is recursive because it contains the loop. A recursive DFG has a fundamental limit on how fast the underlying DSP program can be implemented in hardware. This limit is called iteration bound.

Given that the execution times of nodes A and B are 2 and 4 u.t., respectively one iteration of the loop requires 6 u.t. This is the loop bound, which represents the lower bound on the loop computation time. Formally, the loop bound of loop is defined as , where is the loop computation time and is the number of delays in the loop. For example the loop bound for loop in is $6/11 = 6$ . .

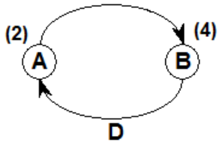


Fig 1

The DFG in contains two loops, namely, the loops = → → and = → → The loop bounds for and are $6/2 = 3$. and $11/1 = 11$. respectively. The critical loop is the loop with the maximum loop bound, so the loop is the critical loop in fig 2. The loop bound of the critical loop is the iteration bound of the DSP program, which is the lower bound on the iteration or sample period of the DSP program regardless of the amount of computing resources available.

Formally, the iteration bound is defined as

$$= \max_{\in} \frac{}{},$$

Where is set of loops in the DFG, is the computation time of the loop , and is the number of delays in the loop . For the DFG in the Fig we have

$$_{\infty} = \max \frac{6}{2}, \frac{11}{1} = 11 \quad . \quad .$$

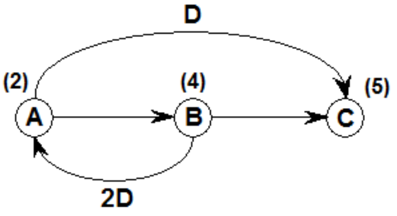


Fig 2

II. ALGORITHMS FOR COMPUTING ITERATION BOUND

A. Longest Path Matrix Algorithm

In Longest Path Matrix (LPM) algorithm, a series of matrices is constructed, and the iteration bound is found by examining the diagonal elements of the matrices. Let d be the number of delays in DFG. These matrices, ( ), $= 1, 2, \ldots\ldots,$ are constructed such that the value of the element ( ) is the longest computation time of all paths from delay element to delay element , that passes through exactly ($- 1$) delays (not including and). If no such path exists, then ( ) $= -1$. The longest path between any two nodes can be computed using any path algorithm.

For example, to determine ( ) for DFG in Fig. all paths from delay element to delay element that passes through exactly zero delay elements must be considered. There is one such path, namely, the path → → → → → → .

This path has computation time 5, so ( ) $= 5$. To determine ( ) we note that there are no paths from the delay element d₄ to delay element d₃ the passes through zero delay elements, so ( ) $= -1$. After determining th rest of the elements of ( ) we find

$$( ) = \begin{matrix} - & & - & - \\ & - & & - \\ & - & - & \\ & - & - & - \end{matrix}$$

The higher order matrices, ( ) $= 1, 2, \ldots\ldots,$, do not need to be determined directly from the DFG. Rather, they can be recursively computed according to the rule

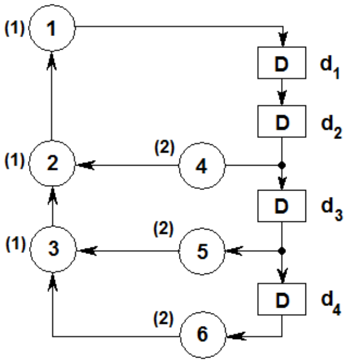


Fig 3

$$( ) = \max_{\in}(-1, ( ) + ( ))$$

Where is the set of integers in the interval [1,] such that neither ( ) $= -1$ nor ( ) $= -1$ holds. For example to compute ( ) first step is to find the set from possible set $\{1,2,3,4\}$. The value 3 is in because ( ) $= 0$ and ( ) $= 5$, and the values $= 1,2,4$ are not in because atleast one of ( ) or ( ) is equal to -1 for each of these. Using $= \{3\}$, the value of ( ) can be computed as

$$( ) = \max_{\in\{\ \}}(-1, ( ) + ( ))$$

$$= \max(-1, 0+5) = 5.$$

Computing the remaining values of ( ) results in

$$( ) = \begin{matrix} & - & & - \\ & & - & \\ & - & - & - \\ - & & - & - \end{matrix}$$

While ( ) is computed using only (,) the matrix ( ) is computed using both ( ) and ( ). To compute (,) = {1} because (,) = 5 and (,) = 0 and for the values of = 2,3,4, at least one of (,) is equal to (−1). The value of (,) is

$$(,) = \max_{\in\{ \}}(-1, (,) + (,))$$

$$= \max(-1 , 5+0) =5.$$

Computing the rest of ( ) **and** ( ) results in

$$( ) = \begin{matrix} & - & \\ & & - \\ & & - \\ - & & - \end{matrix}$$

and

$$( ) = \begin{matrix} - \\ \\ \\ - \end{matrix}$$

Once the matrices ( ) have been computed, the iteration bound can be determined as

$$_\infty = \max_{, \in\{ , ,\dots\}} \{ —', \}$$

Which for this example is

$$_\infty = \max \frac{4}{2},\frac{4}{2},\frac{5}{3},\frac{5}{3},\frac{5}{3},\frac{8}{4},\frac{8}{4},\frac{5}{4},\frac{5}{4} = 2 \quad . \quad .$$

B. Minimum Cycle Mean Algorithm

Construct a graph from the original DFG(G). If is the number of delay elements in , then the graph has nodes where each node corresponds to one of the delays in .The weight (,) of the edge from the node to the node in is the longest path length among all paths in from the delay to the delay that do not pass through any delays. If no zero-delay path exists from the delay to the delay , then the edge → does not exist in . The cycle mean of a cycle in is

$$\frac{\text{Maximum computation time of all cycles in G that contain the delays in c}}{\text{the number of delays in these cycles in G}}$$

To compute the maximum cycle mean of , the graph $\bar{}$ is constructed from by simply multiplying the weights of the edges by −1, . $\bar{}$., has the same topology as and the weights (,) of the edge → in the are given by ( ), = ( ), where (,) is the weight of edge → in . The graph for the DFG in is given in 2.15(e). The maximum cycle mean of is simply the MCM of $\bar{}$ and multiplying it by −1.

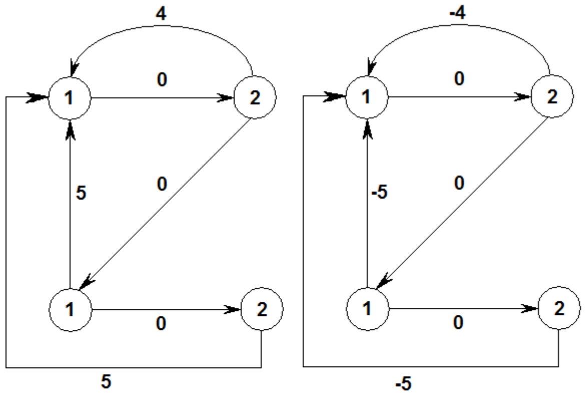


Fig 4 (a) Fig 4 (b)

$$( ) \qquad ( )\bar{}$$

The MCM of $\bar{}$ is found by first constructing the series of +1 vectors, ( ) = 0,1,2 ....., , which are each of dimension d X 1. An arbitrary reference node is chosen in (call this node . The initial vector ( ) is formed by setting ( )( ) = 0 and setting the remaining entries of ( ) to ∞. If node 1 is chosen as the reference node for the graph in 4(then)

$$( ) = \begin{matrix} 0 \\ \infty \\ \infty \\ \infty \end{matrix}$$

The remaining vectors, (,) = 1,2,....., are respectively computed according to

$$( )( ) = \ \ _{\in} (( )( ) + \bar{}(,))$$

where $\bar{}$(,) is the weight of the edge → in $\bar{}$ and is the set of nodes in $\bar{}$ such that there exists an edge from node to node (→)

This series of vectors found from $\bar{}$ in is 4( )

$(\quad) \neq \begin{matrix} \infty \\ 0 \\ \infty \\ \infty \end{matrix} \qquad (\quad) \neq \begin{matrix} -4 \\ \infty \\ 0 \\ \infty \end{matrix}$

$(\quad) \neq \begin{matrix} -5 \\ -4 \\ \infty \\ 0 \end{matrix} \qquad (\quad) \neq \begin{matrix} -8 \\ -5 \\ -4 \\ \infty \end{matrix}$

From the vectors $(\quad) = 0,1,2 \ldots\ldots,$ the iteration bound can be computed as

$$= - \min_{\{\in \; , \; ,\}\ldots,} \; \max_{\in\in\{ \; , \; , \; \ldots,} \frac{(\;)(\;) - (\;)(\quad)}{\}-}$$

Table 2.1 Values of $\frac{(\;)(\;) \quad (\;)(\quad)}{}$ $1 \leq$ 4 $0 \leq 3$

	= 0	= 1	= 2	= 3	$\frac{(\;)(\;) - (\;)(\;)}{-}$
= 1	−2	−∞	−2	−3	−2
= 2	−∞	−5 /3	−∞	−1	−1
= 3	−∞	−∞	−2	−∞	−2
= 4	∞ − ∞	∞ − ∞	∞ − ∞	∞	∞

$$= \quad (- \quad ,- \quad ,-) = ,\infty .$$

III. Iteration bound of Multirate Data Flow Graphs.

Another class of the DFGs, is called Multirate DFGs (MRDFGs), which allows to execute the node more than once per iteration, and two nodes are not required to execute the same number of times in an iteration. The step two process used to compute the iteration bound of Multirate DFG. The process is,

1. Construct a SRDFG(Single Rate DFG) that is equivalent to the MRDFG.
2. Compute the iteration bound of the equivalent SRDFG using the LPM algorithm, or the MCM algorithm.

The iteration bound of the MRDFG is same as the iteration bound of the equivalent SRDFG. Here MRDFGs are defined and an algorithm is presented for constructing equivalent SRDFG from an MRDFG. Then by applying LPM algorithm or MCM algorithm to SRDFG its iteration bound can be determined.

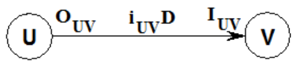


Fig 5

→

An edge from the node to node in an MRDFG is shown in the . The 5value is the number of ≤samples produced on the edge by an invocation of the node , and the value of is the number of samples consumed from the edge by an invocation of node . The value of is the number of delays on the edge.

If the nodes and are invoked times and times, respectively in an iteration, then the number of samples produced on the edge from the node to the node in one iteration. ,and then no of samples consumed from the edge by the node in one iteration is . To avoid a buildup or defiency of samples on the edge, The number of samples produced in 1 iteration must equal the number of samples consumed in one iteration. This relationship can be described mathematically as

$$=$$

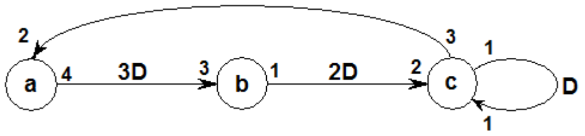


Fig 6

An Algorithm to construct SRDFG from MRDFG

1. For each node in the MRDFG
2. $= 0$
3. Draw a node in the SRDFG with the same computation time as in the MRDFG.
4. For each edge in the MRDFG.
5. $= 0 \; -$
6. Draw an edge $/ \quad \rightarrow \; ((\quad)/ \quad)\%$ in the SRDFG with $(\quad + / \quad)$ delays.

To determine how many times each node must be executed in iteration, the set of equations found by writing = for each edge in an MRDFG must be solved so the number of invocations of nodes are co-prime. For example, the set of equations for MRDFG in . 6

$$= 3$$

$$= 2$$

$$=$$

$$= 2$$

Which has a solution = 3, = 4, = 2. Once the number of invocations of the nodes has been determined, an equivalent SRDFG can be constructed for the MRDFG.

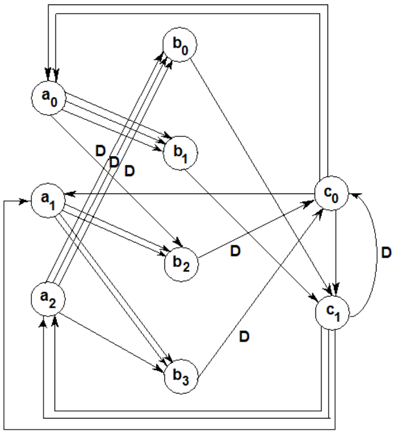


Fig 7

IV. UNFOLDING

Unfolding is a transformation technique that can be applied to a DSP program to create a new program describing more than one iteration of the original program. More specifically, unfolding a DSP program by unfolding factor J creates a new program that describes J consecutive iterations of the original program.

UNFOLDING ALGORITHM

1. For each node in the original DFG, draw the nodes , , … … .
2. For each edge → with delays in the original DFG, draw the edges → ( )% —h delays = 0, 1, 2 … . . , −. 1

A DFG is shown in In8(this D)FG, the nodes A and B represent input and output, respectively and the nodes C and D represent addition and multiplication by a, respectively. To unfold this DFG by unfolding factor 2, the 8 nodes , , , and , are=fi0r,s1t,drawn according to 1 step of unfolding algorithm.

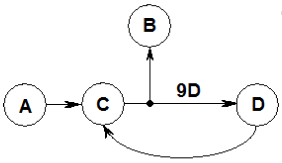


Fig 8 (a)

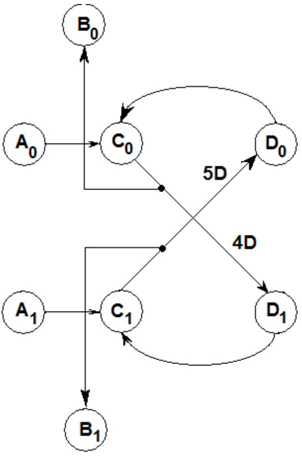


Fig 8 (b)

After these nodes have been drawn, the 2 step of unfolding algorithm needs to be performed. For an edge → with no delays, this step reduces to drawing the edges → with no delays = 0, 1, 2 … . . , −. 1For example, the edge → with no delays in re8s(ults)in the two edges → and → with no delays in the 2 − DFG in ( ). F8or the edge → with = 9 delays in ( ), w8e draw the edges → ( )% with — delays and → ( )% with — delays, which correspond to the edges → with 4 delays and → with 5 delays respectively, in ( ). 8

In some cases, the DSP program cannot be implemented with iteration period equal to the iteration bound without the use of unfolding. There are 3 cases where unfolding allows DSP program to be implemented with an iteration period equal to the iteration bound.

: The iteration period cannot be made equal to the iteration bound is when there is a node in the DFG that has computation time greater than iteration bound $_\infty$

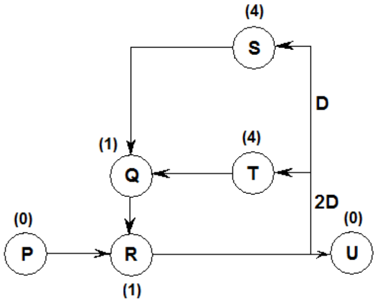


Fig 9 (a)

In this case even retiming cannot be used to reduce the computation time of the critical path of the DFG to $_\infty$. For example the DFG in has9(iteration) bound $_\infty = 3$, but the nodes and each require computation times of 4 . so the minimum sample period after retiming is 4 . This DFG can be unfolded with unfolding factor of 2.

In general, if the computation time of the node , , is greater than the iteration bound $_\infty$, then $[\ /\ _\infty] -$ should be used, where [ ] is the ceiling of , which is the smallest integer greater than or equal to .

The iteration period cannot be made equal to the iteration bound is when the iteration bound is not an integer.

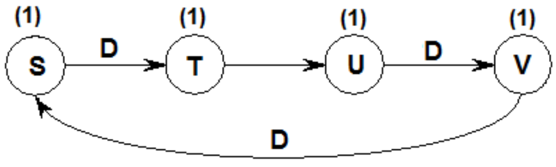


Fig 10

10. h $=$

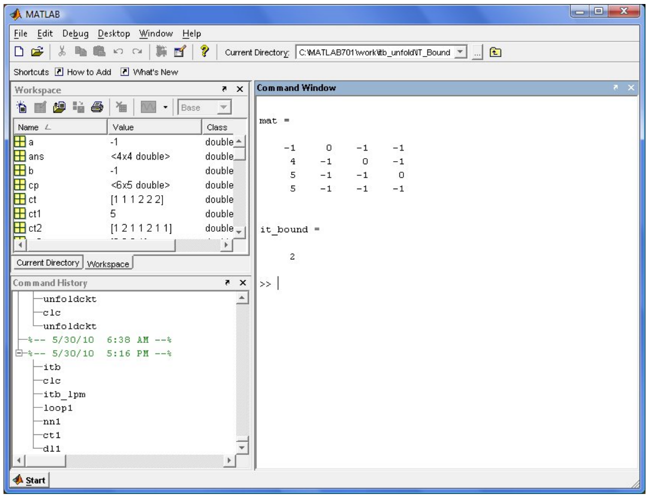
$4/3$

A simple example of this is shown in where10 the DFG has iteration bound $_\infty = 4/3$; however, even retiming cannot be used to achieve a critical path of less than 2 . This DFG can be unfolded with unfolding factor 3.

In general, if a critical loop bound is of the form $/$, where and are mutually coprime, then unfolding should be used.

V. RESULTS

Results of LPM algorithm for DFG in Fig 3.


MATLAB
File Edit Debug Desktop Window Help
Current Directory: C:\MATLAB701\work\itb_unfold\IT_Bound
Shortcuts How to Add What's New
Workspace
Base
Command Window
Current Directory
Workspace
Command History
Start

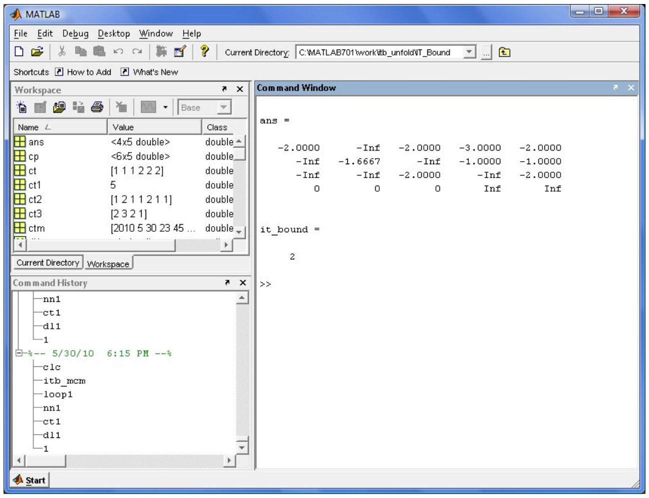
Name	Value	Class
a	-1	double
ans	<4x4 double>	double
b	-1	double
cp	<6x5 double>	double
ct	[1 1 1 2 2 2]	double
ct1	5	double
ct2	[1 2 1 1 2 1 1]	double

```
mat =

    -1     0    -1    -1
     4    -1     0    -1
     5    -1    -1     0
     5    -1    -1    -1


it_bound =

     2

>> |
```

```
  unfoldckt
  clc
  unfoldckt
%-- 5/30/10  6:38 AM --%
%-- 5/30/10  5:16 PM --%
  itb
  clc
  itb_lpm
  loop1
  nn1
  ct1
  dl1
```

In MATLAB command window mat corresponds to and it_bound corresponds to which is equal to the values found theoretically.

Results of MCM algorithm for DFG in Fig 3.


MATLAB
File Edit Debug Desktop Window Help
Current Directory: C:\MATLAB701\work\itb_unfold\IT_Bound
Shortcuts How to Add What's New
Workspace
Base
Command Window
Current Directory
Workspace
Command History
Start

Name	Value	Class
ans	<4x5 double>	double
cp	<6x5 double>	double
ct	[1 1 1 2 2 2]	double
ct1	5	double
ct2	[1 2 1 1 2 1 1]	double
ct3	[2 3 2 1]	double
ctm	[2010 5 30 23 45 ...	double

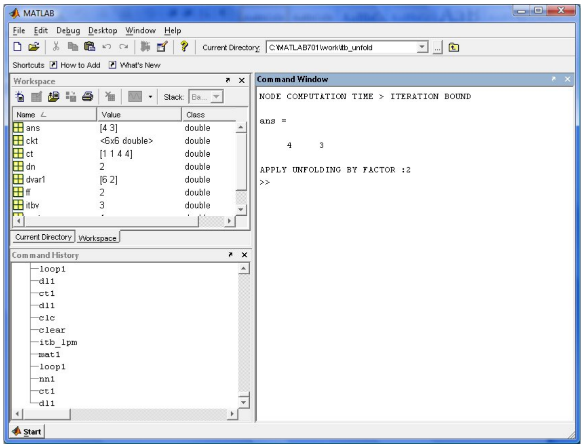
```
ans =

   -2.0000      -Inf   -2.0000   -3.0000   -2.0000
      -Inf   -1.6667      -Inf   -1.0000   -1.0000
      -Inf      -Inf   -2.0000      -Inf   -2.0000
         0         0         0       Inf       Inf


it_bound =

     2

>>
```

```
  nn1
  ct1
  dl1
  1
%-- 5/30/10  6:15 PM --%
  clc
  itb_mcm
  loop1
  nn1
  ct1
  dl1
  1
```

In MATLAB command window ans corresponds to $\frac{(\,)(\,)\;(\,)(\;\;)}{}$ $1 \leq$ ≤ 4 ≤ 3 and0 $\leq$ it_bound corresponds to which is equal to the values found theoretically.

Results of Unfolding for DFG in Fig 9

Case 1:

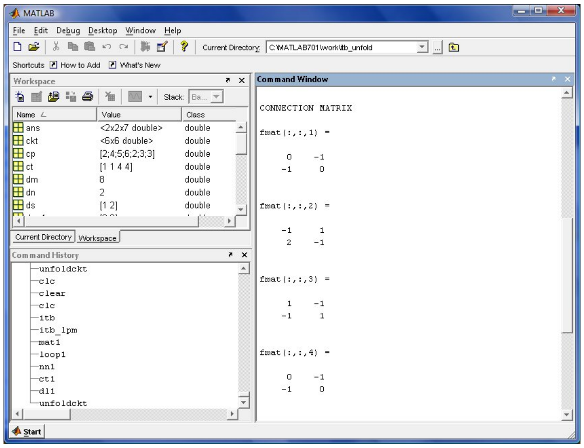


Name	Value	Class
ans	[4 3]	double
ckt	<6x6 double>	double
ct	[1 1 4 4]	double
dn	2	double
dvar1	[6 2]	double
ff	2	double
itbv	3	double

```
loop1
dl1
ct1
dl1
clc
clear
itb_lpm
mat1
loop1
nn1
ct1
dl1
```

```
NODE COMPUTATION TIME > ITERATION BOUND

ans =

     4     3

APPLY UNFOLDING BY FACTOR :2
>>
```

In MATLAB command window ans is a row vector in which the first element corresponds to largest node computation time which is equal to 4 ..and second element corresponds to iteration bound which is equal to 3 . The program also suggest the unfolding factor of 2 to be applied to DFG to achieve iteration period equal to iteration bound.



Name	Value	Class
ans	<2x2x7 double>	double
ckt	<6x6 double>	double
cp	[2;4;5;6;2;3;3]	double
ct	[1 1 4 4]	double
dm	8	double
dn	2	double
ds	[1 2]	double

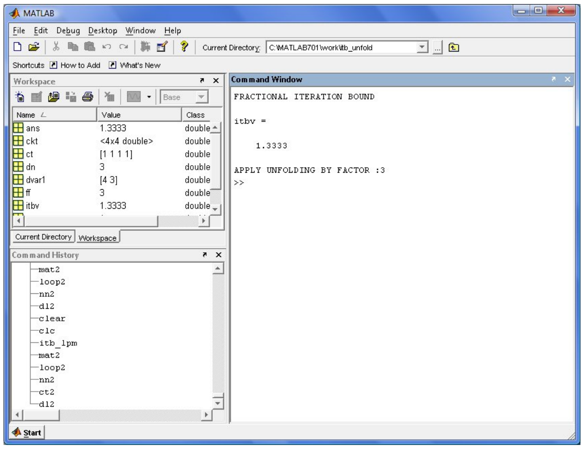
```
unfoldckt
clc
clear
clc
itb
itb_lpm
mat1
loop1
nn1
ct1
dl1
unfoldckt
```

```
CONNECTION MATRIX

fmat(:,:,1) =

     0    -1
    -1     0


fmat(:,:,2) =

    -1     1
     2    -1


fmat(:,:,3) =

     1    -1
    -1     1


fmat(:,:,4) =

     0    -1
    -1     0
```

In command window fmat corresponds to connection matrix of unfolded DFG.

Case 2:

Results of Unfolding for DFG in Fig 10

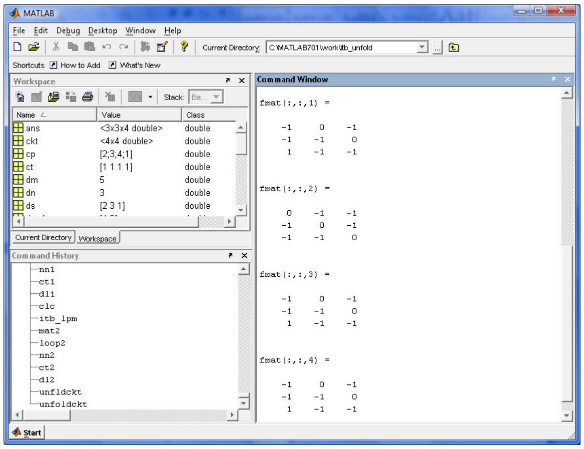


Name	Value	Class
ans	1.3333	double
ckt	<4x4 double>	double
ct	[1 1 1 1]	double
dn	3	double
dvar1	[4 3]	double
ff	3	double
itbv	1.3333	double

```
mat2
loop2
nn2
dl2
clear
clc
itb_lpm
mat2
loop2
nn2
ct2
dl2
```

```
FRACTIONAL ITERATION BOUND

itbv =

    1.3333

APPLY UNFOLDING BY FACTOR :3
>>
```

In MATLAB command window itbv corresponds to fractional iteration bound value =1.33 and the program also suggest the unfolding factor of 3 to be applied to DFG to achieve iteration period equal to iteration bound.



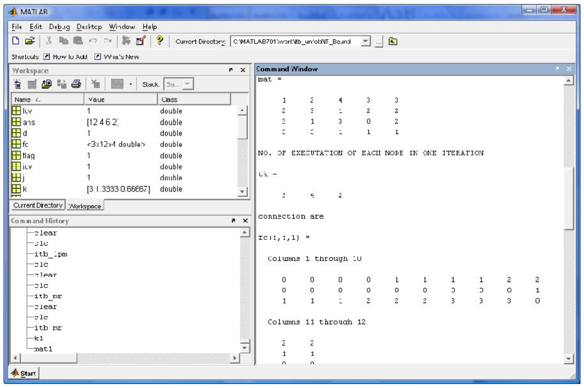
Name	Value	Class
ans	<3x3x4 double>	double
ckt	<4x4 double>	double
cp	[2;3;4;1]	double
ct	[1 1 1 1]	double
dm	5	double
dn	3	double
ds	[2 3 1]	double

```
nn1
ct1
dl1
clc
itb_lpm
mat2
loop2
nn2
ct2
dl2
unfldckt
unfoldckt
```

```
fmat(:,:,1) =

    -1     0    -1
    -1    -1     0
     1    -1    -1


fmat(:,:,2) =

     0    -1    -1
    -1     0    -1
    -1    -1     0


fmat(:,:,3) =

    -1     0    -1
    -1    -1     0
     1    -1    -1


fmat(:,:,4) =

    -1     0    -1
    -1    -1     0
     1    -1    -1
```

In command window fmat corresponds to connection matrix of unfolded DFG.

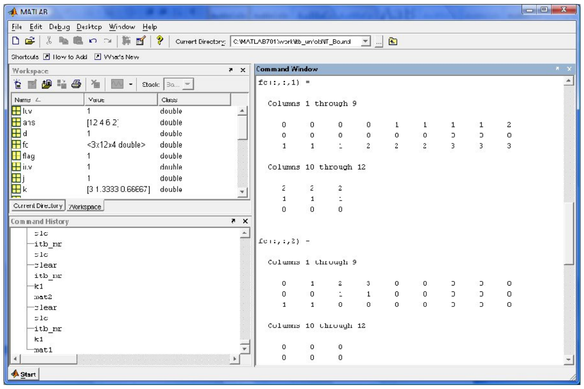
Results of conversion of MRDFG of Fig 6 into SRDFG of Fig 7.



Name	Value	Class
icv	1	double
ans	[1 2 4 6 2]	double
d	1	double
fc	<3x12x4 double>	double
flag	1	double
icv	1	double
j	1	double
k	[3 1.3333 0.66667]	double

```
clear
clc
itb_lpm
clc
clear
clc
itb_mr
clear
clc
itb_mr
k1
mat1
```

```
mat =

     1     2     4     3     3
     2     3     1     2     2
     3     1     3     0     2
     0     0     1     1     1

NO. OF EXECUTATION OF EACH NODE IN ONE ITERATION

kx =

     3     4     2

connection are

fc(:,:,1) =

  Columns 1 through 10

     0     0     0     0     1     1     1     1     2     2
     0     0     0     0     0     0     0     0     0     1
     1     1     1     2     2     2     3     3     3     0

  Columns 11 through 12

     2     2
     1     1
     0     0
```

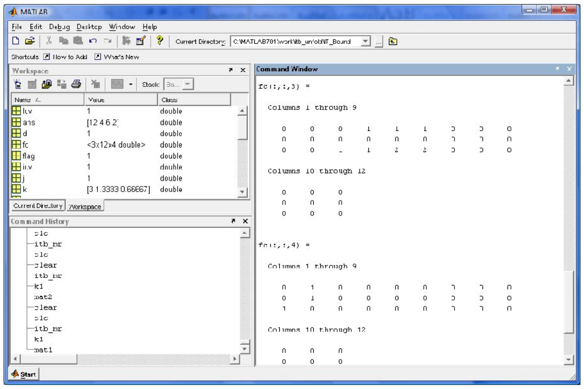


MATLAB
File Edit Debug Desktop Window Help
Current Directory: C:\MATLAB701\work\itb_unfoldNT_Bound
Shortcuts How to Add What's New
Workspace
Stack: Ba...
Current Directory
Workspace
Command History
Command Window
Start

Name	Value	Class
lcv	1	double
ans	[12 4 6 2]	double
d	1	double
fc	<3x12x4 double>	double
flag	1	double
inv	1	double
j	1	double
k	[3 1.3333 0.66667]	double

```
clc
itb_mr
clc
clear
itb_mr
k1
mat2
clear
clc
itb_mr
k1
mat1
```

```
fc(:,:,1) =

  Columns 1 through 9

     0     0     0     0     1     1     1     1     2
     0     0     0     0     0     0     0     0     0
     1     1     1     2     2     2     3     3     3

  Columns 10 through 12

     2     2     2
     1     1     1
     0     0     0


fc(:,:,2) =

  Columns 1 through 9

     0     1     2     3     0     0     0     0     0
     0     0     1     1     0     0     0     0     0
     1     1     0     0     0     0     0     0     0

  Columns 10 through 12

     0     0     0
     0     0     0
```



MATLAB
File Edit Debug Desktop Window Help
Current Directory: C:\MATLAB701\work\itb_unfoldNT_Bound
Shortcuts How to Add What's New
Workspace
Stack: Ba...
Current Directory
Workspace
Command History
Command Window
Start

Name	Value	Class
lcv	1	double
ans	[12 4 6 2]	double
d	1	double
fc	<3x12x4 double>	double
flag	1	double
inv	1	double
j	1	double
k	[3 1.3333 0.66667]	double

```
clc
itb_mr
clc
clear
itb_mr
k1
mat2
clear
clc
itb_mr
k1
mat1
```

```
fc(:,:,3) =

  Columns 1 through 9

     0     0     0     1     1     1     0     0     0
     0     0     0     0     0     0     0     0     0
     0     0     1     1     2     2     0     0     0

  Columns 10 through 12

     0     0     0
     0     0     0
     0     0     0


fc(:,:,4) =

  Columns 1 through 9

     0     1     0     0     0     0     0     0     0
     0     1     0     0     0     0     0     0     0
     1     0     0     0     0     0     0     0     0

  Columns 10 through 12

     0     0     0
     0     0     0
```

Above figures show connection matrix of fig 7.

VI. CONCLUSION

DSP Programs have the property that they are executed from = 0 . These ∞ programs are often represented using DFGs, which can be recursive or non recursive. When the DFG is recursive, the iteration bound is the fundamental limit on the minimum sample period of a hardware implementation of the DSP program. Two algorithms for computing the iteration bound were discussed. The LPM algorithm finds the iteration bound with time complexity (+) and the MCM algorithm finds the iteration bound with time complexity ( ).+The MCM algorithm is usually faster than the LPM algorithm because ≤ holds for most cases.

The iteration bound of Multirate DFG can be determined by first constructing an equivalent single-rate DFG and then computing the iteration bound of the single-rate DFG using LPM and MCM.

Unfolding can be used to reduce the iteration period in DSP algorithms.

VII. REFERENCES

[1] S.H. Gerez**, *"Algorithms for VLSI Design Automation",*** wiley India Edition, John wiley & sons Ltd, ISBN 978-81-265-0821-1, 2007, pp-1-19.

[2] Chao, D.Y. and D.T. Wang, ***Iteration Bounds of Single-Rate Data Flow Graphs for Concurrent Processing",*** IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, Vol.40 (9), (September 1993) pp. 629- 634.

[3] Sabih H. Gerez, Sonia M. Heemstra de Groot, and Otto E. Herrmann**, *"A Polynomial Time Algorithm for computation of the iteration period bound in recursive Data Flow Graphs",*** IEEE Transaction on Circuits and Systems-I: Fundamental Theory and applications, VOL. 39 (January 1992).

[4] Dasdan, A., S.S. Irani and R.K. Gupta, ***"An Experimental Study of Minimum Mean Cycle Algorithms",*** University of California, Irvine, Department of Information and Computer Science, UCI-ICS Technical Report no. 98-32, (1998).

[5] Dasdan A and R.K. Gupta**, *"Faster Maximum and Minimum Mean Cycle Algorithms for System Performance Analysis",*** IEEE Transactions of Computer-Aided Design of Integrated Circuits and Systems, Vol.17 (10), (October 1998).

[6] Karp, R.M.**, *"A Characterization of the Minimum Cycle Mean in a Digraph*** *"*, Discrete Mathematics, Vol.23, (1978), pp. 309-311.

[7] Huang, S.H. and J.M. Rabaey, *"**Maximizing the Throughput of High Performance DSP Applications Using Behavioral Transformations", European Design and Test Conference*, (1994), pp. 25-30.**

[8] Tyan, H.Y. and Y.H. Hu**, "*Minimum Initiation Interval of Multi- Module Recurrent Signal Processing Algorithm Realization with Fixed Communication Delay*",** *International Conference on Acoustics, Speech and Signal Processing, ICASSP '99**, (1999).***

Propagation Effects And Remedies In Microwave Applications

K.Sudhakar[1], Dr.M.V.Subramanyam[2]
[1]Associate Professor, ECE Department, St.Johns College of Engg & Tech., Yemmiganur, Kurnool, A.P.,
[2]Santhiram Engineering College, Nandyal, Kurnool, A.P., India
sudhakar_403@yahoo.co.in

Abstract:

As lower microwave frequency bands become saturated with users, there is a motivation for the development of applications that utilize higher frequencies, especially in the microwave and millimetric range. The major difficulties for a signal in high frequency band are atmospheric absorption and phase dispersion by gases, water vapor and rain drops. Due to its magnetic dipole moment, the gases molecule absorbs power and causes phase dispersion may cause inter symbol interference (ISI). Due to the rain drops the attenuation and differential attenuation occurs for both horizontally polarized and vertically polarized signals. This paper focused on attenuation at horizontal polarization, as well as differential attenuation and differential propagation phase between horizontal (H) and vertical (V) polarization are considered. It is shown that both attenuation and differential attenuation are linearly related to differential propagation phase (ΦDP). This is shown through simulation using (a) gamma raindrop size distributions (RSD) with three parameters (N_0, D_0, m) that are varied over a very wide range representing a variety of rainfall types, and (b) measured raindrop size distributions at a single location using disdrometer.

Key words: *Atmospheric attenuation, Phase dispersion, Differential Propagation Phase.*

I INTRODUCTION

The higher frequencies offer the advantage of lower cost resulting smaller antenna size. Also higher frequencies are preferred, since the power returned by atmospheric scatters is inversely proportional to the fourth power of the wavelength. However, the resulting gain is sensitivity and spatial resolution is vastly offset by attenuating precipitation. An indirect estimate of the specific attenuation A (attenuation in dB over 1 km distance), due to the scatterers can be obtained using empirical relationships such as Z-R (between the reflectivity factor Z and rain rate R) and A-R (between attenuation constant, A and R). In this scheme, the correction for the attenuation of the power received from the nth range location is done using the reflectivity measurements made at all the preceding (n-1) range locations. The attenuation correction is first invoked for the range location nearest to the radar and then at successive range locations.

With the advent of polarization agile radars, a better estimate of R, and hence A, is possible than with Z alone. Aydin etal derived an empirical relationship to estimate A_H (specific attenuation at horizontal polarization) from Z_H (reflectivity at H-polarization) and Z_{DR} (Differential reflectivity) and proposed a correction scheme for C-band data. Dual wavelength radars use both 10-cm and 3-cm wavelengths. This has the advantage that reflectivity estimates unaffected by attenuation are available at all range locations from the 10-cm radar. However , the attenuation correction to the 3-cm reflectivity, and hence to the dual wavelength ratio (DWR) , is done by apportioning the total attenuation along the radial using an empirical A-Z_H (S) relation ; Z_H(S) is the reflectivity factor for horizontal polarization at S-band.

With the recent advent of a number of polarimetric radars at C-and X-bands it is important to consider not only absolute attenuation in rainfall but also differential attenuation (αDP) and differential propagation phase (ΦDP) between the two principal polarization states. These are the horizontal (H) and vertical(V) states for an equioriented, oblate raindrop model whose symmetry axis is closely oriented along the V-direction. The approach adopted here is similar to Holt's (1988) in that both the absolute attenuation (αH or α_V) as well as the differential attenuation (αDP) are linearly related to Φ_{DP} in rainfall. The range profile of the complex observable W/W_2 can be used to directly estimate α_{DP} and Φ_{DP} in a variety of precipitation media. The quantity W is the complex cross-correlation between the two simultaneously received, circularly polarized waves and W_2 is the conventional reflectivity proportional to the cross-polar received power. It is conventional to assume that absolute attenuation at S-band is negligible. However, when Φ_{DP} is large and when accurate reflectivity estimates are needed (±1dB) it is not possible to neglect the attenuation.

A substantial body of literature exists on the measurement and modeling of attenuation and depolarization effects along terrestrial and satellite-earth paths at microwave and millimeter wave frequencies. S-band differential reflectivity radar data have been used to model rain attenuation along satellite earth paths at higher frequencies, to calculate attenuation ratios (frequency scaling) , and to estimate site diversity gain. The basic method is to combine reflectivity and Z_{DR} to estimate the parameters N_0 and D_0 of an exponential raindrop size distribution (RSD) and subsequently to calculate rain path attenuation. In this context it is possible to identify backscatter observables such as Z_H and Z_{DR} that measure the variability of the rain medium as a function of range as opposed to forward scatter observables such as α_H and Φ_{DP} that increase cumulatively with increasing penetration in to the rain medium. Prediction and correction procedures based on backscatter observables are more sensitive to RSD fluctuations as compared to those based on forward scatter observables. Thus prediction and correction procedures based on Φ_{DP} are likely to be more stable than those based on iterative range location-by-range location methods.

II CALCULATIONS BASED ON GAMMA DISTRIBUTIONS

The gamma model can adequately describe many of the natural variations in the RSD. This model has three parameters and is given by

$$N(D) = N_0 D^n \exp[-(3.67 + n)D/D_0] \qquad (1)$$

Where D is the volume equivalent diameter and D_0 is termed median volume diameter of the distribution.

The specific differential phase (K_{DP}) is defined as

$$K_{DP} = \frac{180\lambda}{\pi} \int \mathrm{Re}[f_H(D) - f_V(D)]N(D)dD, \qquad (2)$$

Where λ is the wavelength , and f_H and f_V are the forward scattering amplitudes for horizontally and vertically polarized waves. The one way differential propagation phase (Φ_{DP1}) between two range locations r_1, r_2 is defined as

$$\phi_{DP1} = \int_{r_1}^{r_2} K_{DP}(r)dr \ . \qquad (3)$$

The two-way differential propagation phase Φ_{DP} equals to 2 Φ_{DP1}.

The specific attenuation at H or V polarization is defined as

$$A_{H,V} = 0.4343 \int Q_{H,V}(D)N(D)dD, \quad \mathrm{dB\ km^{-1}} \qquad (4)$$

Where $Q_{H,V}$ (D) are the extinction cross sections for H and V polarized waves. The specific differential attenuation is $A_{DP} = A_H - A_V$. dB km^{-1}

, and the one-way differential attenuation (α_{DP1}) is defined as

$$\alpha_{\rm DP1} = \int_{r_1}^{r_2} A_{\rm DP}(r)dr. \qquad (5)$$

The two-way differential attenuation α_{DP} (= α_H—α_V) is defined as being equal to $2\alpha_{DP1}$.

The reflectivities $Z_{H,V}$ at horizontal (H) and vertical(V) polarizations, respectively are defined as

$$Z_{\rm H,V} = \frac{\lambda^4}{\pi^5|K|^2}\int \sigma_{\rm H,V}(D)N(D)dD, \quad \text{mm}^6\,\text{m}^{-3}$$

Where $\sigma_{H,V}(D)$ are the radar cross sections at H and V polarizations, λ is the wavelength, K = (ε_r-1)/(ε_r+2) and ε_r is the refractive index of water. Differential reflectivity (Z_{DR}) is defined as

$$Z_{\rm DR} = 10 \log(Z_{\rm H}/Z_{\rm V}), \quad \textbf{dB}. \qquad (6)$$

If $S_H(D)$ and $S_V(D)$ are the principal plane elements of the backscatter matrix , then the cross correlation coefficient ρ_{HV} can be defined as

$$\rho_{\rm HV} = \frac{\int S_{\rm H}(D)S_{\rm V}^*(D)N(D)dD}{[\int |S_{\rm H}|^2N(D)dD]^{1/2}[\int |S_{\rm V}|^2N(D)dD]^{1/2}}. \qquad (7)$$

Further , ρ_{HV} = $|\rho_{HV}|$ exp (jδ) where δ is defined as the backscatter different phase shift for a H-V basis. Non zero values of δ imply non-Rayleigh scattering effects.

The scatter reflects the variations imposed on the RSD parameters. Both $A_H(S)$ and $A_H(X)$ are linearly related to $K_{DP}(S)$ with very litter scatter where as $A_{DP}(S)$ shows somewhat larger scatter. The fact that the simulation shows considerable scatter around the mean linear relationship between K_{DP} and A_{DP}. The linear relationship between $A_H(S)$ versus $K_{DP}(S)$ as seen in fig1 can be used to estimate the attenuation at H-polarization experienced by S-band radar (such as NEXRAD) as function of Φ_{DP}. Thefig2 shows two-way attenuation at H-polarization versus the two-way differential propagation phase.

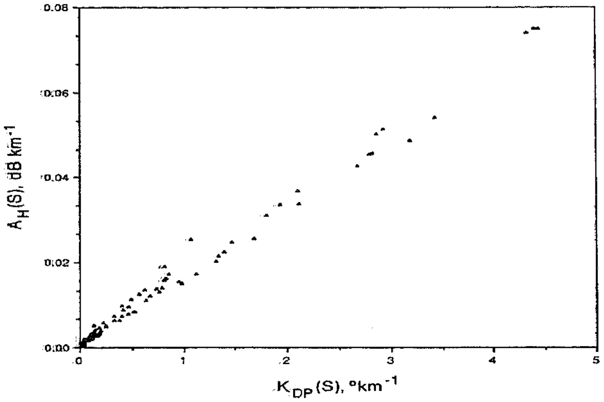


Fig.1: Simulation of specific attenuation at H-polarization versus specific differential phase.

The attenuation is lee than 1 dB if $\Phi_{DP}\leq60^0$, and about 2 dB for Φ_{DP} = 120^0. Thus, if reflectivity estimates with bias less than 1 dB are desired then a correction scheme could be implemented based on the measured Φ_{DP}. In principle, correction for Z_H (C) and Z_{DR}(C) can be attempted if Φ_{DP}(C) can be measured accurately. However Fig.3 shows that the backscatter differential phase shift, δ(C) , may not be negligible when $Z_{DR}\geq2.5$ dB. The correction accuracy will depends on how accurately Φ_{DP}(C) can be estimated given the measured range profile of arg(ρ_{HV}). The DFR range profile is used to estimate X-band attenuation in rainfall assuming that Rayleigh scattering holds at both frequencies. Thus, in the presence of large raindrops or oblate, melting hail the range profile of arg(ρ_{HV}) will be monotonically increasing with "bumps" superimposed near locations of non zero δ. From fig 4 we see that Δ(C) can be estimated if Z_{DR}(C) is know n.

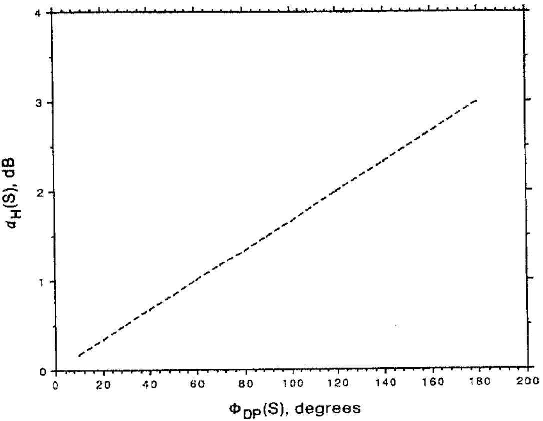


Fig 2: Two-way

attenuation atH-polarization, α_H, versus the two-way differential propagation phase between H and V-polarizations, Φ_{DP}.

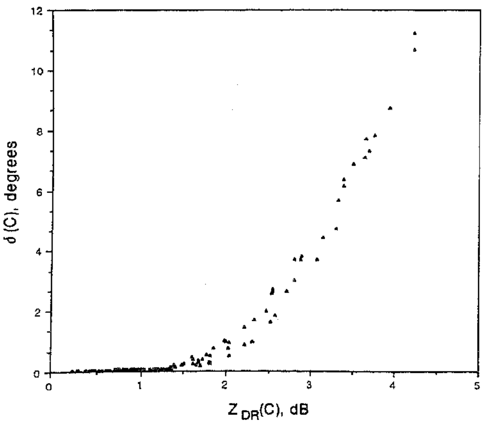


Fig.3. Simulations of backscatter differential phase between H and V-polarizations, δ, versus differential reflectivity. Calculations are at C-band.

III SIMULATIONS BASED ON DISDROMETER MEASUREMENTS:

An impact type disdrometer, developed by Rowland, has been used for measuring the drop-size distributions. This disdrometer has a diameter of two inches, and measures drop sizes in 57 size categories, from 0.2 mm to 5.8 mm with a resolution of 0.1mm. Each distribution is obtained after the disdrometer has sampled 2000 drops. Hence, the sampling time and the sampling volume are dependent on the rain rate. The sampling volume ios equal toηv (D) t where η is the sampling area of the disdrometer in m^2 and v(D) is the terminal velacity in m s^{-1}, and t is the sampling in seconds. The drop diameter to terminal velocity relationship is taken as $v(D) = 3.778\ (D)^{0.67}$m s^{-1} where D is in mm.

Figures 4 show scotterplots of A_H versus K_{DP} at S-band. The S-band simulations show less scatter and hence these data are in good agreement with similar calculations based on the gamma RSD.

The smaller spread seen in fig 4 indicates that the variability of natural rain. The gamma RSD parameter variations cover an unusually wide range of rainfall types where as the disdrometer measurements were required at one location and three of one season. Thus, the gamma simulations show more variability than the disdrometer simulations.

1 Radar measurements :

The radar are used to gather time series data in convective rain shafts from which $A_H(X)$ and $K_{DP}(S)$ can be estimated. Figure 5 AND 6 show range profiles of the DFR and Φ_{DP}. Time series data were generated at low elevation angles through the core of convective rain shafts over a time period of few minutes. The processing technique is used to reduce the statistical fluctuations in both DFR and $\Phi_{DP}(S)$. In this method the reflectivity field (Z_H) was filtered in range using a weighted, moving average filter .

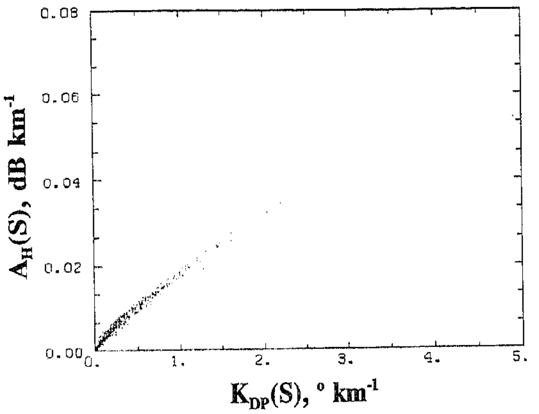


Fig.4: Simulations of specific attenuations at H- polarization versus specific differential phase.

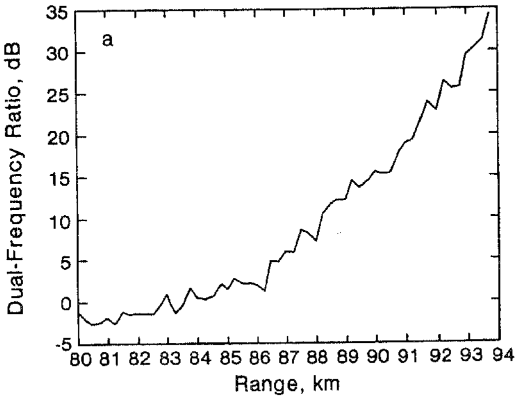


Fig.5:The dual frequency reflectivity ratio in dB versus range.

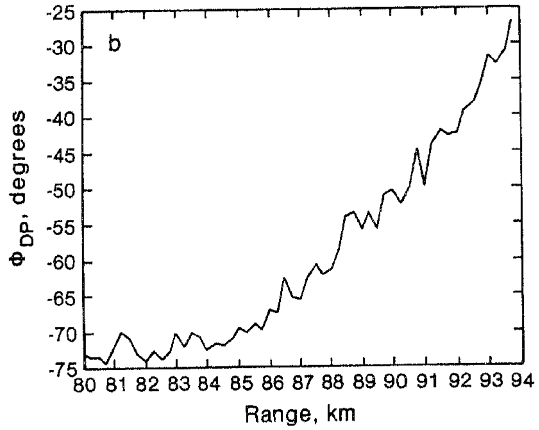


Fig.6 The two-way differential propagation phase between H and V-polarization, Φ_{DP}, versus range.

These radar data show that the attenuation prediction at higher microwave frequencies based on S-band differential propagation phase is feasible.

2 Accuracy of correction procedures at S-band:

The accuracy of correction procedures will depend mainly on three factors, (i) RSD fluctuations, (ii) variability of the estimate of arg(ρ_{HV}) due to measurement fluctuations, and (iii) to nonzero values of δ. The arg (ρ_{HV}) measurement involves "pulse-pair" type algorithms for estimation of differential phase shift. Fluctuations in these estimates can be related to the width (σ_v) of the Doppler spectrum. The bivariate signals can also be used to estimate arg(ρ_{HV}) using Mueller's algorithms. Signals corresponding to two polarizations and having the same Doppler spectrum are independently simulated at two ranges, say, r_1 and r_2. The propagation path is defined to be the range interval (r_2-r_1) which is characterized by a constant K_{DP}. The simulated mean $\Phi_{DP}(r_2)$ is obtained from mean $\Phi_{DP}(r_1)$ by adding $2K_{DP}(r_2-r_1)$.

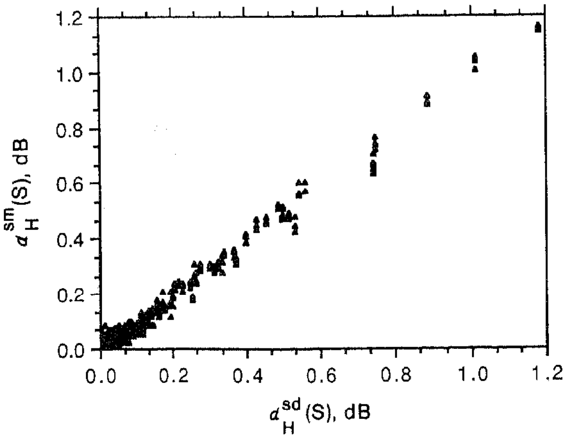


Fig 7: Simulations at S-band of two-way attenuation at H-polarization versus the "exact" attenuation.

Figure 7 shows the simulation at S-band as a scotterplot of α_H^{sm} versus α_H^{sd}. From the simulated Φ_{DP} at r_1 and r_2, $\Delta\Phi_{DP}^{sm}$ is obtained and converted to α_H^{sm} using the mean values in fig 1. This process is repeated for different gamma RSDs . Thus the fluctuations in figure11 include measurement error as well as gamma RSD variations.

If reflectivity estimates biased to less than 1 dB are desired, then the correction procedure can be implemented if $\Phi_{DP}(S) \geq 60^0$.

CONCLUSION

In this paper presentation the propagation effects such as attenuation, differential attenuation, and differential propagation phase in rainfall are examined at microwave frequencies corresponding to S(3.0 GHZ), C(5.5 GHZ), and X (10.0GHZ) bands. Calculations using gamma RSDs and simulation using disdrometer measured RSDs at one location are used to show that both attenuation and differential attenuation can be linearly related to differential propagation phase. A method to correct radar measured reflectivity and Z_{DR} at attenuating frequencies is proposed if the differential propagation phase can be measured by the same radar. The correction accuracy for S-band attenuation was estimated to be ~0.05dB. If S-band reflectivity estimates accurate to within 1 dB are desired, corrections need not be made for $\Phi_{DP} \leq 60^0$. If $\Phi_{DP} > 60^0$, then reflectivity estimates have bias greater than 1 dB and require correction procedures. Simulations indicates that C-band differential attenuation can be corrected to within ~35% of the mean value. This implies that Z_{DR} can be corrected to within, say 0.3 dB if $\Phi_{DP}(C) \leq 60^0$.

REFERENCES

[1] Balakrishnan, N., D. S. Zrnic, J. goldhirsh and J.Rowland, 1989 : Comparison of simulated rain rates from disdrometer data employing polarimetric radar algorithms, J.Atmos. Oceanic. Technol., 6, 476-486.

[2] Bebbington, D.H.O., R.McGuiness and A.R.Holt, 1987; Correction of propagation effects in S-band circular polarization-diversity adars, Proc. IEEE, 34, 431-437.

[3] Bringi, V. N., V. K. Varadan and V.V Varadan , 1983; Average dielectric properties of discerete randam media using multiple scattering theory. IEEE Trans. Antenna and propag., 31, 371- 375.

[4] Chandrasekhar, V., V. N. Bringi and P.J. Brockwell , 1986; Statistical properties of dual- polarized signals. 23rd AMS Conf. Radar Meterology, Amer. Metreor.Soc., 193-196.

[5] Eccles, P. J., 1979; Comparison of remote measurement by a single and dual avelength meteorological radars. IEEE Trans. Geosci Electron., 205-218.

[6] McCormick, G. C., and A. Hendry, 1975; Principals for the radar determination of the polarization properties of precipitation . Radio Science. 10, 421-434.

[7] Marshall, R. E., T. Pratt, E. A. Manus, D. P. Stapor and J.H.Andrews, 1984; S-band radar differential reflectivity measurements in multiple polarization planes along satellite slant paths . Rad. Sci.19,109-114.

[8] Mueller, E. A., 1984; Calculation procedure for differential propagation phase shift. Preprints, 22nd AMS Conf. on Radar Meteorology. Zurich, Amer. Meteor . Soc., 397-399.

[9] Rowland, J. R., 1976; Comparison of two different disdrometers. 17th AMS Conf. Radar meteorology, Seattle, Amer. Meteor Soc.

[10] Twersky, V., 1978; Coherent electromagnetic waves in pair-correlated random distribution of aligned scatters. J. Math. Phys., 19, 215-230.

An Approach to Secure Ad Hoc Networks

Mandlik Sachin B.
Lecturer [PREC, Loni. Pune University]
mandlik.sb@gmail.com
6A, Shivaji Nagar, Sangamner-422605
Cell: 0-9021608173, 0-8975881423

Abstract:

1 Problem Statement

Ad hoc networks are a new wireless networking paradigm for mobile hosts. Unlike traditional mobile wireless networks, ad hoc networks do not rely on any fixed infrastructure. Instead, hosts rely on each other to keep the network connected. One main challenge in design of these networks is their vulnerability to security attacks. In this paper, we study the threats an ad hoc network faces and the security goals to be achieved. We identify the new challenges and opportunities posed by this new networking environment and explore new approaches to secure its communication. In particular, we take advantage of the inherent redundancy in ad hoc networks — multiple routes between nodes — to defend routing against denial of service attacks. We also use replication and new cryptographic schemes, such as threshold cryptography, to build a highly secure and highly available key management service, which forms the core of our security framework.

Introduction

In an ad hoc network, there is no fixed infrastructure such as base stations or mobile switching centers. Mobile nodes that are within each other's radio range communicate directly via wireless links, while those that are far apart rely on other nodes to relay messages as routers. Node mobility in an ad hoc network causes frequent changes of the network topology. Figure 1 shows such an example: initially, nodes A and D have a direct link between them. When D moves out of A's radio range, the link is broken. However, the network is still connected, because A can reach D through C, E, and F

1.1 Security goals

Security is an important issue for ad hoc networks, especially for those security-sensitive applications. To secure an ad hoc network, we consider the following attributes: availability, confidentiality, integrity, authentication, and non-repudiation

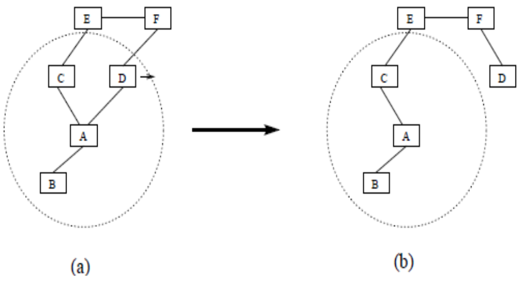


Figure 1: Topology change in ad hoc networks: nodes A, B, C, D, E, and F constitute an ad hoc network. The circle represents the radio range of node A. The network initially has the topology in (a). When node D moves out of the radio range of A, the network topology changes to the one in (b).

Availability ensures the survivability of network services despite denial of service attacks. A denial of service attack could be launched at any layer of an ad hoc network. On the physical and media access control layers, an adversary could employ jamming to interfere with communication on physical channels. On the network layer, an adversary could disrupt the routing protocol and disconnect the network. On the higher layers, an adversary could bring down high-level services. One such target is the key management service, an essential service for any security framework.

Confidentiality ensures that certain information is never disclosed to unauthorized entities.

Integrity guarantees that a message being transferred is never corrupted

Authentication enables a node to ensure the identity of the peer node it is communicating with.

Finally, non-repudiation ensures that the origin of a message cannot deny having sent the message. Non-repudiation is useful for detection and isolation of compromised nodes.

1.2 Challenges

The salient features of ad hoc networks pose both challenges and opportunities in achieving these security goals.

First, use of wireless links renders an ad hoc network susceptible to link attacks ranging from passive eavesdropping to active impersonation, message replay, and message distortion. Eavesdropping might give an adversary access to secret information, violating confidentiality. Active attacks might allow the adversary to delete messages, to inject erroneous messages, to modify messages, and to impersonate a node, thus violating availability, integrity, authentication, and non-repudiation.

Secondly, nodes, roaming in a hostile environment (e.g., a battlefield) with relatively poor physical protection, have non-negligible probability of being compromised. Therefore, we should not only consider malicious attacks from outside a network, but also take into account the attacks launched from within the network by compromised nodes. Therefore, to achieve high survivability, ad hoc networks should have a distributed architecture with no central entities. Introducing any central entity into our security solution could lead to significant vulnerability; that is, if this centralized entity is compromised, then the entire network is subverted.

Thirdly, an ad hoc network is dynamic because of frequent changes in both its topology and its membership (i.e., nodes frequently join and leave the network). Trust relationship among nodes also changes, for example, when certain nodes are detected as being compromised. Unlike other wireless mobile networks, such as mobile IP, nodes in an ad hoc network may dynamically become affiliated with administrative domains. Any security solution with a static configuration would not suffice. It is desirable for our security mechanisms to adapt on-the-fly to these changes.

Finally, an ad hoc network may consist of hundreds or even thousands of nodes. Security mechanisms should be scalable to handle such a large network

1.3 Research Methodology

Scope and roadmap

Traditional security mechanisms, such as authentication protocols, digital signature, and encryption, still play important roles in achieving confidentiality, integrity, authentication, and non-repudiation of communication in ad hoc networks. However, these mechanisms are not suffi- cient by themselves.

We further rely on the following two principles. First, we take advantage of redundancies in the network topology (i.e., multiple routes between nodes) to achieve availability. The second principle is distribution of trust. Although no single node is trustworthy in an ad hoc network because of low physical security and availability, we can distribute trust to an aggregation of nodes. Assuming that any t + 1 nodes will unlikely be all compromised, consensus of at least t + 1 nodes is trustworthy.

All key-based cryptographic schemes (e.g., digital signature) demand a key management service, which is responsible for keeping track of bindings between keys and nodes and for assisting the establishment of mutual trust and secure communication between nodes. We will focus our discussion in Section 3 on how to establish such a key management service that is appropriate for ad hoc networks

2 Secure Routing

To achieve availability, routing protocols should be robust against both dynamically changing topology and malicious attacks. Routing protocols proposed for ad hoc networks cope well with the dynamically changing topology. However, none of them, to our knowledge, have accommodated mechanisms to defend against malicious attacks. Routing protocols for ad hoc networks are still under active research. There is no single standard routing protocol. Therefore, we aim to capture the common security threats and to provide guidelines to secure routing protocols.

In most routing protocols, routers exchange information on the topology of the network in order to establish routes between nodes. Such information could become a target for malicious adversaries who intend to bring the network down.

There are two sources of threats to routing protocols. The first comes from external attackers. By injecting erroneous routing information, replaying old routing information, or distorting routing information, an attacker could successfully partition a network or introduce excessive traffic load into the network by causing retransmission and inefficient routing.

The second and also the more severe kind of threats come from compromised nodes, which might advertise incorrect routing information to other nodes. Detection of such incorrect information is difficult: merely requiring routing information to be signed by each node would not work, because compromised nodes are able to generate valid signatures using their private keys.

To defend against the first kind of threats, nodes can protect routing information in the same way they protect data traffic, i.e., through the use of cryptographic schemes such as digital signature. However, this defense is ineffective against attacks from compromised servers. Worse yet, as we have argued, we cannot neglect the possibility of nodes being compromised in an ad hoc network. Detection of compromised nodes through routing information is also difficult in an ad hoc network because of its dynamically changing topology: when a piece of routing information is found invalid, the information could be generated by a compromised node, or, it could have become invalid as a result of topology changes. It is difficult to distinguish between the two cases.

On the other hand, we can exploit certain properties of ad hoc networks to achieve secure routing. Note that routing protocols for ad hoc networks must handle outdated routing information to accommodate the dynamically changing topology. False routing information generated by compromised nodes could, to some extent, be considered outdated information. As long as there are sufficiently many correct nodes, the routing protocol should be able to find routes that go around these compromised nodes. Such capability of the routing protocols usually relies on the inherent redundancies — multiple, possibly disjoint, routes between nodes — in ad hoc networks. If routing protocols can discover multiple routes (e.g., protocols in ZRP , DSR, TORA, and AODV all can achieve this), nodes can switch to an alternative route when the primary route appears to have failed.

Diversity coding takes advantage of multiple paths in an efficient way without message retransmission. The basic idea is to transmit redundant information through additional routes for error detection and correction.

3 Key Management Service

We employ cryptographic schemes, such as digital signatures, to protect both routing information and data traffic. Use of such schemes usually requires a key management service.

We adopt a public key infrastructure because of its superiority in distributing keys and in achieving integrity and non-repudiation. Efficient secret key schemes are used to secure further communication after nodes authenticate each other and establish a shared secret session key.

In a public key infrastructure, each node has a public/private key pair. Public keys can be distributed to other nodes, while private keys should be kept confidential to individual nodes. There is a trusted entity called Certification Authority (CA) for key management. The CA has a public/private key pair, with its public key known to every node, and signs certificates binding public keys to nodes.

The trusted CA has to stay on-line to reflect the current bindings, because the bindings could change over time: a public key should be revoked if the owner node is no longer trusted or is out of the network; a node may refresh its key pair periodically to reduce the chance of a successful brute-force attack on its private key.

It is problematic to establish a key management service using a single CA in ad hoc networks. The CA, responsible for the security of the entire network, is a vulnerable point of the network: if the CA is unavailable, nodes cannot get the current public keys of other nodes or to establish secure communication with others. If the CA is compromised and leaks its private key to an adversary, the adversary can then sign any erroneous certificate using this private key to impersonate any node or to revoke any certificate.

A standard approach to improve availability of a service is replication. But a naive replication of the CA makes the service more vulnerable: compromise of any single replica, which possesses the service private key, could lead to collapse of the entire system. To solve this problem, we distribute the trust to a set of nodes by letting these nodes share the key management responsibility.

3.1 Analysis

System model: Our key management service is applicable to an asynchronous ad hoc network; that is, a network with no bound on message-delivery and message-processing times. We also assume that the underlying network layer provides reliable links. The service, as a whole, has a public/private key pair. All nodes in the system know the public key of the service and trust any certificates signed using the corresponding private key. Nodes, as clients, can submit query requests to get other clients' public keys or submit update requests to change their own public keys.

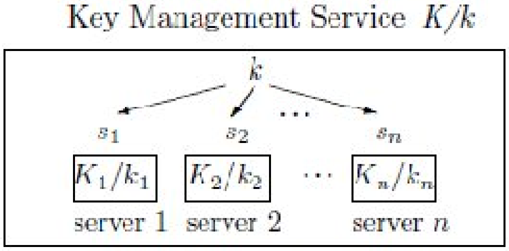


Figure 2: The configuration of a key management service: the key management service consists of n servers. The service, as a whole, has a public/private key pair K/k. The public key K is known to all nodes in the network, whereas the private key k is divided into n shares $s_1, s_2, \ldots, s_n$, one share for each server. Each server i also has a public/private key pair K_i/k_i and knows the public keys of all nodes

Internally, our key management service, with an (n, t +1) configuration ($n \geq 3t +1$), consists of n special nodes, which we call servers, present within an ad hoc network. Each server also has its own key pair and stores the public keys of all the nodes in the network. In particular, each server knows the public keys of other servers. Thus, servers can establish secure links among them. We assume that the adversary can compromise up to t servers in any period of time with a certain duration.

If a server is compromised, then the adversary has access to all the secret information stored on the server. A compromised server might be unavailable or exhibit Byzantine behavior (i.e., it can deviate arbitrarily from its protocols). We also assume that the adversary lacks the computational power to break the cryptographic schemes we employ.

The service is correct if the following two conditions hold:

a.(**Robustness**) The service is always able to process query and update requests from clients. Every query always returns the last updated public key associated with the requested client, assuming no concurrent updates on this entry.

b.(**Confidentiality**) The private key of the service is never disclosed to an adversary.

Thus, an adversary is never able to issue certificates, signed by the service private key, for erroneous bindings

3.2 Threshold cryptography

Distribution of trust in our key management service is accomplished using threshold cryptography . An (n, t + 1) threshold cryptography scheme allows n parties to share the ability to perform a cryptographic operation (e.g., creating a digital signature), so that any t + 1 parties can perform this operation jointly, whereas it is infeasible for at most t parties to do so, even by collusion.

In our case, the n servers of the key management service share the ability to sign certificates. For the service to tolerate t compromised servers, we employ an (n, t + 1) threshold cryptography scheme and divide the private key k of the service into n shares (s_1, s_2, . . . , s_n), assigning one share to each server. We call (s_1, s_2, . . . , s_n) an (n, t + 1) sharing of k. Figure 2 illustrates how the service is configured.

For the service to sign a certificate, each server generates a partial signature for the certificate using its private key share and submits the partial signature to a combiner. With t + 1 correct partial signatures the combiner is able to compute the signature for the certificate. However, compromised servers (there are at most t of them) cannot generate correctly signed certificates by themselves, because they can generate at most t partial signatures. Figure 3 shows how servers generate a signature using a (3, 2) threshold signature scheme.

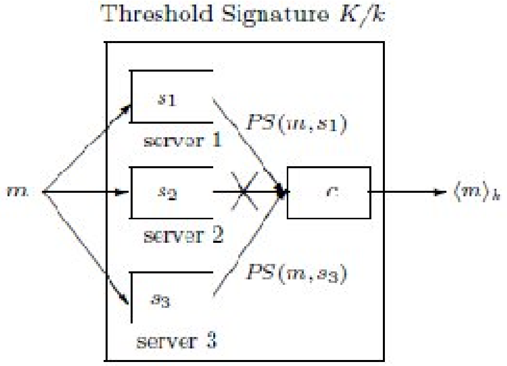


Figure 3: Threshold signature: given a service consisting of 3 servers. Let K/k be the public/private key pair of the service. Using a (3, 2) threshold cryptography scheme, each server i gets a share s_i of the private key k. For a message m, server i can generate a partial signature PS (m, s_i) using its share s_i. Correct servers 1 and 3 both generate partial signatures and forward the signatures to a combiner c. Even though server 2 fails to submit a partial signature, c is able to generate the signature $(m)_k$ of m signed by service private key k

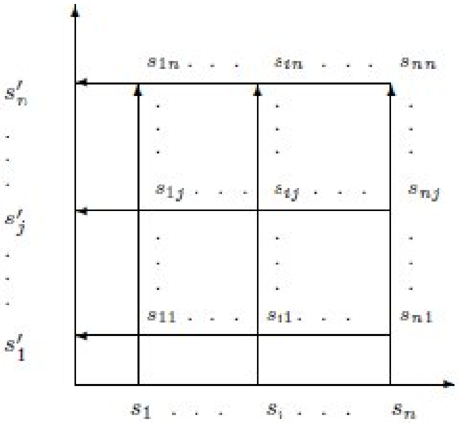
When applying threshold cryptography, we must defend against compromised servers. For example, a compromised server could generate an incorrect partial signature. Use of this partial signature would yield an invalid signature. Fortunately, a combiner can verify the validity of a computed signature using the service public key. In case verification fails, the combiner tries another set of t + 1 partial signatures. This process continues until the combiner constructs the correct signature from t + 1 correct partial signatures. More efficient robust combining schemes are proposed [13, 12]. These schemes exploit the inherent redundancies in the partial signatures (note that any t +1 correct partial signatures contain all the information of the final signature) and use error correction codes to mask incorrect partial signatures. In [13], a robust threshold DSS (Digital Signature Standard) scheme is proposed. The process of computing a signature from partial signatures is essentially an interpolation. The authors uses the Berlekamp and Welch decoder, so that the interpolation still yields a correct signature despite a small portion (fewer than one fourth) of partial signatures being missing or incorrect.

3.3 Proactive security and adaptability

Besides threshold signature, our key management service also employs share refreshing to tolerate mobile adversaries and to adapt its configuration to changes in the network.

Mobile adversaries are first proposed by Ostrovsky and Yung to characterize adversaries that temporarily compromise a server and then move on to the next victim (e.g., in form of viruses injected into a network). Under this adversary model, an adversary might be able to compromise all the servers over a long period of time. Even if the compromised servers are detected and excluded from the service, the adversary could still gather more than t shares of the private key from compromised servers over time. This would allow the adversary to generate any valid certificates signed by the private key.

Proactive schemes are proposed as a countermeasure to mobile adversaries. A proactive threshold cryptography scheme uses share refreshing, which enables servers to compute new shares from old ones in collaboration without disclosing the service private key to any server. The new shares constitute a new (n, t + 1) sharing of the service private key. After refreshing, servers remove the old shares and use the new ones to generate partial signatures. Because the new shares are independent of the old ones, the adversary cannot combine old shares with new shares to recover the private key of the service. Thus, the adversary is challenged to compromise t + 1 servers between periodic refreshing



s_{1n} . . . s_{in} . . . s_{nn}
s'_n
s_{1j} . . . s_{ij} . . . s_{nj}
s'_j
s_{11} . . . s_{i1} . . . s_{n1}
s'_1
s_1 . . . s_i . . . s_n

Figure 4: Share refreshing: given an (n, t+1) sharing (s_1, . . . , s_n) of a private key k, with share s_i assigned to server i. To generate a new (n, t+1) sharing (s^0_1, . . . , s^0_n) of k, each server i generates subshares s_{i1}, s_{i2}, . . . , s_{in}, which constitute the ith column in the figure. Each subshare s_{ij} is then sent securely to server j. When server j gets all the subshares s_{1j} , s_{2j} , . . . , s_{nj} , which constitute the jth row, it can generate its new share s^0_j from these subshares and its old share s_j .

Share refreshing relies on the following homomorphic property. If (s^1_1, s^1_2, . . . , s^1_n) is an (n, t + 1) sharing of k_1 and (s^2_1, s^2_2, . . . , s^2_n) is an (n, t + 1) sharing of k_2, then $(s^1_1 + s^2_1, s^1_2 + s^2_2, . . . , s^1_n + s^2_n)^v$ is an (n, t + 1) sharing of $k_1 + k_2$. If k_2 is 0, then we get a new (n, t + 1) sharing of k_1.

Given n servers. Let (s_1, s_2, . . . , s_n) be an (n, t + 1) sharing of the private key k of the service, with server i having s_i. Assuming all servers are correct, share refreshing proceeds as follows: first, each server randomly generates (s_{i1}, s_{i2}, . . . , s_{in}), an (n, t+1) sharing of 0. We call these newly generated s_{ij} 's subshares. Then, every subshare s_{ij} is distributed to server j through a secure link. When server j gets the subshares s_{1j} , s_{2j} , . . . , s_{nj} , it can compute a new share from these subshares and its old share ($s^0_j = s_j + \Sigma^n_{i=1} s_{ij}$). Figure 4 illustrates a share refreshing process.

Share refreshing must tolerate missing subshares and erroneous subshares from compromised servers. A compromised server may not send any subshares. However, as long as correct servers agree on the set of subshares to use, they can generate new shares using only subshares generated from t + 1 servers. For servers to detect incorrect subshares, we use verifiable secret sharing schemes, for example, those in [7, 33]. A verifiable secret sharing scheme generates extra public information for each (sub)share using a one-way function. The public information can testify the correctness of the corresponding (sub)shares without disclosing the (sub)shares.

A variation of share refreshing also allows the key management service to change its configuration from (n, t + 1) to (n^0 , t^0 + 1). This way, the key management service can adapt itself, on the fly, to changes in the network: if a compromised server is detected, the service should exclude the compromised server and refresh the exposed share; if a server is no longer available or if a new server is added, the service should change its configuration accordingly.

3.4 Asynchrony

Existing threshold cryptography and proactive threshold cryptography schemes assume a synchronous system (i.e., there is a bound on message-delivery and message-processing times). This assumption is not necessarily valid in an ad hoc network, considering the low reliability of wireless links and poor connectivity among nodes. In fact, any synchrony assumption is vulnerability in the system: the adversary can launch denial of service attacks to slow down a node or to disconnect a node for a long enough period of time to invalidate the synchrony assumption. Consequently, protocols based on the synchrony assumption are inadequate.

To reduce such vulnerability, our key management service works in an asynchronous setting. Designing such protocols is hard; some problems may even be impossible to solve . The main difficulty lies in the fact that, in an asynchronous system, we cannot distinguish a compromised server from a correct but slow one.

One basic idea underlying our design is the notion of weak consistency: we do not require that the correct servers be consistent after each operation; instead, we require enough correct servers to be up-to-date. For example, in share refreshing, without any synchrony assumption, a server is no longer able to distribute the subshares to all correct servers using a reliable broadcast channel. However, we only require subshares to be distributed to a quorum of servers. This suffices, as long as correct servers in such a quorum can jointly provide or compute all the subshares that are distributed. This way, correct servers not having certain subshare(s) could recover its subshare(s) from other correct servers.

Another important mechanism is the use of multiple signatures for correct servers to detect and to reject erroneous messages sent by compromised servers. That is, we require that certain messages be accompanied with enough signatures from servers. If a message contains digital signatures from a certain number (say, t + 1) of servers testifying its validity, at least one correct server must have provided one signature, thus establishing the validity of the message.

Significance

This paper focuses on how to secure routing and how to establish a secure key management service in an ad hoc networking environment. These two issues are essential to achieving our security goals. Besides the standard security mechanisms, we take advantage of the redundancies in ad hoc network topology and use diversity coding on multiple routes to tolerate both benign and Byzantine failures. To build a highly available and highly secure key management service, we propose to use threshold cryptography to distribute trust among a set of servers. Furthermore, our key management service employs share refreshing to achieve proactive security and to adapt to changes in the network in a scalable way. Finally, by relaxing the consistency requirement on the servers, our service does not rely on synchrony assumptions. Such assumptions could lead to vulnerability. A prototype of the key management service has been implemented, which shows its feasibility.

The paper represents the first step of our research to analyze the security threats, to understand the security requirements for ad hoc networks, and to identify existing techniques, as well as to propose new mechanisms to secure ad hoc networks. More work needs to be done to deploy these security mechanisms in an ad hoc network and to investigate the impact of these security mechanisms on the network performance

5 Conclusion

In this paper, we have analyzed the security threats an ad hoc network faces and presented the security objectives that need to be achieved. On one hand, the security-sensitive applications of ad hoc networks require high degree of security; on the other hand, ad hoc networks are inherently vulnerable to security attacks. Therefore, security mechanisms are indispensable for ad hoc networks. The idiosyncrasy of ad hoc networks poses both challenges and opportunities for these mechanisms.

References

[1]Y. Desmedt. Threshold cryptography. European Transactions on Telecommunications, 5(4):449–457, July–August 1994

[2]Y. Desmedt and S. Jajodia. Redistributing secret shares to new access structures and its applications. Technical Report ISSE TR-97-01, George Mason University, July 1997

[3]A. Ephremides, J. E. Wieselthier, and D. J. Baker. A design concept for reliable mobile radio networks with frequency hopping signaling. Proceedings of the IEEE, 75(1):56–73, January 1987.

[4]M. J. Fischer, N. A. Lynch, and M. S. Peterson. Impossibility of distributed consensus with one faulty processor. Journal of the ACM, 32(2):374–382, April 1985

Integrated Real-Time Application for ATM Security System

Praveen Kumar Y G[#1], B A Sujathakumari[*2]
[#] Sri Jayachamarajendra College Of Engineering
Department of Electronics and Communication
Manasa Gangothri, Mysore – 570 006
[1]praviraj.star@gmail.com, [2]basujatha@yahoo.com

Abstract

The use of mobile handheld devices is expanding rapidly both within the business and individual context. These devices are now essential tools that offer competitive business advantages in today's growing world of ubiquitous computing environment. The technology advancement has made it possible to embed more facilities in mobile phones. Though they provide benefits, they also pose new risks on security either by the information they contain or information that they can access remotely. Secure money transaction is of serious concern in growing use of cash cards and internet transactions. However, there has been limited research focus on security and flexibility. This project allows the users to use their mobile phones to securely withdraw the cash from ATM machines.

In our proposed system, in every transaction with the ATM card, a handshaking signal is achieved with the cardholder. The handshaking method is achieved by transmitting the randomly generated code to the mobile of the cardholder by means of a GSM modem. From the acknowledgement received from the cardholder further transaction proceeds.

Keywords ATM, GSM Modem, RFID, Security, Authentication

I. INTRODUCTION

The present day world has been termed as "highly networking based world". ATM technology perhaps, is the most complex networking technology we ever have. To secure such a complex system is even more difficult than designing it. This project gives design architecture of a money withdrawal system with higher security using GSM technology. There has been a growing use of GSM in different environments. A typical examples includes telephony ((Teleservices (TS), Bearer services (BS), Supplementary services (SS)). The major use of GSM is being the high degree of flexibility, high quality signal and link integrity and its low cost infrastructure. In this project, this technology has been used in a new application called Secured Money Withdrawal System.

A. Proposed work

GSM technology allows higher security into any design, regardless of microprocessor used. This project develops a system, which employs GSM technology, mobile and ATM cards equipped with RFID tags for secured cash transaction. This project uses ARM7 based controller and incorporates various hardware and software technologies like GSM, RFID technology and Magnetic fields of RFID tags in hardware and Hi-top, H-JTAG emulator and OrCAD in software.

B. Scope of the Work

This project sets sight on authenticating the conventional Credit card transaction system. In the prevailing system though the Credit card paves a convenient mode of transactions, it is subjected to security jeopardy. As technology extends its limit, the way of hacking and cracking also goes along the road. Hence there is a call for higher security in money transaction.

In my proposed system, in every transaction with the Credit card, a handshaking signal is achieved with the cardholder. The handshaking method is achieved by transmitting the randomly generated code to the mobile of the cardholder by means of a GSM modem. Until and unless the cardholder enters the random code, further transactions cannot be achieved. The main purpose of interacting via mobile phones is to improve the security of transactions.

II. DESCRIPTION

The main objective of the project is to develop a system for secured money transaction using GSM technology and RFID. The general block diagram of the overall project is as shown in Fig.1

When RFID tag is brought near the RFID module, the corresponding tag ID is compared with the database, which is stored in the controller. If the tag's ID matches with any one of the database stored, then the card is said to be valid and controller instructs the GSM Modem to send the code (random) to the mobile of the cardholder. Once the user enters the correct code, the electromagnetic door gets unlocked with the LED in ON state. On the other hand if the card is invalid or the card holder enters incorrect code, the controller puts ON the buzzer. The status of the system in each step is displayed using LCD.

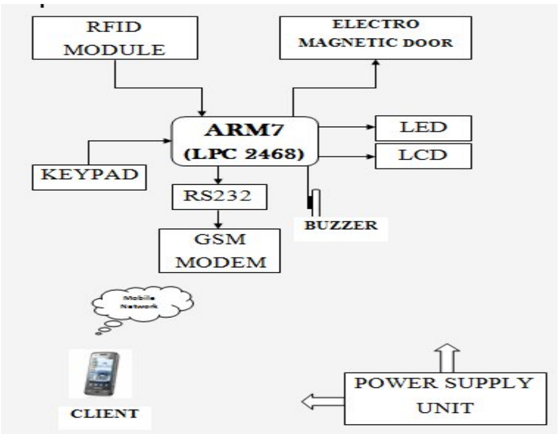


Fig 1: General block diagram of the overall project

The success of GSM and RFID in recent days makes them ideal for my system which provides higher security.

RFID communicates with controller serially at the baud rate of 19200bps.The command “DTGU” is used to get the tag’s ID. RFID Commands ends with <LF><CR> and response from RFID module starts with 0x02 and ends with 0x03. There are generally three types of RFID tags: active RFID tags, which contain a battery and can transmit signals autonomously, passive RFID tags, which have no battery and require an external source to provoke signal transmission, and battery assisted passive (BAP) RFID tags, which require an external source to wake up but have significant higher forward link capability providing greater range.

The command response flow of RFID is as shown in Fig 2. Upon reset, the system will check for the RFID tag in the reader’s range. The range of the reader is <=70mm. This avoids the problem of skimming. If the RFID tag is present in the readers range, RFID reader will reads the unique ID. Tag’s ID is of 4-Byte length. If it is not present in the reader’s range, the response of the reader is NT meaning ‘No Tag’.

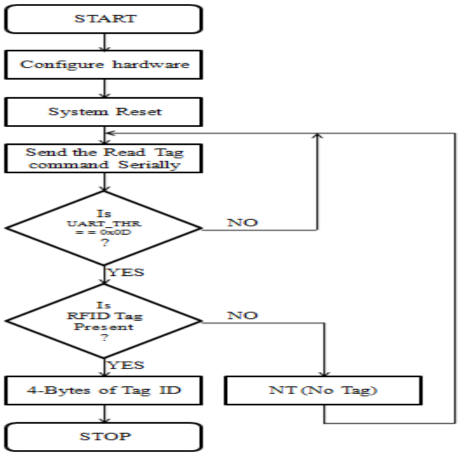


Fig 2: Command Response flow of RFID reader

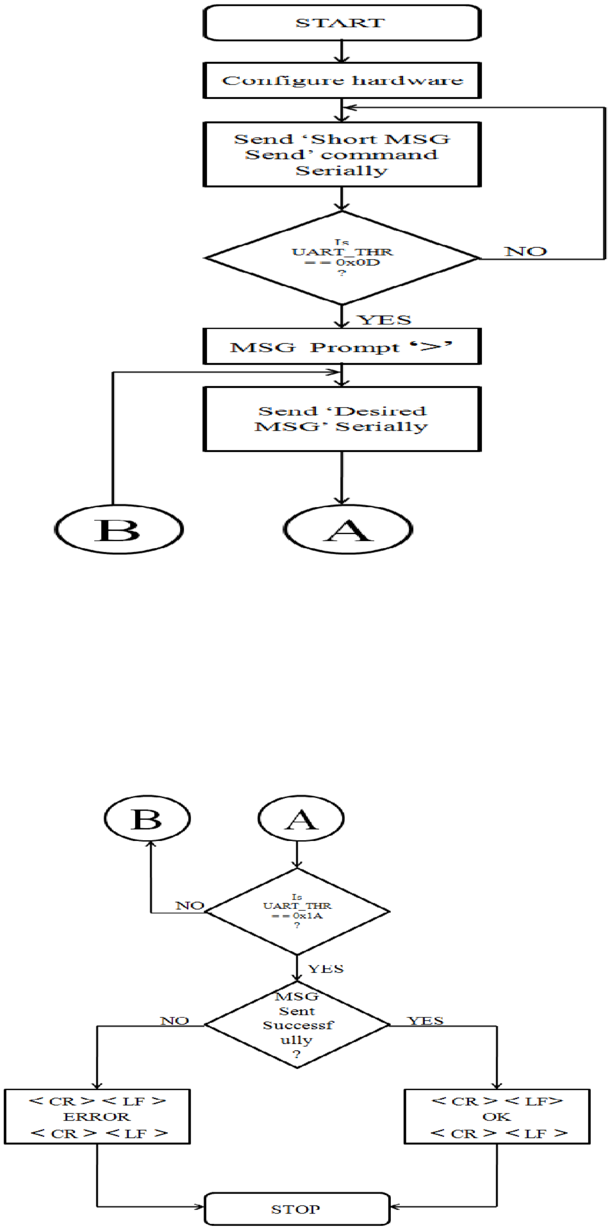

START
Configure hardware
Send ‘Short MSG Send’ command Serially
Is UART_THR = = 0x0D ?
NO
YES
MSG Prompt ‘>’
Send ‘Desired MSG’ Serially
B
A
B
A
NO
Is UART_THR = = 0x1A ?
YES
MSG Sent Successfully ?
NO
YES
< CR > < LF > ERROR < CR > < LF >
< CR > < LF> OK < CR > < LF >
STOP

Fig 3: Command Response flow of GSM modem

GSM modem communicates with the controller serially at the baud-rate of 9600bps. It is programmed by AT commands. In my design GSM modem is used to send the random code to the mobile of the cardholder. The command “AT+CMGS” is sent serially by writing into the transmit hold register and the response from the modem can be read from receiver buffer register. Command ends with <CR> character and responses starts and ends with<CR><LF>. The command response flow of GSM modem is shown in Fig 3

III FUNCTIONAL FLOW

The functional flow of our system is as shown in Fig 4. Once the hardware is configured for our application and upon reset, the system checks whether the config key is pressed. If it is pressed, the system gets into the configuration mode in which we can create and update our database. If config key is not pressed, the system operates in normal mode. When the RFID tag is brought near the reader, it reads the tag's ID and compares the tag's ID with the database. If the tag's ID is not present in the database, the system displays "Invalid Card" and gets back to original state. If it is present in database, randomly generated secret code is sent to the mobile of the cardholder and the system waits for the user to enter the code. When the user presses the enter key after entering the code that he/she received, the entered code is compared with the code that has been sent to the cardholder's mobile. If the code which was sent to the cardholder and code entered by the user matches, the user is allowed to access the ATM and this is indicated by LED and displaying "Accessible". If not, the user can not access the ATM and this is indicated by using buzzer and displaying "Not Accessible".

The status of the system for each individual process is displayed on the LCD. Accessibility of ATM is indicated by unlocking the electro-magnetic door on the other hand inaccessibility of ATM is indicated by locking the door.

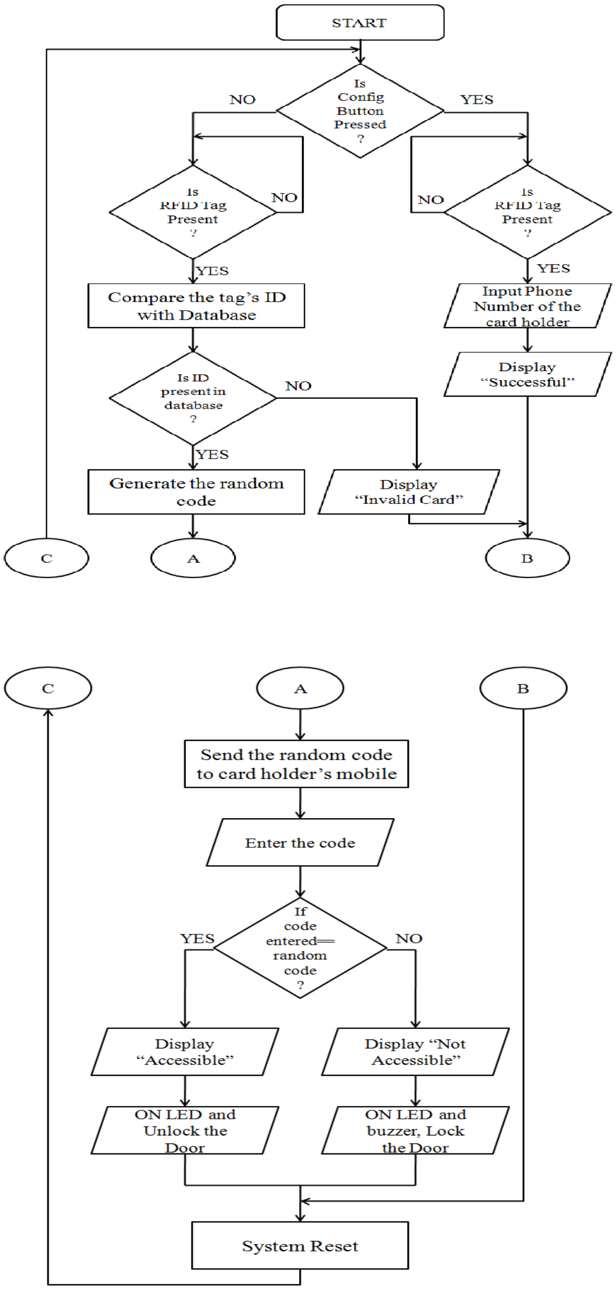

START
Is Config Button Pressed ?
NO
YES
Is RFID Tag Present ?
NO
NO
Is RFID Tag Present ?
YES
YES
Compare the tag's ID with Database
Input Phone Number of the card holder
Display "Successful"
Is ID present in database ?
NO
YES
Generate the random code
Display "Invalid Card"
C
A
B
C
A
B
Send the random code to card holder's mobile
Enter the code
If code entered== random code ?
YES
NO
Display "Accessible"
Display "Not Accessible"
ON LED and Unlock the Door
ON LED and buzzer, Lock the Door
System Reset

Fig 4: Function Flow of the System

IV Results

The source code has been developed in C language, then the code is converted into HEX file using HiTop followed by target programming. Now the target board is ready for our application. The following step shows the step wise operation of our system.

A. System on Power up

When we connect the set up to the supply, it resets. Now the system is completely operative and the overall system on power up is as shown in Fig 5.The system waits until a RFID tag is brought near the RFID reader. The status is displayed on LCD, showing NT which means No Tag.

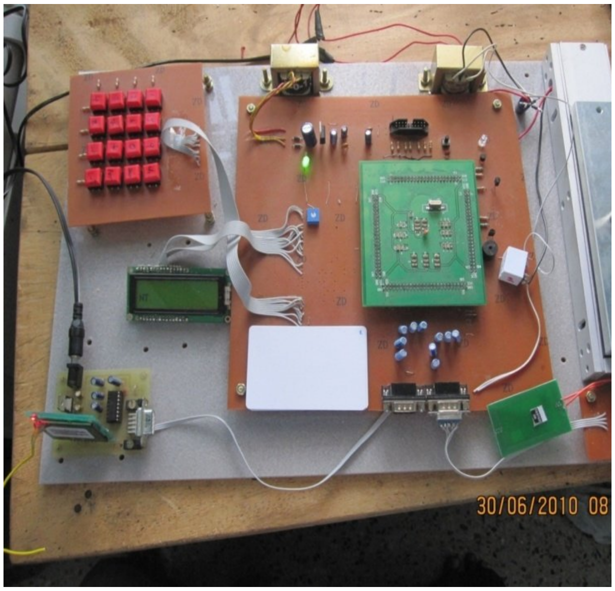


Fig 5: System on Power up

B. Display after Reading the RFID Tag

If the RFID tag is not brought near the reader the display shows NT which means ‘No Tag’. When the RFID Tag is brought near the RFID module, it reads the Tag ID and the ID is displayed on the LCD and is as shown in Fig 6. Then controller compares the tag’s ID with the database stored and if it is a valid card, it instructs the GSM modem to send the random code to the cardholder’s mobile. If not, the system gets back to the initial status.

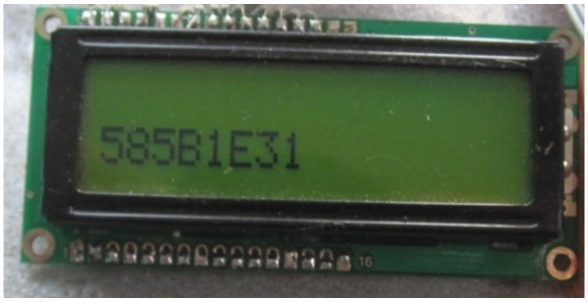


Fig 6: Display after Reading the RFID Tag

C. Display after entering the Pin-Code

If the tag’s ID is present in the database, the GSM modem sends the random code to the cardholder’s mobile and the system waits till he/she enters the code. Once after the user enters the pin-code using keypad, the entered code appears as shown in Fig 7.

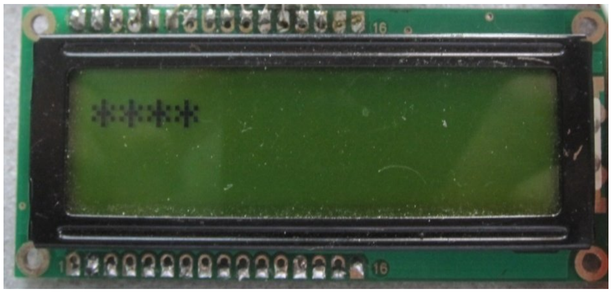


Fig 7: Display after Entering the Pin-Code

D. Accessibility of the ATM

The pin-code entered by the user is compared with the pin-code which has been sent to cardholder’s mobile. If both the code matches, the user is allowed for accessing the ATM and is displayed as shown in Fig 8.

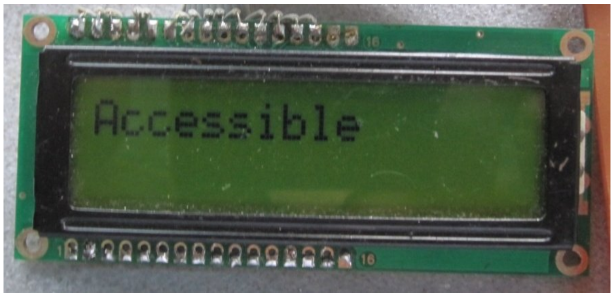


Fig 8: Display after the User is allowed for Accessing the ATM

It is required to maintain and update the database of our system in daily basis. The users are not allowed to have access to the database. Only an authorized person who maintains the system has the authority to change and update the database.

If the new user wants to access the ATM, an authorized person who maintains the system can make the system to operate in configuration mode. Then the database of the new user is updated thereby making him/her as a valid user. The system can be switched back to its normal mode from configuration mode by resetting the system.

Any user can access the ATM if once their database is updated in our system's database and can assure the higher security, because of the success of RFID and GSM technologies in recent times.

III. CONCLUSION

This project presents an architecture that can be used as a means of interaction between mobile phone, ATM machine and the cardholder for the purpose of withdrawing cash. The proposed design, allows the use of mobile phones as a tool of interaction and provides higher security.

In current systems, the message will be sent to the cardholder's mobile only after the money has been withdrawn thus posing a problem, if the ATM card is lost somewhere and the password is hacked. Thus security for the ATM is a critical issue.

This problem is overcome by our system where without the authentication from the cardholder, the money transaction is a word of impossibility. There by providing the higher security, the users can assure that their account balance is safe even if the ATM card is lost accidentally.

Our system uses LPC2468 controller as it is ideal for multi-purpose communication application and supports multiple UARTs (4) which is required for our system. With this system, I have tried sincerely to assure the security concerns as the technologies used in our system are GSM and RFID. We can be sure of the security issues because of the success of GSM and RFID technology in this modern world.

IV. FUTURE ENHANCEMENT

The world around is rapidly changing and in the same way rapidly developing. Today's technology is being outdated is next day. Thus, we cannot be sure of any systems being occupied or used for a very long duration. Therefore, we have to design a system in such a fashion that it should be flexible for advancing technologies without compromising on the basic objective of the work.

Our system uses both RFID and GSM technology. The simplicity of the RFID and GSM protocol and the flexibility makes them ideal for use with small microcontrollers.

However, because the ATM is more susceptible to security threats, we can incorporate other security measures such as making use of biometric data as the further authentication layer. Any of the biometric information that measures behavioural or physical traits can be used to identify the user, by making use of individual anatomy or physiology. Hence the biometric data can be used as the unique personal attribute for security and authentication purposes.

Also it is important to point out that single sign-on process can improve the reliability of identity management and access control. Having this application as the security measure, could remove the current risk of identity theft and meet the required security standards of implementing the secured ATM systems.

REFERENCES

[1] Abdullahi Arabo: "Secure Cash Withdrawal through Mobile Phone/Device" *International Conference on Compute and Communication Engineering 2008,volume 978-1-4244-1692-9/08, 13-15 Sept. 2008*

[2] Valkkynen,P. Korhonen,I. Plomp,J., Tuomisto,T., cluitmane,L., Ailisto,H. and Seppa,H: *"A user interaction paradigm for physical Browsing and near-object control based on tags":in proc.Physical Interaction Workshop on Real World User Interfaces,2003.*

[3] Lauri Pohjanheimo, Heikki Keränen and Heikki Ailisto:"Implementing TouchMe Paradigm with a Mobile Phone",*ACM International Conference Proceeding Series; Vol. 121.*

[4] David Seal, *ARM Architecture Reference Manual,* 2nd Edition, Addison - Wesley, December, 2000.

[5] Mahesh Bhuptani, "Deploying radio frequency identification systems".

[6] Steven Shepard, "Radio frequency identification" Mcgraw Hill networking professional.

[7] "GSM World statistics" *GSM Association. 2010. Retrieved 2010-06-08.*

[8] "Two Billion GSM Customers Worldwide"*3G Americas. June 13, 2006. Retrieved 2007-01-08.*

[9] "Texas Instruments Executive Meets with India Government Official to outline Benefits of Open Standards to drive mobile phone penetration". *Texas Instruments. July 12, 2006. Retrieved 2007-01-08.*

[10] http://www.arm.com

[11] http://www.nxp.com

[12] http://www2.nowsms.com/gsm%20modems.htm

[13] http://www.rfidjournal.com

[14] http://www.datasheetcatalog.com

TIME REDUCTION FOR WIRELESS COMMUNICATION USING OFDM

[1]Balraj B, [2]Sivakumar D, [3]Thamarai Selvi D

[1]Departement of Electronics & Instrumentation Engineering,
Annamalai University,Annamalai Nagar – 608002, Tamilnadu, India.
Email : balrajece@yahoo.com
[2]Professor / Electronics & Instrumentation Engineering,
Annamalai University,Annamalai Nagar – 608002, Tamilnadu, India.
Email : dsk2k5@gmail.com
[3]Assistant Professor / Electronics & Communication Engineering,
Krishnasamy College of Engineering and Technology, Cuddalore – 607109, Tamilnadu, India.
Email : thamarai_vlsi@yahoo.com

Abstract — Orthogonal Frequency Division Multiplexing (OFDM) is a multi–carrier modulation system employing Frequency Division multiplexing of orthogonal sub–carrier, each modulating a low bit – rate digital stream. OFDM symbol is generated by computing IFFT. The design of IFFT / FFT is the main consideration in OFDM transceiver design. The parallel – pipelined FFT architecture based on multiplier less implementation targeting in High – Speed Wireless Communication application such as IEEE 802.11 wireless base band chip and CDMA. These have advantages of high throughput and high power efficiency. The multiplier – less architecture uses shift and addition operations for complex multiplication. By using low power butterfly the resulting power and area savings are increased upto 20%. 64 point FFT is used compared to Wallace tree multiplier.

Key-Word : FFT, IFFT, OFDM, CDMA

I. INTRODUCTION

OFDM is a multichannel modulation system employing FDM of orthogonal sub-carriers each modulating a low bit rate digital stream. In OFDM, to overcome the problem of bandwidth wastage, N overlapping but orthogonal sub-carriers each carrying a baud rate of 1/T and spaced 1/T apart are used. For recent wireless systems, such as IEEE 802.11a providing 54Mbps data rate, increased throughput requires further parallelization. It means more than one Processor Element needs to be assigned per column to the FFT. Parallel - pipelined FFTs are suitable for both high throughput and high power efficiency. Parallel pipelined FFTs have not significant area and it can operate at lower frequency and low power consumption. This paper discus the application of common sub expression sharing across coefficients to the second stage of 64-point or the first stage of 16-point FFTs. In the past complex multiplication shifters and adders were used and some special constant coefficients.

.

In this FFT architecture, the Complex multiplications are replaced by minimum number of shift and addition operations. Hence both area and power consumptions for the multiplier unit are reduced.

II. ALGORITHMS

A. FFT Algorithm

The Discrete Fourier Transform of N complex data point x(n) is defined by

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk} \qquad \text{k = 0, 1,……N-1;}$$

Where $W_N = e^{-j(2\pi/N)}$, WN is twiddle factor or coefficient. Previously [7] R4SDC pipelined FFT algorithm for word sequential data radix r, the equation (1) can be written as

$$X(k) = \sum\nolimits_{q1=0}^{N1-1} W_N q_1 k \sum\nolimits_{p=0}^{r_1-1} x(N_1 p + q1)\, W_{r_1}^{pk}$$

This N point can be decomposed into V stages where N=r1, r2,………..rv. The final stage is X(r1r2….rv-1mv+r1r2….rv-2mv-1+r….r1m1+m1)

$$= \sum\nolimits_{q_{v-1}}^{r_{v-1}} X_{v-1}(q_v - 1, m_v - 1) W_{rv}^{q_{-1v}mv} .$$

The intermediate stages are given by recursive equation

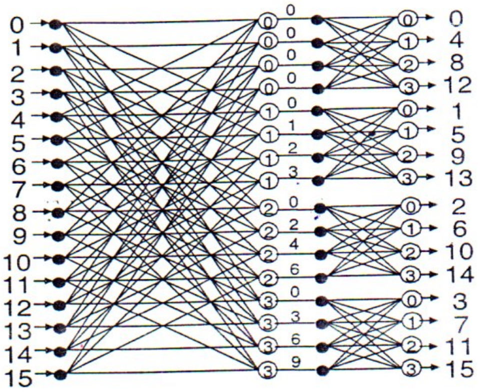
Xt (qt-mt) $= W_{N_{t-1}}^{q_t mt} \sum\nolimits_{p=0}^{r_t-1} x_{t-1}(N_t P + q_t, m_{t-1}) W_{rt}^{pmt}$

0≤qi≤ni-1, 2≤i≤v and Nt=N1(r1r2…..rv), 2≤t≤v-1,0≤mi≤ri-1. For r1=4, the 16 point FFT is shown in figure. Dots define stage borders. Open cycle denotes summations. Number outside the open circle is twiddle factor. Architecture in figure, It has 75% utilization of complex multiplier and 100% Butterfly.

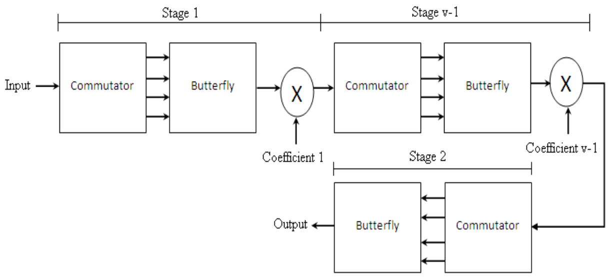
2.2 Common Subexpression Sharing

Common subexpression sharing shares the subexpression among several multiplication-accumulation operations in order to reduce the total number of operations. This approach is very effective for reducing the hardware cost of multiple constant multiplications, especially for the filter-like operation. For example, for a 3-tap FIR filter, the output Y(2) is given as follow.

$Y(2) = \sum_{i=0}^{2} A_i \times X_{n-i}$ The weights A1 are the filter coefficients. Suppose the coefficients are given as A0 = 00111011, A1= 00101011, and A2 = 10110011. The coefficients are represented in two's complement format. According to the equation Y(2) = A2 ×X2+A1×X1+A2×X0. Using shifts and additions to replace the multiplications, gives: Y(2) = X2+X2 << 1+X2 << 3+X2 << 4+X2 << 5+X1+X1 << 1+X1 << 3+X3 << 5+X0+X0<<1+X0<< 4+X0<<5-X0<< 7.

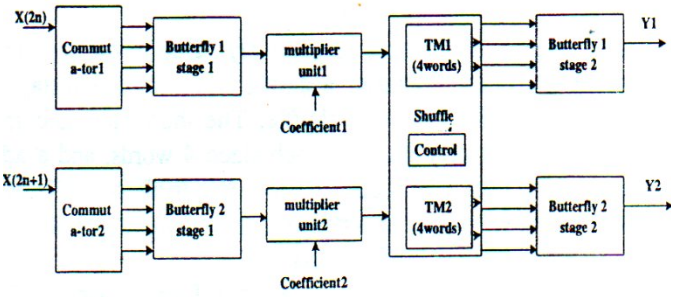


(Signal Flow Graph of a radix – 4, 16 point FFT)



(N-point radix-4 pipelined FFT processor architecture)

The computation required twelve additions, one subtraction and eleven shifts. However, if pre-computing X02 = X0+X2; X12 = X1+X2; X012 = X12+X0, the output can be shown as : Y(2) = X012+X02 << 4+X012 << 1+X012 << 5+X12 << 3- X0 << 7.



(16-point 2-parallel pipelined FFT architecture)

This computation only needs seven additions, one subtraction and five shifts. X02, D12, X012 are the common subexpression for this case.. From the above examples, it can be shown that common subexpression sharing can reduce the number of additions and subtraction from 13 to 8 (i.e. 38% reduction).

2.3 Canonic Signed Digit (CSD)

It is common to use the redundancy of signed digit code to replace the conventional multiplier digits, such that addition operations in a multiplication can be reduced with the increase of average shift length across the zeros in the multiplier. Canonical Signed-Digit (CSD) is a widely used signed digit approach. In CSD code of a number, each bit is set to 0, 1 or -1 and no two consecutive bits are nonzero. The advantages of CSD form is that no value has more than (N+1)/2 nonzero bits, often fewer, and so the multiplication by a constant requires no more than that number of additions for its implementation.

III. IMPLEMENTATION

3.1 Reduction Multiplier R4SDC FFT

In this algorithm it consists of four real multipliers, one adder and one subtractor. The complex co-efficient for all stages are precomputed. The calculated co-efficients are shown in table. For the trivial co-efficient (7FFF, 0000) and (0000, 8000), the complex multiplication is not necessary. An additional unit, which swaps the real and imaginary parts of input data and inverts the imaginary part for (0000, 8000). The rest of the co-efficients are composed of only 6 constants (7641, 5A82, 30FB, A57D, 89BE, CF04). For example, a multiplication with the constant A57d could be realized by first multiplying the data with 5A83, and then two's complementing the result.

The coefficients for 16 point R4SDC FFT Table

Coefficient sequence m1=0, 1	Original quantized coefficient	Coefficient sequence m1=2, 3	Original quantized coefficient
W0	7FFF, 0000	W0	7FFF, 0000
W0	7FFF, 0000	W2	5A82, A57D
W0	7FFF, 0000	W4	0000, 8000
W0	7FFF, 0000	W6	A57D, A57D
W0	7FFF, 0000	W0	7FFF, 0000
W1	7641, CF04	W3	30FB, 89BE
W2	5A82, A57D	W6	A57D, A57D
W3	30FB, 89BE	W9	89BE, 30FB

Note that a multiplication by the constant 5A82 already existents. Therefore, the multiplication with the constant 5A83 can simply be obtained by adding the data to the already existing multiplication with 5A82. The other two constants (89BE and CF04) can be realized in a similar manner, using constants 7641 and 30FB respectively. 5A82 is represented by two's complement format, 7641 and 30fb are represented by CSD format as follows,

5A82	0101101010000010
7641	1000-10-1001000001
30FB	010-1000100000-10-1

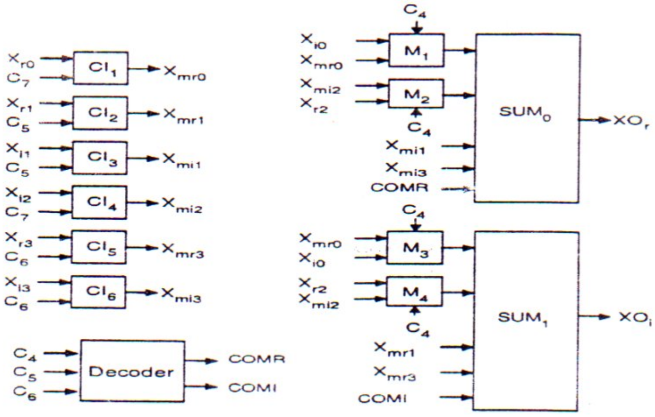
The mixed use of CSD and two's complement is for minimizing the number of addition/shift operations. We can use shifters and adders based on the three constants to carry out those nontrivial complex multiplications as shown below:

5A82X = 5X<<12+5X<<9+65X<<1
7641X = X<<15+65X-5X<<9
30FBX = 65X<<8-X<<12-5X

Where X means input data. The common subexpressions for the three constants are 101(5) and 1000001(65).

3.2 Conventional Butterfly

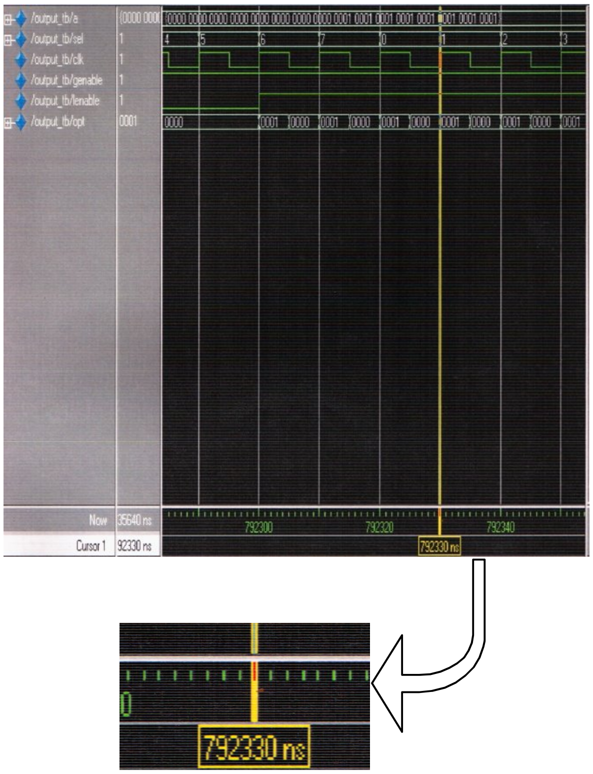
The conventional butterfly architecture consists of 6 adder / subtracters. In this paper, we proposed a low power butterfly architecture which employs two 5-input summation blocks to replace six adder / subtracters.



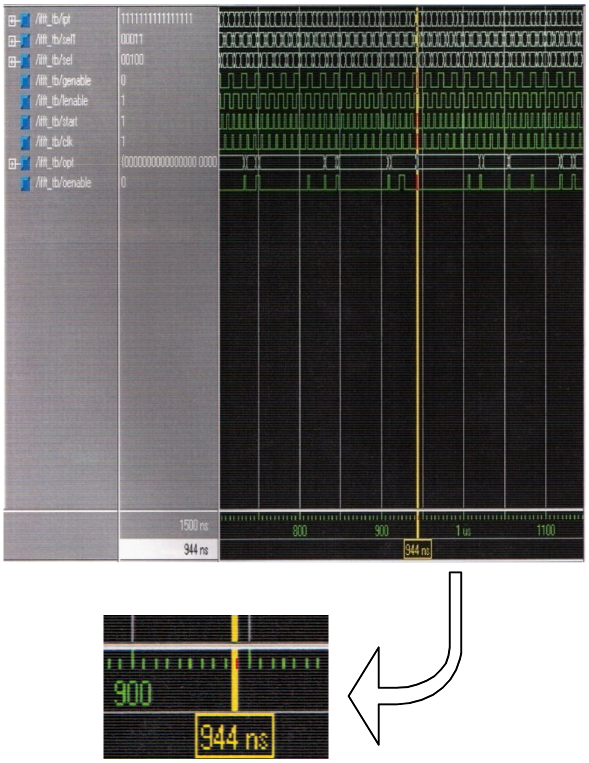
(Block Diagram for Conventional Butterfly)

Figure shows the conventional butterfly architecture. Inverters (CI1 to CI6) are used to generate the normal or the one's complement form under the control of c5, c6 and c7. The signal C4 controls the four multiplexers (M1 to M4) for directing appropriate data to the inputs of the summation blocks. Two 5-input summation blocks (SUM0 to SUM5) are employed to generate the real and imaginary parts of the output respectively. An additional decoder unit is used to generate compensation for eliminating the error which results from the one's complement inversion controllable inverters.

IV. Simulation Results



(Sımulatıon Result for FFT)



(Simulation Result for IFFT)

V. Conclusion

The Paper presents parallel pipelined architecture for 64 point FFTs. This has multiplier-less and Low power butterfly. By synthesizing the power consumption can be compared with the normal FFTs.

REFERENCES

[1] S. He and M. Torkelson, “Design and implementation of 1024-point pipeline FFT processor” Custom Integrated Circuits Conference, 1998. Processing of the IEEE 1998, 11-14 May 1998. pp. 131-134.

[2] K. Maharatna, E. Grass and U. Jagdhold “ A 64-point Fourier Transform Chip for High speed Wireless LAN Application using OFDM” IEEE Journal of Solid-State Circuit. Vol. 39, No.3, March 2004.

[3] Wei Han, T. Arslan, A.T. Erdogan and M. Hasan “ Multiplier Less based Parallel Pipelined FFT architecture for Wireless Communication Applications” in IEEE 2005, CASSP 2005, Pp V-45 – V-48.

[4] G. Bi and E.V. Jones, “A pipelined FFT processors for word-sequential data”, IEEE Transactions on acoustics, speech and signal processing, Vol.37, no:12, Dec.1989, pp.1982-1985.

[5] Wei Han, T. Arslan, A.T. Erdogan and M. Hasan “ A novel low power pipelined FFT based on subexpression sharing for wireless LAN applications”, in IEEE signal processing systems workshop, 2004. (SIPS 2004), Oct 2004,pp. 83-88,

Cooperative Black Hole Node Detection in Mobile Ad-Hoc Network

Author Name- Dipankar Chatterjee
Designation- Lecturer, Department of Computer Applications, Guru Nanak Institute of Technology
Email – dipankarchat102@yahoo.co.in

Abstract - *Wireless networks can be basically either infrastructure based networks or infrastructure less networks. The infrastructure based networks uses fixed base stations, which are responsible for coordinating communication between the mobile nodes. The ad-hoc network falls under the class of infrastructure less networks, where the mobile nodes communicate with each other without any fixed infrastructure between them. Ad-hoc network is a collection of nodes that do not rely on a predefined infrastructure to keep the network connected. So, the functioning of ad-hoc network is dependent on the trust and cooperation between nodes. Nodes help each other in conveying information about the topology of the network and share the responsibility of managing the network. In addition to acting as hosts, each mobile node does the function of routing and relaying messages for other mobile nodes. Nodes within each others radio range communicate directly via wireless links, while those that are far apart use other nodes as relays. Nodes usually share the same physical media. They transmit and acquire signals at the same frequency band. Due to their inherent characteristics of dynamic topology, lack of centralized infrastructure and dependence on other nodes for transmission, MANET is vulnerable to various kinds of attacks. One such attack is a black-hole attack. A black-hole is a malicious node that incorrectly replies the route requests that it has a fresh enough route to destination and then it drops all the receiving packets. The damage will be very serious if black-holes work together as a group.*

I: INTRODUCTION

There will be tremendous growth over the next decade in the use of wireless communication, from satellite transmission into many homes to wireless personal area networks. As the cost of wireless access drops, wireless communications could replace wired in many settings. One advantage of wireless is the ability to transmit data among users in a common area while remaining mobile. However, the distance between participants is limited by the range of transmitters or their proximity to wireless access points. Ad-hoc wireless networks mitigate this problem by allowing out of range nodes to route data through intermediate nodes. MANET or mobile ad-hoc network is a multi hop infrastructure less system comprised of many mobile nodes with wireless transmission capacity. Unlike wireless network which controlled by a fixed base station or access point, mobile ad-hoc network has no such central control. For this unique characteristic of MANET security has become one of the major concerns.

Ad-hoc networks have a wide array of military and commercial applications. Ad-hoc networks are ideal in situations where installing an infrastructure is not possible because the infrastructure is too expensive or too vulnerable, the network is too transient, or the infrastructure was destroyed. For example, nodes may be spread over too large an area for one base station and a second base station may be too expensive. An example of vulnerable infrastructure is a military base station on a battlefield. Networks for wilderness expeditions and conferences may be transient if they exist for only a short period of time before dispersing or moving. Finally if network infrastructure has been destroyed due to a disaster, ad-hoc networks could be used to coordinate relief efforts.

Due to limited memory and computational power, nodes in MANET have limited services and security provision. Unlike wired networks which have a higher level of security for gateways and routers, ad hoc networks have characteristics such as dynamically changing topology, weak physical protection of nodes, no established infrastructure or centralized administration and highly dependence on inherent node cooperation. The routing protocols used in the current generation of mobile ad hoc networks, like Dynamic Source Routing (DSR) (Johnson et al. 2001), and Ad-hoc On Demand Distance Vector Routing Protocol (AODV) (Perkins and Royer, 1999), are based on the principle that all nodes will cooperate, but dynamic and cooperative nature of MANETS presents substantial challenges to this assumption. Without node cooperation in a mobile ad-hoc network, neither routes can be established, nor can packets be forwarded. As a consequence, access control mechanisms (similar to firewalls in wired networks) are not feasible. However, cooperative behavior, such as forwarding other node's messages, cannot be taken for granted since any node could misbehave. Misbehavior means deviation from regular routing and forwarding protocol assumption. It may arise for several reasons, non-intentionally when a node is faulty or intentional when a node may want to save its resources. Cooperation in mobile ad-hoc networks is a big issue of consideration. To save battery, bandwidth, and processing power, nodes should not forward packets for others. If this dominant strategy is adopted, the outcome is a non-functional network when multi-hop routes are needed. The effects of misbehavior have been shown to dramatically decrease network performance. Depending on the proportion of misbehaving nodes and their strategies, network throughput could decrease, and there could be packet losses, denial of service or network portioning. These detrimental effects of misbehavior can endanger the entire network.

Wireless ad-hoc networks are vulnerable to various attacks. These include passive eavesdropping, active interfering, impersonation, modification of packets
and denial-of-service. Intrusion prevention measures, such as strong authentication and redundant transmission can be used to tackle some of these attacks. However, these techniques can address only a subset of the threats, and moreover, are costly to implement due to the limited memory and computation power of nodes. We can identify two types of uncooperative nodes: faulty or malicious and selfish. Faulty or malicious behavior refers to the broad class of misbehavior in which nodes are either faulty and can therefore not follow a protocol, or are intentionally malicious and try to attack the system. Selfishness refers to no cooperation in certain network operations. In mobile ad-hoc networks, the main threat from selfish nodes is dropping of packets (black hole), which may affect the performance of the network severely.

II. AODV Routing Protocol

In Ad-Hoc On-Demand Distance Vector routing protocol, every node maintains a routing table, which contains information about the route to a particular destination. Whenever a packet is to be sent by a node, it first checks with its routing table to determine whether a route to the destination is already available. If so, it uses the route to send the packets to the destination. If a route is not available or the previously entered route is inactivated, then the node initiates a route discovery process. A RREQ (Route Request) packet is broadcasted by the source node. Every node that receives the RREQ packet first checks if it is the destination for that packet and if so, it sends back an RREP (Route Reply) packet back to the source. If it is not the destination, then it checks with its routing table to determine if it has got a route to the destination. If not, it relays the RREQ packet by broadcasting it to its neighbors. If its routing table does contain an entry to the destination, then the next step is the comparison of the destination sequence number in its routing table to that present in the RREQ packet. This destination sequence number is the sequence number of the last sent packet from the destination to the source. If the destination sequence number present in the routing table is lesser than or equal to the one contained in the RREQ packet, then the node relays RREQ further to its neighbors. If the destination sequence number in the routing table is higher than the destination sequence number contained in the RREQ packet, it denotes that the route is a fresh route and packets can be sent through this route. This intermediate node than sends a RREP packet to the node through which it received the RREQ packet. The RREP packet gets relayed back to the source through the reverse route. The source node than updates it's routing table and sends its packets through this route. During the operation if any node identifies a link failure it sends a RERR (Route Error) packet to all other nodes that uses this link for their communication with other nodes.

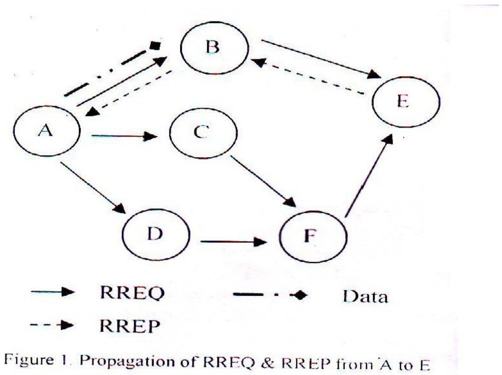


Figure 1. Propagation of RREQ & RREP from A to E

III. BLACK HOLE ATTACK

A Black-hole attack is one kind of **denial of service** attack where a malicious node can attract all packets by falsely claiming a fresh route to the destination and then absorb them without forwarding them to the destination. Cooperative black-hole means the malicious nodes act in a group. As an example, consider the scenario in the following figure. Here, node S is the source node and D is the destination node. Node 1 to 5 act as the intermediate nodes. Node 4 and 5 act as the cooperative black-holes. When the source node wishes to transmit a data packet to the destination, it first sends out the RREQ packet to the neighboring nodes. The malicious nodes being part of the network, also receive the RREQ. Since the black-hole nodes have the characteristic of responding first to any RREQ, it immediately sends out the RREP. The RREP from the black-hole node 4 reaches the source node well ahead of the other RREPs, as it can be seen in the following figure. Now on receiving the RREP from node 4, the source starts transmitting the data packets. On the receipt of data packets, black-hole node 4 simply drops them instead of forwarding to the destination D or node 4 forwards all the data to black-hole node 5. Node 5 simply drops it instead of forwarding to the destination. Hence, the data packets get lost and never reach to the intended destination.

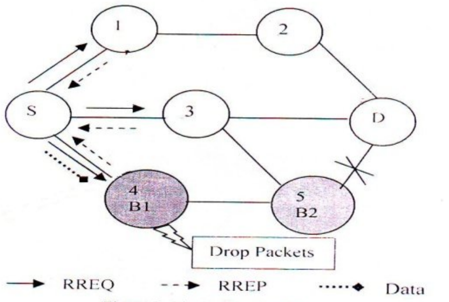


Figure 2. Black Hole Attack

IV. REVIEW OF PAPERS ON BLACK HOLE NODE DETECTION

Over the past few years, a lot of research has focused on the black-hole node detection in MANET. Several related works are briefly presented here.

Deng, Li and Agarwal [1] proposed a solution for single black-hole node detection. In the proposed method, each intermediate node, which has a fresh enough route to the destination, send back the next hop information when it sends back an RREP packet.

The source node when receives the reply message, it does not send data packets right away, but extracts the next hop information from the reply packet and then sends a further request to the next hop via a different route other than through the intermediate node to verify that the next hop has a route to the intermediate node and it has a route to the destination node. The next hop sends back further reply message to the source node.

If both the routes are available then the source node transmits data packets, otherwise the source node does not send data packets and detect the intermediate node as malicious.

This method can only detect single black-hole node, but if the intermediate node and its next hop both are malicious, then this method does not work because the next hop then falsely claim both the routes are available when it has no route to the destination and the source node send data packets towards malicious node. Another drawback is that the source node needs to communicate with the next hop node of the intermediate node via a route other than trough the intermediate node. If the cache table of the source node does not have such a root, then the source have to find an alternative path to reach next hop node of the intermediate node. That means then source must initiate a root discovery process to discover a root. This process yields overheads such as increasing network traffic (due to broadcast of RREQ and unicast of RREP) and time delay. Also a RREP may come from a black-hole node (if there exists another black-hole node in the network) and we must detect it as stated previously. So, the procedure then becomes recursive.

S.Ramaswamy, H.Fu, M.Sreekantaradhya, J.Dixon and K.Nygard [2] proposed a method for identifying multiple black-hole nodes. The methodology works with slightly modified AODV protocol by introducing Data Routing Information (DRI) table and cross checking. Every node maintains this table. DRI table contains [Node id, from, through] columns. From stands for whether the node route data packets from the node in the node field or not. Through stands for whether the node route data packets through the node in the node field or not. When an intermediate node (IN) replies a RREP to a given source node, the next hop node and DRI entry of next hop node (NHN) should also be sent together. The source node will then use the information together with its own DRI table to check whether the intermediate node is reliable node or not. If SN has used IN before to route data, then IN is a reliable node and SN starts transmitting data packets through this route. If it is not reliable, then it sends a further request packet to NHN and asks NHN: (1) if IN has routed data packets through NHN, (2) who is the current NHN's next hop to destination and (3) has the current NHN routed data through its own next hop.

The NHN in turn responds with further reply message including (1) DRI entry for IN,
(2) the next hop of current NHN and (3) the DRI entry for the current NHN's next hop.

Based on the further reply message from NHN, source node checks whether NHN is a reliable node or not. If source node has routed data through NHN before, NHN is reliable, otherwise unreliable. If NHN is reliable, source node will check whether IN is a black-hole or not. If the second bit of the DRI entry of IN (IN has routed data from NHN) is equal to 1 and the first bit of the DRI entry of IN (NHN has routed data from IN) is equal to 0, IN is a black-hole node. If IN is not a black-hole and NHN is a reliable node, the route is secure, and the source node will update its DRI entry for IN with 01 and starts routing data via IN. If IN is a black-hole, source node does not send data packets by this route, ignores any other RREP coming from IN from now and broadcasts an alarm message that contains the black-hole node id to the network. If NHN is an unreliable node, source node treats current NHN as IN and sends further request message to the updated IN's next hop node and by looking the further reply message from updated IN's next hop(which we assume reliable), we decide whether both IN(before update) and its next hop(NHN) are consecutive cooperative black-hole or not. Maintaining DRI table (which contains node_id, from, trough records of all the nodes in the network) for each node increase overhead. Another drawback is that the source node needs to communicate with the next hop node of the intermediate node and if next hop node is unreliable then the source node needs to communicate with the next hop node of the current next hop node via a route other than trough the intermediate node. If the cache table of the source node does not have such roots, then the source have to find an alternative path to reach next hop node of the intermediate node (or even next hop node of the current next hop node if current next hop node is unreliable). That means then source must initiate root discovery processes to discover such roots. This process yields overheads such as increasing network traffic (due to broadcast of RREQ and unicast of RREP) and time delay. Also a RREP may come from a black-hole node and we must detect it as stated previously. So, the procedure then becomes recursive.

Mohammad Ai-Shurman, Seong-Moo-Yoo and Seungjin Park [3] propose the following two approaches to solve the black-hole attack problem. The first solution is to find more than one route to the destination (redundant routes, at least three different routes). Then, the source node unicasts a ping packet to the destination using these three routes (we should assign different packet Ids, so any node who receives the first packet will not drop the second one if it exists in both paths). The receiver and the malicious in addition to any intermediate node might have a route to the destination will reply to this ping request. The source will check those acknowledgements, and process them in order to figure out which one is not safe and might have the black-hole node. This method only identifies the route in which a black-hole node present but failed to detect the black-hole node. So, all the nodes that are part of this route cannot be used by the source (Because source does not know which one (or more) is a black-hole.). You can say that the one from which the route reply comes is the black-hole. (Because the inherent nature of black-hole is to send a route reply with a higher destination sequence number as early as possible against each RREQ received). But you can not guarantee that the node in question is a black-hole node (Because due to network traffic or some other reasons either the ping packet or the ACK packet may be lost). You can only say that this node is suspicious. Another drawback is the time delay.

The second solution proposed by Mohammad Ai-Shurman, Seong-Moo-Yoo and Seungjin Park [3] to solve the black-hole attack problem is as follows. Every node stores the last sent packet sequence number for every node and last received packet sequence number for every node it two separate tables. The sender broadcasts the RREQ packet to its neighbors. Once this RREQ reach the destination or reach to a intermediate node which has a fresh enough route to a destination, it will reply to the sender with a RREP contains the last packet sequence number received from the source. When a source node receives a RREP from another node, it checks the last sent packet sequence number and received packet sequence number, if there is any mismatch then it generates an alarm indicating the existence of a black-hole node. If the black-hole node uses the same sequence number that the source node matches with, then source node can not detect the black-hole and sent data packets towards black-hole. Another drawback is the control packet overhead (Every RREP packet must also contain an extra header which stores last received packet sequence number from the source).

Chang Wu Yu, Tung-Kuang Wu, Rei Heng Cheng and Shun Chao Chang [4] proposed a distributed and cooperative procedure to detect black-hole node. First each node detects local anomalies, then after finding the local anomalies the sender node calls for a cooperative detection by sending a message to the 1 hop neighbors of the infected node. In local data collection, each node collects information through overhearing packets to evaluate if there is any suspicious node in the neighborhood. If finding one, the detecting node would initiate the local detection procedure to analyze whether the suspicious node is a malicious black-hole node. Subsequently, the cooperative detection procedure is initiated by the initial detection node, which proceeds by first broadcasting and notifying all 1hop neighbors of the possible malicious node to cooperatively participate in the decision process confirming that the node in question is indeed a malicious one. They use a voting scheme to identify the black-hole node. If all the nodes vote for the infected node, then the node is declared as black-hole node. As soon as a confirmed black-hole node is identified, the global reaction is activated immediately to establish a proper notification system to send warnings to the whole network. Due to cooperative global detection mechanism and broadcasting nature, network traffic is increased enormously. It cannot detect cooperative group black-hole attack (if one or more of the 1hop neighbors of the malicious node is malicious). Another drawback is that the voting scheme is not good.

Payal N.Raj and P.B. Swadas [5] proposes the following solution for black-hole detection. In normal AODV, the node that receives the RREP packet first checks the value of sequence number in its routing table. The RREP packet is accepted if it has RREP sequence number higher than the one in routing table. This solution does an additional checking to find whether the RREP sequence number is higher than some threshold value. The threshold value is dynamically updated in every time interval. As the RREP sequence number is found to be higher than the threshold value, the node is suspected to be malicious and it adds the node to the black-hole list. As the node detected a malicious node, it sends an ALARM to its neighbors and they will not use this malicious node from now then on to send packets and each of the neighbors also broadcast ALARM message to their neighbors and this process continues until all the nodes of the network blacklist the malicious node. The threshold value is the average of the difference of destination sequence number in each time slot between the sequence number in the routing table and the RREP packet. The threshold value is dynamically updated using the data collected in the time interval. As a new node receives a RREP packet for the first time, it gets the updated value of the threshold. There may arise some nodes which are malicious and they may select a destination sequence number in the RREP packet which is less then the threshold value and so source node does not detect that malicious node and send data packets towards it (False negative).There may also arise some nodes which are not malicious and they put the actual destination sequence number in the RREP packet and the number happens to be more than the threshold value, then source node will detect it as malicious node and does not send data packets towards it (False positive).

V. OUR PROPOSAL FOR BLACK-HOLE NODE DETECTION AND PREVENTION FROM BLACK-HOLE ATTACK.

We propose a solution for black-hole node detection using the concept of threshold value [5] [9] and using the concept of reputation [7] [8] and weight of a node. This solution can detect single as well as cooperative black-hole node with less network overhead compared to paper [2] and reduce the number of false positive and false negative compared to paper [5].

Our proposed solution is divided into four steps:

Step 1 :

Creating and Populating Reputation table:

We maintain a reputation table that contains the following fields: node_id, packet_forwarded, packet_received, weight and reputation. We give the description of each field below:

(i) node_id is the id of the node whose information we are storing in the row.

(ii) Packet_forwarded is the number of data packet forwarded by this node.

(iii) Packet_received is the number of data packets received by this node.

(iv) Weight is an integer value that depends on the ratio of packet_transmitted / packet_received and weight is calculated by the following formula

Weight (of a node) = (packet_forwarded /packet_received) × 10

[If the value is float then up to .5 choose the previous integer value otherwise choose the next integer value].

(v) Reputation is of three types: (i) reliable, (ii) suspicious and (iii) malicious.

We calculate the reputation by the following way:

weight>=8 then reputation is reliable

4<=weight<=7 then reputation is suspicious

Weight<4 then reputation is malicious.

In the reputation table we maintain all the above information for each node in the network. The reputation table is shared by all the nodes in the network.

<u>Step 2 :</u>

Choosing a route:

The source node transmits the RREQ to all its neighbors. Then the source waits for a "timer" seconds to collect the RREP's. All the RREP's are collected in a table called Reply Collection table until the timer will go off. Each route reply must contain the following information: source address, destination address, hop count, next hop and destination sequence number. This information is stored in the Reply Collection table for each RREP. A reply is chosen by the source to send data packets based on the following criteria: In each of the received RREP, the reputation of the responding node (from where the RREP generates) and its next hop (along the path) are checked. If both the reputation values are reliable, then take average of the weight values of both the nodes. If there are more than one such root replies available where the responding node and its next hop both are reliable, then choose the route with the highest average weight value. [If the average weight values are same for more than one route and this value happens to be the highest then choose the route with the least hop count]. If at least one of the reputation values of the responding node and its next hop is unreliable, then the source node will not send data packets along the route and buffer its packets until a safe route is discovered.

<u>Step 3 :</u>

Updating the Reputation table:

After choosing the right route, send data packets along the route. Every destination node sends back an acknowledgement to the source node along the reverse path upon the reception of data packets. The receipt of the acknowledgement enables the source node to increment the weight of all the intermediate nodes along the path by one in the reputation table, for it has proved to be reliable and safe. [Because a safe route means every node along the route is reliable]. In case the source node does not receive the acknowledgement within a "timer" event, the source node will decrement the weight of the responding node that replied and also the weight of the next hop node of the responding node by one in the reputation table because they may be cooperative black-hole node. [Because if only one node (single black-hole) or two (consecutive cooperative black-hole) nodes is malicious then acknowledgement is not received and the path become unsafe.

Step 4 :

Detecting black-hole and generating alarm message:

The reputation table is periodically checked to identify black-hole. When the weight of a node drops below 4 (reputation becomes malicious) in the reputation table, it implies it has not forwarded data packets faithfully for a period of time and hence a black-hole. The detection of a black-hole has to be intimated to other participating nodes in the network. This is accomplished by sending alarm packet that contains the black-hole node id. When a node receives an alarm packet, it will identify the black-hole and so can eliminate the use of that node from then on.

Algorithm for the proposed solution

Notations used :

TIMER_TIME – Time of the timer clock
RREQ – Route Request
RREP_COLLECT_TIME – Time for which route replies are collected
RCT – Reply Collection Table **RACK** – Route Acknowledgement
AVG_WT – Dynamic array that contains average weight values for each route
ACK_TIMEOUT – Time for which a node waits for ACK
IN – Intermediate Node **NHN –** Next Hop Node
REP_TIME – Time interval after which reputation table must be checked.

Algorithm :

```
Set TIMER_TIME to zero.
    While ( TIMER _TIME < RREP_COLLECT_TIME )
        Store RREP's in RCT.
  If ( size of RCT = = 0 )
        Retransmit RREQ.
Else
    {
        Store the value of number of RREP's in N.
        For (I=0; I < N; I++)   / For each route reply /
            {
                If ( Reputation (IN) = = Reliable and Reputation (NHN) = = Reliable )
                        AVG_WT [I] = Weight (IN) + Weight (NHN).
                Else
                        AVG_WT [I] = 0.
            }
        If ( all the entries of AVG_WT are zero )
                No valid route is found.
        Else if ( AVG_WT contains one non zero entry )
                Send data packets by the route corresponding to that entry.
        Else if ( AVG_WT contains more than one non zero entry )
```

```
                    Send data packets by the route
corresponding to the highest AVG_WT
                    entry value.

        }


Again Set TIMER_TIME to zero.

   While  ( TIMER_TIME < ACK_TIMEOUT )

        {

             If  ( RACK is received )

                    Increment the weight values of all
the intermediate nodes along the path
                    By 1 in the reputation table.

        }

    If  ( no RACK is received )

          Decrement the weight values of IN and
NHN by 1 in the reputation table.

Check reputation table after every REP_TIME.

      If  ( Weight of a node < 4)

            Detect the node as black-hole and
broadcast alarm packets.
```

VI. CONCLUSION

In this project the routing security issues of MANETs are discussed. One type of attack, the black-hole, which can be easily be deployed against the MANET is described. I have studied and reviewed some of the works that attempts to detect single and simultaneous cooperative black-holes. in my third proposal, I proposed a feasible solution for detection of simultaneous cooperative black-holes and prevention from black-hole attack in AODV protocol using the concept of threshold and trust/reputation of a node. As future work, I intend to develop simulations to analyze the performance of my proposed solution. Also, as future work, I will try to enhance my solution further so that they can detect grey-holes (partial packet dropping) as well as black-holes (full packet dropping).

REFERENCES

[1] "Routing Security in Wireless Ad-hoc Network" by Hongmei Deng, Wei Li and Dharma P. Agrawal, University of Cincinnati, 2002.

[2] "Prevention of Cooperative Black Hole Attack in Wireless Ad Hoc Networks" by Sanjay Ramaswamy, Huirong Fu, Manohar Sreekantaradhya, John Dixon and Kendall Nygard, 2003.

[3] "Black Hole Attacks in Mobile Ad Hoc Networks" by Mohammad AL- Shurman, Seon-Moo Yoo and Seungiin Park, ACMSE, April 2-3, 2004.

[4] "A Distributed and Cooperative Black Hole Node Detection and Elimination Mechanism for Ad Hoc Network" by Chang Wu Yu, Tung-Kuang Wu, Rei Heng Cheng and Shun Chao Chang, Springer 2007.

[5] "A Dynamic Learning System Against Blackhole Attack in AODV Based MANET" by Payal N Raj and Prashant B. Swadas, IJCSI, Vol 2, 2009.

[6] " Mitigating Routing Misbehavior in Mobile Ad-hoc Networks" by Sergio Marti, T.J.Giuli, Kevin Lai and Mary Baker, Department of Computer Science, Standford University.

[7]] "Performance Analysis of the CONFIDANT Protocol (Cooperation of Nodes: Fairness in Dynamic Ad-hoc Networks" by Sonja Buchegger and Jean-Yves Le Boudec, ACMSE, 2002.

[8] "Core: A collaborative reputation mechanism to enforce node cooperation in mobile ad-hoc network" by P.Michiardi and R.Molva , 6th IFIP communications and multimedia security conference, September 2002.

[9] "Observation based cooperation enforcement in ad-hoc networks" by S.Bansal and M.Baker, July 2003.

[10] "A Cooperative Blackhole Node Detection Mechanism for ADHOC Networks", Moumita Deb, WCECS 2008.

[11] "Enhanced Intrusion Detection System for Discovering Malicious Nodes in Mobile Ad-hoc Networks" by Nidal Nasser and Yunfeng Chen, IEEE 2007.

[12] " Mitigating Routing Misbehavior in Mobile Ad-hoc Networks" by Sergio Marti, T.J.Giuli, Kevin Lai and Mary Baker, Department of Computer Science, Standford University.

MULTI-HOP CO-OPERATIVE WIRELESS SENSOR NETWORKS

D.Praveen Kumar[1], K.Swapna Reddy[2], P.Krishna[3] and M.A.H.Shamshi [4]

[1,2,3,4] ECE Dept.Vaagdevi College of Engineering, Jawaharlal Nehru Technological University, Hyderabad.

praveenandsuma@gmail.com, +91-9985822005[1], swapna409@gmail.com, +91-9491319626[2], kpatteti@gmail.com, +91-996330906[3] and himayaht@yahoo.co.in, +91-9866013800[4]

***Abstract* - We develop and analyze Distributive Space–time coded Multi-Hop diversity protocols for combating multipath fading across multiple protocol layers in a wireless Sensor networks. The protocols exploit spatial diversity available among a collection of distributed terminals that relay messages for one another in such a manner that the destination terminal can average the fading. As opposed to conventional channel coding schemes, distributed coding constructs the whole codeword in a distributed manner among the sensors. The aim of this article is to present an overview of recent development in distributed coding design in cooperative Multi-Hop wireless sensor networks.**

***Index Terms*— Spatial diversity, fading channels, Distributive Space–time coding, Multipath fading, Distributive Low Density Parity Coding.**

I. INTRODUCTION

Wireless sensor networks are dragging immense interest in modern communication architectures. They are always evolving research topics to fulfill the requirements of higher data rates in compromise of low implementation cost. In Multi-hop sensor networks, each node acts as both a source and a relay. That is, each node not only transmits its own information but also helps other nodes to transmit signals. In relay networks, the relay nodes are explicitly built nodes only for the purpose of relaying and forwarding information. They do not have their own information to transmit. Despite these differences, as far as signal processing at relays is concerned, they are almost the same. This architecture reduces the use of multiple antennas at the source and destination, but the architecture resembles as virtual MIMO.

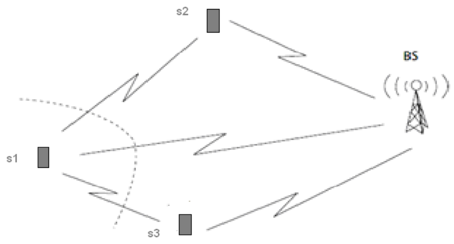


Figure1: Multi-Hop communication system

The Fig.1 depicts a model of Multi-hop network with single relay stage. Due to the practical constraints in hardware implementation, it is usually assumed that each node cannot transmit and receive at the same time. The overall transmission can be divided into two phases. The source node first broadcasts its messages to both relays and destination. Upon receiving signals from the source, each relay processes the received signals and then forwards them to the destination.

The further paper is organized as follows: the next section provides details with System modeling. Section III, explains about the different types of Relay protocols. Section IV, provides an analysis of Multi-hop networks with Distributed coding. Section V, we conclude.

II. SYSTEM MODELLING

We consider a wireless network with a set of transmitting terminals denoted M={1,2,…,m}. Each transmitting source terminal s $\in M$ has information to transmit to a single destination terminal, denoted d(s)$\notin M$, potentially using terminals M-d(s) as relays. Thus, there are 'm' co-operating terminals communicating to d(s). For algorithms in which we require the relays to fully decode the source message, we define the decoding set D(s) to be the set of relays that can decode the message of source s. In the case of amplify-and-forward cooperative diversity, we take D(s)=M-{s}.

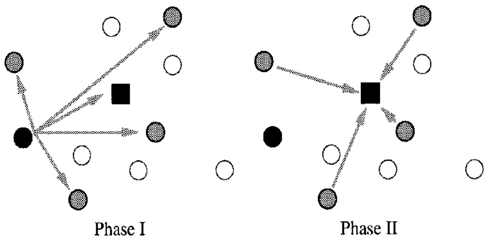
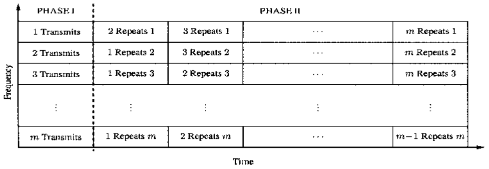


Figure 2: Illustration of the two phases of repetition-based and space–time-coded cooperative diversity algorithms.

Both classes of algorithms consist of two transmission phases, as in [1-2] Fig. 2 illustrates these two phases, and allows us to point out the similarities and differences between the algorithms. In the first phase, the source broadcasts to its destination and all potential relays. During the second phase of the algorithms, the other terminals relay to the destination, either on orthogonal subchannels in the case of repetition-based cooperative diversity, or simultaneously on the same subchannel in the case of space–time-coded cooperative diversity.

A silent source in the second phase, orthogonal relays can transmit through orthogonal channels or nonorthogonal channels. In the prior transmissions, all relays transmit to the destination through orthogonal channels, such that the transmitted signals from each relay can be separated at the destination, without any interference from other relays. For non-orthogonal transmissions, relays transmit to the destination at the same time and same frequency. Therefore the received signal at the destination is the superposition of signals transmitted from the source and relays.

Fig. 3 illustrates example channel and sub channel allocations for repetition-based cooperative diversity, in which relays either amplify what they receive or fully decode and repeat the source signal, as in [3].



PHASE I	PHASE II			
1 Transmits	2 Repeats 1	3 Repeats 1	···	m Repeats 1
2 Transmits	1 Repeats 2	3 Repeats 2	···	m Repeats 2
3 Transmits	1 Repeats 3	2 Repeats 3	···	m Repeats 3
⋮	⋮	⋮	⋮	⋮
m Transmits	1 Repeats m	2 Repeats m	···	$m-1$ Repeats m

Figure 3: Repetition-based medium-access control

In order for the destination to combine these signals and achieve diversity gains, the repetitions must occur on essentially orthogonal subchannels. For simplicity, Fig. 3 shows channel allocations for different source terminals across frequency, and subchannel allocations for different relays across time. At the destination, depending on the transmission mode and relay protocols, various processing methods will be used to recover the source messages based on the received packets transmitted from the source and relays, such as linear combinations and iterative decoding. Since signals received at the destination arrive via different paths from the source and various relays, spatial diversity can be achieved at the destination. Based on operations at the relays, there are several commonly used relay protocols.

III. RELAY PROTOCOLS

A. AMPLIFY AND FORWARD

Amplify and Forward (AAF) is one of the simplest relay protocols [3]. In AAF, upon receiving signals from the source, each relay just simply forwards to the destination a scaled version of the received signals, including both information and noise. By properly combining received signals from the source and relays, the destination node makes a final decision. Since the destination receives multiple copies of signals transmitted from the source and relays through multiple independent paths, spatial diversity can always be achieved by the AAF protocol at high signal-to- noise ratios (SNRs). Obviously the major drawback of AAF protocols is noise amplification at the relays. The outage probability and end-to-end bit error rate (BER) performance of AAF have been widely investigated [3]. Given a fixed transmission power at the source, relay, and destination, it is shown that the performance of AAF depends on the position of the relay relative to the source and destination. When the relay is positioned midway between the source and destination, AAF achieves its optimum performance and worsens as the relay moves closer or further to the source [4].

B. DECODE AND FORWARD

Decode and Forward (DAF) is another commonly used protocol for eliminating the noise effect, especially for coded systems. The relay decodes the received signals and re-encodes them before forwarding to the destination. When the channel quality in the link between the source and relay is good, the process of decoding and re-encoding provides more powerful error correcting capabilities than DemAF. Thus, the method can considerably outperform both AAF and DemAF. However, when the link from the source to the relay suffers from deep fading, decoding errors may occur at the relay. In this case, if the relay re-encodes these incorrect bits, error propagation will occur and lead to even worse performance.

C. ADAPTIVE RELAY PROTOCOL

So far, we have seen that different protocols mentioned above all have their advantages as well as disadvantages. Intuitively, we would like to ask one question: are there any possible protocols, which can not only effectively mitigate the noise amplification, but also avoid error propagations? Adaptive Relay Protocol (ARP) is one of the protocols developed to meet this need [5]. ARP has advantages of both AAF and DAF and minimizes their negative effects at the same time. In ARP, each relay adaptively selects the AAF or DAF protocol based on whether its decoding result is correct or not. All the relays that fail to decode correctly use the AAF protocol to amplify the received signals and forward them to the destination. On the other hand, all the relays that can successively decode the received signals use the DAF protocol. The signals received at the destination, forwarded from all relays by using either AAF or DAF protocol, are combined into one signal to recover the source information. It has been shown in [5] that ARP considerably outperforms the AAF scheme and simultaneously avoids error propagation due to the imperfect decoding at relays in a DAF protocol. Thus, it outperforms both AAF and DAF protocols. The performance gain grows as the number of relays increases, and it approaches the perfect DAF scheme at high SNRs. Recently, some threshold-based ARP protocols have also been developed [6]. In such protocols the relay adaptively selects the relay protocols by comparing the received SNR to a threshold. If the SNR is lower than the threshold, the relay uses the AAF protocol. Otherwise, the relay switches to the DAF protocol.

IV. DISTRIBUTED CODING STRUCTURE

The performance of relayed transmission can be further optimized if joint signal design and coding are performed at the source and relays. We refer to such a coding scheme as distributed coding. The major difference between distributed coding and conventional channel coding schemes is that in distributed coding, the overall codeword is constructed in a distributed manner. That is, different parts of the codeword in distributed coding are transmitted by different nodes through independent wireless links. This creates additional degrees of freedom, but also poses challenges in code construction. Although we can directly apply the concepts of conventional channel coding to construct distributed coding in wireless relay networks, some practical issues in designing these distributed coding schemes have to be taken into account, such as decoding errors at relays, channel variations in different parts of codeword, and rate and power allocations at the source and relays. In this section we present an overview of various distributed coding structures that have been successively developed over the past several years for wireless relay networks.

A. DISTRIBUTED SPACE TIME CODING

Let us consider a wireless relay network in which the source and relays cooperatively communicate with a common destination. This cooperative transmission among the source and relays forms a virtual antenna array. Therefore, conventional space-time coding schemes can be applied to relay networks for achieving the cooperative diversity and coding gain. Two types of distributed space -time coding (D-STC) schemes have been developed, including distributed space-time block codes (DSTBCs) and distributed spacetime trellis codes (DSTTCs).

Several DSTBC schemes have been proposed. A simple DSTBC scheme was proposed by Laneman based on orthogonal STBCs. In such a scheme different relays transmit different columns of the STBC code matrix, and at the destination a DSTBC codeword is formed. Since the orthogonal STBC does not always exist for any number of antennas, such a DSTBC scheme has certain requirements on the number of active relays. The design of such a DSTBC becomes especially difficult for a large network with a large number of relays. To solve this problem, several solutions have been proposed. In [7] proposed a DSTBC scheme, which selects a subset of nodes for transmission, and each active node transmits a linear transformation of DSTBC codewords. The transformation is unique for each relay and is represented by a signature vector. It has been shown that by proper code design, DSTBC can achieve a diversity order equal to the number of active relay nodes. It is well known that STTCs can provide a higher coding gain than STBCs in MIMO systems. It is therefore natural to consider DSTTC in wireless relay networks. Figure 4 shows the system structure of a DSTTC. In the DSTTC the overall transmission of a codeword is divided into two time slots. In the first time slot, the source encodes the information symbols by using encoder **A**1 to generate a codeword **X***s*1 and broadcasts it to both the relay and destination. Upon receiving the signals transmitted from the source, the relay calculates another codeword **X***r*1. Here **X***r*1 could be the estimated version of the signals transmitted by the source or a new codeword generated by another encoder **B**. Simultaneously, the source encodes the same information symbols by using the encoder **A**2 and generates another codeword **X***s*2. In the second time slot, the source and relay transmit **X***s*2 and **X***r*1 to the destination. The total received signals at the destination in these two time slots form a DSTTC codeword **X**, given by

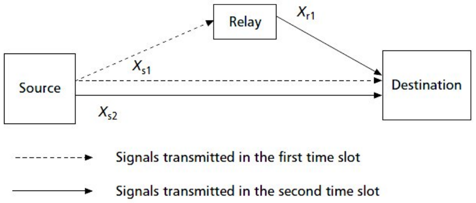

Relay
X_{r1}
Source
X_{s1}
Destination
X_{s2}
Signals transmitted in the first time slot
Signals transmitted in the second time slot

Figure4: Distributed space time coding scheme

$$X = \begin{bmatrix} X_{s1} & X_{s2} \\ 0 & X_{r1} \end{bmatrix}$$

We should note that various relay protocols can be used to construct DSTTCs, such as AAF, DemAF, or DAF. When DAF or DemAF is used in DSTTCs, detection or decoding errors have to be taken into account in the code design. Most papers usually assume that the relay can decode correctly, and thus the construction of DSTTC can be done in a similar way as in conventional STTC schemes. In order to achieve optimum performance in DSTTCs, encoders **A**1, **B**, and **A**2 should be jointly optimized. It has been shown that the code construction rules of DSTTC still follow the rank and determinant design criteria of conventional STTC [9], but the codeword difference matrix between two code words **X** and $\hat{X}$ here is an extended matrix, given by

$$B(X,\hat{X}) = X - \hat{X} = \begin{bmatrix} X_{S1} - \hat{X}_{S1} & X_{S2} - \hat{X}_{S2} \\ 0 & X_{r1} - \hat{X}_{r1} \end{bmatrix}$$

In [8] a practical DSTTC scheme, which takes into consideration detection errors at relays in the design of DSTTCs, has been proposed. In this scheme an equivalent link, representing the source-relay-destination path, has been proposed by using the equivalent SNR of the link model to take into account the detection errors at the relay [8]. It has been shown that the optimum code design still follows the rank and determinant criteria. In Fig.5, the performance of multi-hop relays are analyzed with trellis encoding over distributed code words. However, the codeword difference matrix has to be modified to take into consideration the detection errors at relays [9].

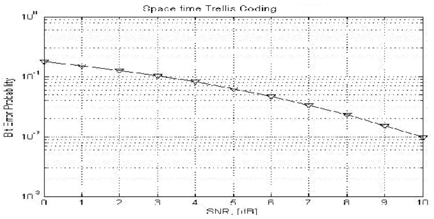


Figure5: Distributive Space time Trellis coding

B. DISTRIBUTED LOW DENSITY PARITY CHECK CODE

In DSTTC the convolutional codes are used as the constituent codes at the source and relay nodes. In order to further improve the system performance, some Distributed Low Density Parity Check (D-LDPC) coding schemes have been developed recently [10]. In D-LDPC the constituent codes at the source and relay are LDPC codes. An LDPC code with a predetermined code rate is first generated. The whole codeword consists of three parts. The first

slot. The codeword is chosen from an LDPC codebook, *CSR*1. The other two parts are chosen from two other LDPC codebooks, *CRD*2 and *CSD*2. They are transmitted by the relay and source in the second time slot. The bits chosen from *CSR*1 and *CRD*2 are transmitted by the source and relay to form another LDPC code, *CSD*1.

As a result, the design of D-LDPC codes requires joint optimization of the code profiles of *CSR*1 and *CSD*1 [10]. The density evolution (DE) has been widely used for the optimization of LDPC codes. DE can accurately track the evolution of the probability densities in a belief propagation decoding algorithm. For the conventional LDPC codes, it has been shown that good check node distributions (CNDs) are concentrated. That is, all parity check nodes should have nearly equal degrees. This property can be used to simplify DE implementation by selecting several concentrated CNDs and searching the best variable node distributions for each CND. Unfortunately, for DLDPC such an assumption is not valid, because the rates of *CSR*1 and *CSD*1 are very different [10]. This creates a significant challenge for DE implementation. Some simplified DE algorithms have been developed in [10] by using Gaussian approximation of the DE distribution and putting some constraints on the CNDs of *CSR*1 and *CSD*1 to reduce the search space. The search for good code profiles can be made using linear programming, and near optimum codes can be found by selecting the code with the optimum convergence threshold. The LDPC code has been implemented over the distributed data and the numerical results are analyzed as shown in Fig.6.It has been shown that through proper code design, an D-LDPC scheme over wireless relay channels can perform very close to the theoretical limit [10].

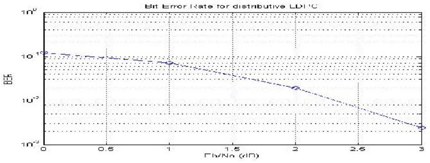


Figure 6: Distributed Low Density Parity Check Coding

V. CONCLUSION

one is transmitted by the source in the first time

In this article we have given an overview of recent research achievement in distributed coding technology. Through proper design, distributed coding can achieve both spatial diversity and a significant coding gain. Some efficient distributed coding schemes have been proposed in the past several years, but there are still many issues in both theory and practical implementation that have not been addressed. Most existing distributed coding schemes are developed based on conventional channel coding schemes, such as STC, turbo coding, and LDPC coding. Furthermore, most distributed coding schemes rely on some assumptions, such as error free decoding at relays. Some initial work on modeling detection errors in DemAF has been done by Wang *et al.* in [9]. This model has been applied to construct DSTTCs in [8]. It has been shown that detection errors do have some effects in constructing practical distributed coding. Design of a distributed coding scheme that can adaptively allocate the rate and power and distribute the code bits among the source and relays, as well as adaptively select the relay protocols, has the potential to increase overall network throughput, reduce network power consumption, and improve reliability.

REFERENCES

[1] J. N. Laneman, “Limiting analysis of outage probabilities for diversity schemes in fading channels,” in *Proc. IEEE Global Communications Conf. (GLOBECOM),* San Francisco, CA, [Online]. Available: http://www.nd.edu/~jnl/pubs/globecom2003.pdf, to be published.

[2] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, “An efficient protocol for realizing cooperative diversity in wireless networks,” in *Proc. IEEE Int. Symp. Information Theory,* Washington, DC, June 2001.

[3] J. Laneman, D. Tse, and G. Wornell, “Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior,” *IEEE Trans. Info. Theory*, vol. 50, no. 12, Dec. 2004, pp. 3062–80.

[4] F. Lin *et al.*, “Impact of Relay Location According to SER for Amplify-and- Forward Cooperative Communications,” *IEEE Int’l. Wksp. Anti-Counterfeiting, Security, Identification*, Apr. 2007, pp. 324–27.

[5] Y. Li and B. Vucetic, “On the Performance of a Simple Adaptive Relaying Protocol for Wireless Relay Networks,” *Proc VTC-Spring 2008*, Singapore, May 2008.

[6] F. A. Onat *et al.*, “Optimum Threshold for SNR-based Selective Digital Relaying Schemes in Cooperative Wireless Networks,” *Proc. IEEE WCNC*, 2007.

[7] S. Yiu, R. Schober, and L. Lampe, “Distributed Space- Time Block Coding,” *IEEE Trans. Commun.*, vol. 54, no 7, July 2006, pp. 1195–2006.

[8] J. Yuan *et al.*, “Distributed Space-Time Trellis Codes for a Cooperative System,” to appear, *IEEE Trans. Wireless Commun*.

[9] T. Wang *et al.*, “High-Performance Cooperative Demodulation with Decode-and-Forward Relays,” *IEEE Trans. Commun.*, vol. 55, no. 7, July 2007, pp. 1427–38.

[10]A. Chakrabarti *et al.*, “Low Density Parity Check Codes for the Relay Channel,” *IEEE JSAC*, vol. 25, no. 2, Feb.2007, pp. 280–91.

MULTI-USER BER PERFORMANCE ANALYSIS IN W-CDMA INDOOR APPLICATIONS

B.Rajanna[1], B.Sreedevi[2], V.Ugendar[3], J.Suman Kumar kaundinya[4]

[1,2,3,4] ECE Dept. Vaagdevi College of Engineering, JNTU Hyderabad

[1]rajannabattula@gmail.com ,+91-9866576013 , vaagvijs15@gmail.com, +91-9704024475,

ugi405@gmail.com , +91-9177553301, sumankaundinya@gmail.com,+91-9866261090

Abstract**− This paper considers the Bit Error Rate (BER) performance of the 3GPP WCDMA-FDD system with space-time transmit diversity in the indoor environments by employing a Resolution-Reduced Maximum Ratio Combing (RR-MRC) Rake receiver. An average uncoded BER lower bound is used for the BER performance analysis, which is based on a Proper Complex Gaussian (PCG) multi-channel assumption without Inter-Symbol Interference (ISI). Numerical results show that substantial diversity diversity gain can still be obtained with the suboptimal RR-MRC receiver even in a mixed-mode fading multi-path channel with the rms delay spread around 50 nanoseconds**.

Keywords- Space Time Transmit Diversity, AWGN, Multi-path fading, Rayleigh Fading, Ricean Fading.

I. INTRODUCTION

There has been a tremendous growth in wireless communication technology over the past decade. The significant increase in subscribers and traffic, new bandwidth consuming applications such as gaming, music down loading and video streaming will place new demands on capacity. The answer to the capacity demand is the provision of new technology-Wideband CDMA or hereinafter referred to as WCDMA

Direct-Sequence Wideband Code Division Multiple Access (WCDMA) has been specified by the 3rd Generation Partnership Project (3GPP) as the air interface standard for the Universal Mobile Telecommunication System (UMTS) Universal Terrestrial Radio Access (UTRA). This system is expected to provide up to 2Mbps on the downlink for the indoor high-speed applications by using an advanced channel coding technique such as Turbo coding. In addition, the twofold block-coded space-time transmit diversity (STTD) proposed by Texas Instruments [1], which is based on Alamouti's scheme [2], is adopted in the UTRA Frequency Division Duplex (FDD) mode [3], [4] to increase the system downlink throughput. Theoretically, twice the diversity order can be achieved with STTD compared to systems without antenna diversity in a flat fading channel, which is often the case for indoor channels. A portable mobile slowly moving through an indoor environment, experiences Ricean or Rayleigh fading depending on the existence of Line of Sight (LoS) conditions. The Rice factor takes values in the range of 2 to 10dB for LoS cases [7]. However, with the increased chip rate to 3.84Mcps in 3GPP UMTS, the indoor radio channels become complicated. A typical indoor channel may present richer multi-path structure than it does in a 2G system of a much lower chip rate, with the same delay spread. If the delay spread is less than the chip width (260ns for the 3GPP WCDMA FDD system), a conventional maximum-ratio- combining (MRC) Rake receiver simply sets to unity the number of branches and the multi-path diversity is not exploited. In this paper, we assume the channel power delay profile (PDP) is obtained through the wideband ray-tracing channel modeling. We exploit the multi-path diversity by using the suboptimal resolution-reduced maximum ratio combing (RR-MRC). The resolution reduction (RR) technique consists of using the Rake receiver with branches spaced less than one chip period apart. Previous reported research [5], [6] has shown that a RR to one quarter of the chip interval for the IS-95 parameters is sufficient to provide almost all the achievable improvement from the irresolvable multi-path components.

The further paper is organized as; in next section, we extend the analytical average uncoded Bit Error Rate (BER) derived in [9] to both the pure Rayleigh correlated fading and the mixed-mode Ricean/Rayleigh correlated fading multi-channels (Ricean fading on the first arriving path and Rayleigh fading on the following paths from each transmit antenna), and indicate that this can be used as a lower bound BER for the indoor channels with limited Inter-Symbol Interference (ISI). In Section III, the numerical results with a ray-tracing channel model are given for the spatial and multi-path diversity gain obtained with RR-MRC Rake receiver and the effectiveness of the lower bound. Section IV concludes the paper.

II. BER PERFORMANCE EVALUATION

A simple STTD transceiver model with RR-MRC receiver is shown in Fig. 1 at the top of next page, as used in a rich scattering environment, where the channel PDP is captured by an offline wideband ray tracer. In this paper, we assume that the multi-path channel has the same delay structure from each transmit antenna, though the corresponding multi-path component power is not necessarily the same. The channel is normalized such that the received signal power is equal to the total transmitted power. Suppose that the number of multi-path components from each transmit antenna is L, and each channel coefficient is a proper complex Gaussian (PCG) random variable. For this case, from [9], the average uncoded BER P_b is given by

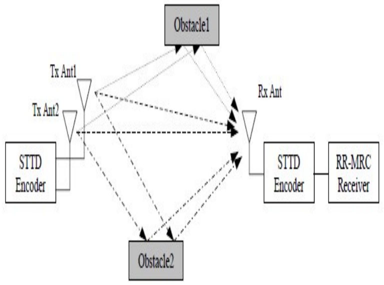


Figure 1: A Transceiver with STTD

$$\overline{P}_b = \frac{1}{\pi}\int_0^{\pi/2}\left[\det\left(\frac{C}{\sin^2\theta}+I\right)\right]^{-1}.\exp\left[-m^H\left(C+\sin^2\theta I\right)^{-1}m\right]d\theta \tag{1}$$

for the BPSK signal with the MRC Rake receiver. In (1), **C** and **m** represent the complex multi-channel covariance matrix and the mean vector of the channel coefficients, respectively. **I** is the identity matrix. The superscript *H* denotes Hermitian operation. The total number of multi-channels from the two transmit antennas is 2L, which is the dimension of the square matrix **C** and the column vector **m**.

A. PURE RAYLEIGH FADING

In the absence of LoS components ,the multi-channel s are pure RAYLEIGH faded.In this case m=O^T and (1) can be simplified to

$$\overline{P_b} = \frac{1}{\pi}\int_0^{\pi/2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{p_{i,j}}{\left(1+\dfrac{\lambda_i}{\sin^2\theta}\right)^j}d\theta$$

$$=\sum_{i=0}^{N}\sum_{j=0}^{n_i}p_{i,j}\frac{1}{\pi}\int_0^{\pi/2}\frac{1}{\left(1+\dfrac{\lambda_i}{\sin^2\theta}\right)^j}d\theta \tag{2}$$

Where $\lambda_1,\ldots\ldots\lambda_N$ are the distinct eigen values of the covariance matrix C each with a multiplicity of $n_1,\ldots\ldots n_N$ respectively, which are subjected to the constraint

$$\sum_{i=0}^{N} n_i = 2L \tag{3}$$

The coefficients P$_{ij}$ are partial fraction coefficient given by

$$\prod_{i=1}^{N}\left(1+\frac{\lambda_i}{\sin^2\theta}\right)^{-n_i} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{p_{i,j}}{\left(1+\dfrac{\lambda_i}{\sin^2\theta}\right)^j} \tag{4}$$

According to the Equation (10)in (9)the integral in (2) has a closed form solution , therefore $\overline{P_b}$ can be evaluated as

$$\overline{P}_b = \sum_{i=0}^{N}\sum_{j=0}^{n_i}p_{i,j}\left[P(\lambda_i)\right]^j\sum_{k=0}^{j-1}\binom{j-1+k}{k}\left[1-P(\lambda_i)\right]^k \tag{5}$$

Where

$$P(\lambda_i)=\frac{1}{2}\left[1-\sqrt{\frac{\lambda_i}{1+\lambda_i}}\right] \tag{6}$$

For the case of N=2, P$_{i,j}$ can be calculated from

$$p_{i,n_i-q} = \frac{\binom{n_j+q-1}{q}\left(-\dfrac{\lambda_j}{\lambda_i}\right)^q}{\left(1-\dfrac{\lambda_j}{\lambda_i}\right)^{n_j+q}}, j \neq i, i, j = 1,2 \tag{7}$$

With q=0,1,…..(n$_i$-1)

B. MIXED REYLEIGH AND RICEAN FADING

In this case we assume the first arriving path from each transmit antenna has a Ricean distribution envelope an the remaining path are Rayleigh faded .The channel coefficient has a mean value in the form(10)

$$m_k = \sqrt{\frac{K_k\Omega_k}{K_k+1}}.e^{j\theta_k}, k=1,2....2L \tag{8}$$

Where K$_k$ is the Rice factor and is zero for the Rayleigh faded path Ω_k the average SNR on the k-th path θ_k is the arrival angle of LoS Component with respect to the mobile moving direction on the k-th path in the presence of LoS.

$$C(k,l)=\frac{\sqrt{\Omega_k\Omega_l}}{\sqrt{K_k+1}\sqrt{K_l+1}}\rho_{k,l},k,l=1,2,\ldots,2L \tag{9}$$

Where P_{kj} is the correlation coefficient between the k-th path and i-th path In the WCDMA downlink each user is assigned an orthogonal channelization code . However, when multi-path propagation is present, the code orthogonality is destroyed. The average uncoded BER in (2) assumes a PCG multi-channel environment without ISI .It therefore gives a lower bound evaluate the effectiveness Of this lower bound for the indoor radio channels, in which both the Inter path interference and nonzero path signal correlation at he Rake combiner output due to the RR combing contribute to the BER increase.

III. IMPLEMENTATION ISSUES

So far in this report, we have shown, mathematically, that the new transmit diversity scheme with two transmit and receive antennas is equivalent to MRRC with one transmit antenna and receive antennas. From practical implementation aspects, however, the two systems may differ. This section discusses some of the observed difference between the two schemes.

A. Power Requirements

The new scheme requires the simultaneous transmission of two different symbols out of two antennas. If the system is radiation power limited, in order to have the same total radiated power from two transmit antennas the energy allocated to each symbol should be halved. This results in a 3-dB penalty in the error performance. However, the 3-dB reduction of power in each transmit chain translates to cheaper, smaller, or less linear power amplifiers. A 3-dB reduction in amplifiers power handling is very significant and may be desirable in some cases. It is often less expensive (or more desirable from inter modulation distortion effects) to employ two half-power amplifiers rather than a single full power amplifier. More over, if the limitation is only due to RF power handling (amplifier sizing, linearity, etc.), then the total radiated power may boubled and no performance penalty is incurred.

B .Sensitivity to Channel Estimation Errors

Throughout this paper, it is assumed that the receiver has perfect knowledge of the channel. The channel information may be derived by pilot symbol insertion and extraction [7],[8]. Known symbols are transmitted periodically from the transmitter to the receiver. The receiver extracts the samples and interpolates them to construct an estimate of the channel for every data symbol transmitted. There are many factors that may degrade the performance of pilot insertion and extraction techniques, such as mismatched interpolation coefficients and quantization effects. The dominant source of estimation errors for narrowband systems, however, is time variance of the channel. The channel estimation error is minimized when the pilot insertion frequency is greater or equal to the channel Nyquist sampling rate, which is two times the maximum Doppler frequency. Therefore, as long as the channel is sampled at a sufficient rate, there is little degradation due to channel estimation errors. For receive diversity combining schemes with antennas, at a given time, independent samples of the channels are available. With transmitters and a single receiver, however, the estimates of the channels must be derived from a single received signal. The channel estimation task is therefore different. To estimate the channel from one transmit antenna to the receive antenna the pilot symbols must be transmitted only from the corresponding transmit antenna. To estimate all the channels, the pilots must alternate between the antennas (or orthogonal pilot symbols have to be transmitted from the antennas). In either case, times as many pilots are needed. This means that for the two-branch transmit diversity schemes discussed in this report, twice as many pilots as in the two-branch receiver combining scheme are needed.

C .The Delay Effects

With branch transmit diversity, if the transformed copies of the signals are transmitted at distinct intervals from all the antennas, the decoding delay is symbol periods. That is, for the two-branch diversity scheme, the delay is two symbol periods. For a multi carrier system, however, if the copies are sent at the same time and on different carrier frequencies, then the decoding delay is only one symbol period.

D .Antenna Configurations

For all practical purposes, the primary requirement for diversity improvement is that the signals transmitted from the different antennas be sufficiently uncorrelated (less than 0.7 correlation) and that they have almost equal average power (less than 3-dB difference). Since the wireless medium is reciprocal, the guidelines for transmit antenna configurations are the same as receive antenna configurations. For instance, there have been many measurements and experimental results indicating that if two receive antennas are used to provide diversity at the base station receiver, they must be on the order of ten wavelengths apart to provide sufficient decorrelation. Similarly, measurements show that to get the same diversity improvement at the remote units it is sufficient to separate the antennas at the remote station by about three wavelengths. This is due to the difference in the nature of the scattering environment in the proximity of the remote and base stations. The remote stations are usually surrounded by nearby scatterers, while the base station is often placed at a higher altitude, with no nearby scatterers. Now assume that two transmit antennas are used at the base station to provide diversity at the remote station on the other side of the link. In other words, to provide sufficient de correlation between the signals transmitted from the two transmit antennas at the base station, we must have on the order of ten wavelengths of separation between the two transmit antennas. Equivalently, the transmit antennas at the remote units must be separated by about three wavelengths to provide diversity at the base station. It is worth noting that this property allows the use of existing receive diversity antennas at the base stations for transmit diversity. Also, where possible, two antennas may be used for both transmit and receive at the base and the remote units, to provide a diversity order of four at both sides of the link.

E. Soft Failure

One of the advantages of receive diversity combining schemes is the added reliability due to multiple receive chains. Should one of the receive chains fail, and the other receive chain is operational, then the performance loss is on the order of the diversity gain. In other words, the signal may still be detected, but with inferior quality. This is commonly referred to as soft failure. Fortunately, the new transmit diversity scheme provides the same soft failure. Therefore, the diversity gain is lost but the signal may still be detected. For the scheme with two transmit and two receive antennas, both the transmit and receive chains are protected by this redundancy scheme. The separation required depends on many factors such as antenna heights and the scattering environment. The figures given apply mostly to macrocell urban and suburban environments with relatively large base station antenna heights.

F. Impact on Interference

The new scheme requires the simultaneous transmission of signals from two antennas. Although half the power is transmitted from each antenna, it appears that the number of potential interferers is doubled, i.e., we have twice the number of interferers, each with half the interference power. It is often assumed that in the presence of many interferers, the overall interference is Gaussian distributed [10]. Depending on the application, if this assumption holds, the new scheme results in the same distribution and power of interference within the system. If interference has properties where interference cancellation schemes (array processing techniques) may be effectively used, however, the scheme may have impact on the system design. It is not clear whether the impact is positive or negative. The use of transmit diversity schemes (for fade mitigation) in conjunction with array processing techniques for interference mitigation has been studied for space-time trellis codes [9]. Similar efforts are under way to extend these techniques to the new transmit diversity scheme.

G. As an example, we consider the fallowing multi-channel covariance matrix with L=2

$$C = \frac{1}{\Omega_1^2}\begin{bmatrix} 1 & \gamma\rho_1 & \rho_2 & 0 \\ \gamma\rho_1 & \gamma^2 & 0 & \gamma^2\rho_2 \\ \rho_2 & 0 & 1 & \gamma\rho_1 \\ 0 & \gamma^2\rho_2 & \gamma\rho_1 & \gamma^2 \end{bmatrix}$$

This corresponds to a correlated multi-channel with temporal correlation ρ_1 and spatial correlation coefficient ρ_2 where $\gamma^2 = \Omega_2^2/\Omega_1^2$ is the multi-path component average SNR ratio. This channel has a positive definite Hermitian covariance matrix if $\rho_1 + \rho_2 < 1$ with the assumption that both correlation coefficients are positive. As can be noted in results the temporal correlation improve the BER performance in the mixed mode Ricean/Rayleigh fading Channel condition Which is contrary to that of pure Rayleigh fading channel. The BER improvement becomes more obvious when the total transmit E_b/N_o is large.

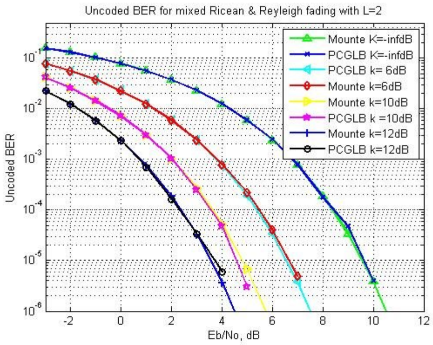


Figure 2: BER Performance for Mixed Ricean and Rayleigh fading

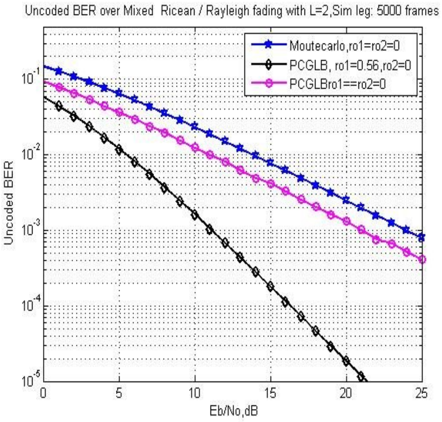


Figure 3: BER Performance for Mixed Ricean and Rayleigh fading for 5000 frames

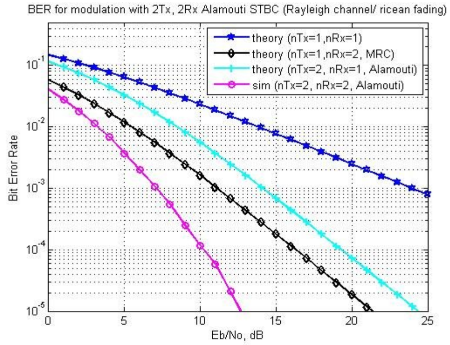


Figure 4: BER performance of 2x2 Alamouti Space Time Coding.

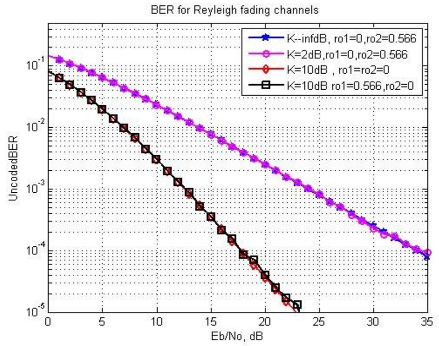


Figure 5: BER performance of Rayleigh fading in Indoor Applications

IV. CONCLUSION

This paper provide a PCG lower bound for the uncoded average BER for the 3GPP WCDMA FDD downlink ,taking into account both the temporal and spatial multi-channel correlation . A link- level simulation with a wideband ray tracing channel modeling shows that substantial spatial and temporal diversity gain can still be obtaind with the suboptimal RR-MRC receiver even in a mixed mode spread around 50 ns. The simulation shows the result analysis of Alamouti Space Time Block Coding, system optimization with 2x2 diversity under Rayleigh and Rican fadings. The advanced modulation schemes can be further used to improve the BER rates with utilization of proper Spatial Diversity.

REFERENCES

(1) Texas instruments , “space time block coded transmit diversity for WCDMA ,”Tdoc 662/98 ETSI SMG2L1,Espoo,Finland Dec.1998.

(2) S.M. Alamouti, “A simple transmit diversity technique for wireless communications” ,”IEEE J.select Areas commun ,vol .16,pp.1451-1458.Oct 1998.

(3) 3GPP TS 25.211 V4.2.0 (2001-09), Third Generation Partnership Project . Technical Specification Group Radio Access Network s Release 4, September 2001.

(4) 3GPP TS 25.213 v4.1.0 (2001-06),Third Generation Partnership Project Technical Specification Group Radio Access Network Spreading and modulation (FDD) ,Release 4, jun 2001.

(5) M.R Hueda ,G .Corral –Briones and C.E Rodriguez “MMSEC-RAKE receivers with RR of the diversity branch analysis , simulation and applications,” IEEE Trans Commun ,vol 47 ,no .2 pp.272 -280 Feb 1999.

(6) J.Yang ,”Diversity receiver scheme and system performance calculation for a CDMA system ,”IEEE Trans Commun,vol 49 no 6,pp 1073-1081, jan 2001.

(7) A. Neskovic, N. Neskovic and G. Paunovic “Modern apporches in modeling of mobile radio systems propagation environment ,”IEEE Commun surveys and Tutorials vol 3 ,pp.2-12 ,Third Quarter 2000 .http://www.comsoc.org/pubs/surveys.

(8) H. Hashemi , “The indoor radio propagation channel ,” proc .IEEE ,VOL 81, NO.7. PP.943-968,JUL 1993.

(9) V.V.Veeravelli “On performance analysis for signaling on correlated fading channels ,”IEEE Trans Commun, vol. 49 no 11 ,pp 1879-1883 , Nov 2001.

(10) G.L Stuber ,Principles of mobiles Communications ,kluwer Academic publisher,1996.

CALCULATION OF CYCLIC CONVOLUTION USING FNT TO PROVIDE HIGH SPEED ARCHITECTURE

S.Swathi[1], B.Sreedevi[2], V.Vijaya[3], S.Vaishali [4]

[1,2,3,4] ECE Dept. Vaagdevi College of Engineering, JNTU, Hyd

sreeramswathi210@gmail.com, +91-9703777645, vaagvijs_15@yahoo.co.in, +91-9704024475,

vsrtej@yahoo.co.in, +91-9849997298 and vaishalimtech@gmail.com, +91-9949061133

Abstract- This paper presents a high speed parallel architecture for cyclic convolution based on Fermat Number Transform (FNT) in the diminished-1 number system. A code conversion method without addition (CCWA) and a butterfly operation method without addition (BOWA) are proposed to perform the FNT and its inverse (IFNT) except their final stages in the convolution. The point wise multiplication in the convolution is accomplished by modulo 2^n+1 partial product multipliers (MPPM) and output partial products which are inputs to the IFNT. Thus modulo 2^n+1 carry propagation additions are avoided in the FNT and the IFNT except their final stages and the modulo 2^n+1 multiplier. The execution delay of the parallel architecture is reduced evidently due to the decrease of modulo 2^n+1 carry propagation addition. Compared with the existing cyclic convolution architecture, the proposed one has better throughput performance and involves less hardware complexity.

Keywords - *NTT, FNT, modulo 2^n+1 operation, diminished-1 number system.*

I. INTRODUCTION

Convolution is a mathematical way of combining two signals to form a third signal. Convolution helps to determine the effect a system has on an input signal. The cyclic convolution based on FFT is a widely used operation in signal processing, which needs to be performed in a complex domain even if both of the sequences to be performed would be real [1][2].Additionally, the dynamic range of the numbers varies widely so that one need to use floating point numbers to avoid scaling and quantization problems. Some architectures for efficient cyclic convolution have been developed to overcome the problems based on Number Theory Transform (NTT) [3-6]. They replace the complex domain with a finite field or a finite residue ring and can be defined by the FFT-like formula. All arithmetic operations are performed modulo *m* and the convolution results are exact without rounding errors. When the modulus in NTT is a Fermat number ($F_t = 2^{2^t}+1$, the t^{th} Fermat; *t* is an integer), the NTT turns into the Fermat Number Transform (FNT). The multiplication in the FNT and its inverse (IFNT) can be converted into bit shifts when the transform kernel is 2 or its integer power. Though the modulus of the FNT has a strict relationship with its maximum transform, the cyclic convolution based on FNT is more attractive than the conventional method in some areas.

Most cyclic convolution architectures based on FNT are implemented for the operands in the diminished-1 representation. Thus the code conversion (CC) stage which converts the normal binary numbers into their diminished-1 representation is compulsory. Other arithmetic operations described originally by Leibowitz includes modulo 2^n+1 negation, addition, subtraction, multiplication operations in the diminished-1 number system [8]. These operations constitute the butterfly operation (BO) which is the most important element in the FNT [7]. The CC and the BO are both mainly composed of modulo 2^n+1 adders of which the fastest one in the diminished-1 number system is proposed by Vergos so far [9]. The fast modulo 2^n+1 adder involves the carry-propagation addition computation and is used in the recent FNT implementations [3-4].

In the structure of FNT, it doesn't need multipliers when doing transform, because multiply by 2^K can be accomplished by only left shift. Thus the FNT seems to provide a more efficient transform domain convolution than utilizing FFT. To avoid the problem of n+1 bits usage instead of n bits diminished-1 number system is used.

In this paper, a code conversion method without addition (CCWA) and a butterfly operation method without addition (BOWA) which takes full advantage of the carry-save adder are proposed to accomplish the cyclic convolution with the unity root 2 or its integer power. The modulo 2^n+1 partial product multiplier (MPPM) is used to accomplish the point wise multiplication so that the final carry-propagation addition of two partial product in the multiplier is avoided. Thus the execution delay of the architecture is reduced evidently. Model estimations and experiment results show that the proposed architecture is faster than the existing one when the modulus of the FNT is no less than 2^8+1. For wider modulus, the proposed parallel architecture leads to considerably faster hardware implementations than those presented in [3-4].

The rest of this paper is organized as follows: the foundations of cyclic convolution based on FNT are formulated in the next section. The important operations in cyclic convolution are presented in section 3. In section 4, the parallel architecture for cyclic convolution based on FNT is illustrated. Comparative results that show the efficiency of the proposed architecture against the existing solution are presented in section 5. Finally we conclude this paper in section 6.

II. FOUNDATIONS

The cyclic convolution via the FNT is composed of the FNTs, the point wise multiplications and the IFNT. FNTs of two sequences $\{a_i\}$ and $\{b_i\}$ will produce two sequences $\{Ai\}$ and $\{Bi\}$. Modulo 2^n+1 multipliers are employed to accomplish the point wise multiplication between $\{A_i\}$ and $\{B_i\}$ and produce the sequence $\{P_i\}$. The final resulting sequence $\{p_i\}$ can be obtained by taking the inverse FNT of the product sequence $\{P_i\}$. Each element in the $\{p_i\}$ is in the diminished-1 representation.

The FNT of a sequence of length N $\{x_i\}(i=0,1,2,3....N-1)$ is defined as

$$X_k = \sum_{i=0}^{N-1} x_i \alpha_N^{<ik>} \bmod F_t$$

$$(k = 0,1,.....N-1) \tag{1}$$

Where $F_t = 2^{2^t}+1$,the t^{th} Fermat; N is a power of 2 and α is an Nth root unit (*i.e.* $\alpha_N^N \bmod F_t = 1$ and $\alpha_N^m \bmod F_t \neq 1$, $1 \le m \le N$). The notation <*ik*> means *ik* modulo *N*.

The inverse FNT is given by

$$x_i = \frac{1}{N}\sum_{i=0}^{N-1} X_k \alpha_N^{-<ik>} \bmod F_t$$

$$(i=0,1,...,N-1) \tag{2}$$

Where 1/*N* is an element in the finite field or ring of integer and satisfies the following condition:

$$(N\cdot \tfrac{1}{N}) \bmod F_t = 1 \tag{3}$$

Parameters α , F_t , *N* must be chosen carefully and some conditions must be satisfied so that the FNT possesses the cyclic convolution property [3]. In this paper, we choose α=8, $F_t = 2^{2^t}+1$ and N=2. 2^t where t is an integer.

III. IMPORTANT OPERATIONS IN CYCLIC CONVOLUTION BASED ON FNT

Important operations of the cyclic convolution based on FNT with the unity root 2 include the CCWA, the BOWA and the MPPM. The CCWA and the BOWA both consist of novel modulo 2^n+1 4-2 compressors mainly which are composed of the 4-2 compressor introduced by Nagamatsu [10]. The 4-2 compressor, the novel modulo 2^n+1 4-2 compressor and the BOWA are shown in Fig. 1. In the figure, "X^*" denotes the diminished-1 representation of X, i.e. $X^* = X - 1$.

A. Code conversion without addition

The CC converts normal binary numbers (NBCs) into their diminished-1 representation. It is the first stage in the FNT. Delay and area of CC of a 2*n*-bit NBC are no less than the ones of two *n*-bit carry propagation adders. To reduce the cost, we propose the CCWA that is performed by the modulo 2^n+1 4-2 compressor.

Let *A* and *B* represent two operands whose widths are no more than 2*n* bits. We define two new variables:

$$\begin{cases} A = 2^n A_H + A_L \\ B = 2^n B_H + B_L \end{cases} \tag{4}$$

and

$$\begin{cases} M_0 = (2^n-1) - A_H = \overline{A}_H \\ M_1 = (2^n-1) - B_H = \overline{B}_H \\ M_2 = (2^n-1) - B_L = \overline{B}_L \end{cases} \tag{5}$$

If the subsequent operation of CC is modulo 2^n+1 addition, assign A_L, M_0, B_L and M_1 to I_0, I_1, I_2, I_3 in the Modulo 2^n+1 4-2 compressor respectively. I_0, I_1, I_2, I_3 are defined as follows:

$$\begin{cases} I_0 = I_{0(n-1)} I_{0(n-2)} \cdots\cdot I_{01} I_{00} \\ I_1 = I_{1(n-1)} I_{1(n-2)} \cdots\cdot I_{11} I_{10} \\ I_2 = I_{2(n-1)} I_{2(n-2)} \cdots\cdot I_{21} I_{20} \\ I_3 = I_{3(n-1)} I_{3(n-2)} \cdots\cdot I_{31} I_{30} \end{cases} \tag{6}$$

We obtain the sum vector H_0^* and carry vector H_1^* in the diminished -1 number system. The most significant bit of H_1^* is complemented and connected back to its least significant bit. That is to say

$$\begin{cases} H_0^* = H_{0(n-1)}H_{0(n-2)}\ldots.H_{01}H_{00} \\ H_1^* = H_{1(n-2)}\ldots.H_{11}H_{10}\overline{H}_{1(n-1)} \end{cases} \tag{7}$$

The result of modulo 2^n+1 addition of A^* and B^* is equal to the result of modulo 2^n+1 addition of H_0^* and H_1^*. In this way, *A* and *B* are converted into their equivalent diminished-1 representations H_0^* and H_1^*.

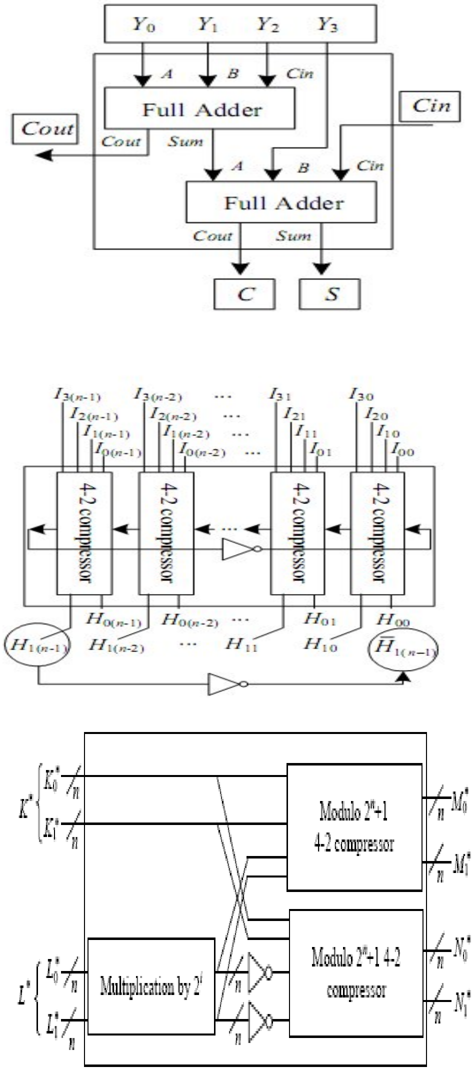


Fig.1. Elementary operations of FNT architecture with unity root 2
(a) 4-2 compressor (b) Modulo 2^n+1 4-2 compressor
(c) Butterfly operation without addition

The result of modulo 2^n+1 addition of *A** and *B** is equal to the result of modulo 2^n+1 addition of H_0^* and H_1^* .In this way, *A* and *B* are converted into their equivalent diminished-1 representations H_0^* and H_1^*.

Let

$$\left|A^* + B^*\right|_{2^n+1}, \left|\overline{A^*}\right|_{2^n+1}, \left|A^* - B^*\right|_{2^n+1}, \left|A^* \times B^*\right|_{2^n+1}$$

Denote modulo 2^n+1 addition, negation, subtraction and multiplication by the power of 2 respectively which are proposed by Leibowitz originally [8]. The CCWA for subsequent modulo 2^n+1 addition can be described as follows:

$$\left|A^* + B^*\right| = \left|A_L + M_0 + B_L + M_1\right|_{2^n+1} = \left|H_0^* + H_1^*\right|_{2^n+1} \tag{8}$$

If the subsequent operation is modulo 2^n+1 subtraction, we assign A_L, M_0, M_2 and B_H to I_0, I_1, I_2, I_3 respectively. Then $H_0^*$and H_1^* in the modulo 2^n+1 4-2 compressor constitute the result of the CCWA. The conversion is described as follows

$$\begin{aligned} \left|A^* - B^*\right|_{2^n+1} &= \left|A - B\right|_{2^n+1} = \left|A - \overline{B}\right|_{2^n+1} = \left|A_L + M_0 + M_2 + B_H\right|_{2^n+1} \\ &= \left|H_0{}^* + H_1^*\right|_{2^n+1} \end{aligned} \tag{9}$$

After CCWA, we obtain the result consisting of two diminished-1 numbers. The result also includes the information of modulo 2^n+1 addition or subtraction in the first stage of previous BO.

B Butterfly operation without addition

After the CCWA, we obtain the results of modulo 2^n+1 addition and subtraction in the diminished-1 representation. Each result consists of two diminished-1 values. The subsequent butterfly operation involves four operands. The proposed BOWA involves two modulo 2^n+1 4-2 compressors, a multiplier and some inverters as shown in Fig. 1(c). The multiplication by an integer power of 2 in the diminished-1 number system in the BOWA is trivial and can be performed by left shifting the low-order *n-i* bits of the number by *i* bit positions then inversing and circulating the high order *i* bits into the *i* least significant bit positions [7].

Thus the BOWA can be performed without the carry-propagation chain so as to reduce the delay and the area obviously. K^*, L^*, M^*, N^* are corresponding to two inputs and two outputs of previous BO in the diminished-1 number system respectively and given by

$$\begin{cases} M^* = \left|M_0^* + M_1^*\right|_{2^n+1} = \left|K_0^* + K_1^* + L_0^* \times 2^i + L_1^* \times 2^i\right|_{2^n+1} = \left|K^* + L^* \times 2^i\right|_{2^n+1} \\ N^* = \left|N_0^* + N_1^*\right|_{2^n+1} = \left|K_0^* + K_1^* - L_0^* \times 2^i - L_1^* \times 2^i\right|_{2^n+1} = \left|K^* - L^* \times 2^i\right|_{2^n+1} \\ = \left|K^* + \overline{L^* \times 2^i}\right|_{2^n+1} \end{cases} \tag{10}$$

Where

$$K^* = \left|K_0^* + K_1^*\right|_{2^n+1}, \quad L^* = \left|L_0^* + L_1^*\right|_{2^n+1}$$

C Modulo 2^n +1 partial product multiplier

For the modulo 2^n+1 multiplier proposed by Efstathiou, there are n+3 partial products that are derived by simple AND and NAND gates [11]. An FA based Dadda tree that reduces the n+3 partial products into two summands is followed. Then a modulo 2^n+1 adder for diminished-1 operands is employed to accept these two summands and produce the required product.

In the proposed parallel architecture for cyclic convolution based on FNT, the BOWA can accept four operands in the diminished-1 number system. Every point wise multiplication only needs to produce two partial products rather than one product. The operation can be accomplished by taking away the final modulo 2^n+1 adder of two partial products in the multiplier. Thus the final modulo 2^n+1 adder is omitted and the modulo 2^n+1 partial product multiplier is employed to save the delay and the area.

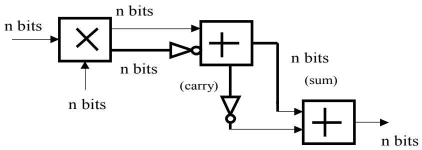


Fig.2 Modulo 2^n +1 partial product multiplier

IV. PARALLEL ARCHITECTURE FOR CYCLIC CONVOLUTION

Based on the CCWA, the BOWA and the MPPM, we design the whole parallel architecture for the cyclic convolution based on FNT as shown in Fig. 3. It includes the FNTs, the point wise multiplication and the IFNT mainly. FNTs of two input sequences $\{a_i\}$ and $\{b_i\}$ produce two sequences $\{A_i\}$ and $\{B_i\}$ (i=1,2, …N-1). Sequences $\{A_i\}$ and $\{B_i\}$ are sent to N MPPMs to accomplish the point wise multiplication and produce N pairs of partial products. Then the IFNT of the partial products are performed to produce the resulting sequence $\{p_i\}$ of the cyclic convolution.

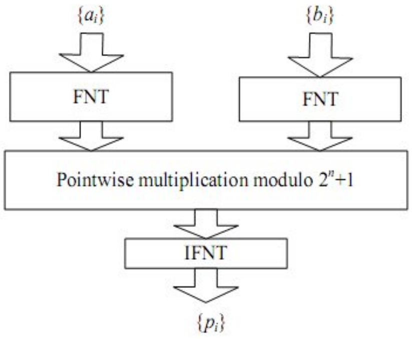
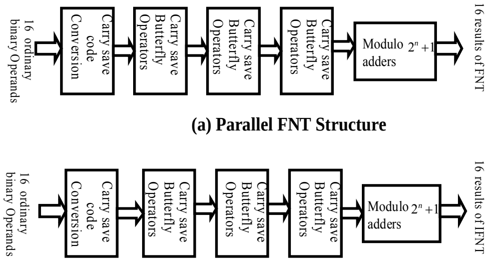


Fig.3. Parallel-architecture for cyclic-convolution

In the architecture, the radix-2 decimation-in-time (DIT) algorithm which is by far the most widely used algorithm is employed to perform the FNT and the IFNT [12]. Illustrative examples of the FNT and the IFNT are shown in Fig. 4 in the case the transform length is 16 and the modulus is $2^8 + 1$


16 ordinary binary Operands
Carry save code Conversion
Carry save Butterfly Operators
Carry save Butterfly Operators
Carry save Butterfly Operators
Modulo 2^n+1 adders
16 results of FNT
16 ordinary binary Operands
Carry save code Conversion
Carry save Butterfly Operators
Carry save Butterfly Operators
Carry save Butterfly Operators
Modulo 2^n+1 adders
16 results of IFNT

(a) Parallel FNT Structure

(b) Parallel IFNT Structure

Fig. 4 Structures for FNT and IFNT ($Ft = 2^8 + 1$)

The efficient FNT structure involves $\log_2 N$+1 stages of operations. The original operands are converted into the diminished-1 representation in the CCWA stage, containing the information of modulo 2^n+1 addition or subtraction in the first butterfly operation stage of the previous FNT structure. Then the results are sent to the next stage of BOWA. After log2N-1 stages of BOWAs, the results composed of two diminished-1 operands are obtained. The final stage of FNT consists of modulo 2^n+1 carry-propagation adders which are used to evaluate the final results in the diminished-1 representation. The CCWA stage, the BOWA stage and the modulo 2^n+1 addition stage in the FNT involves N/2 couples of code conversions including the information of modulo 2^n+1 addition and subtraction, N/2 butterfly operations and N/2 couple of modulo 2^n+1 additions respectively.

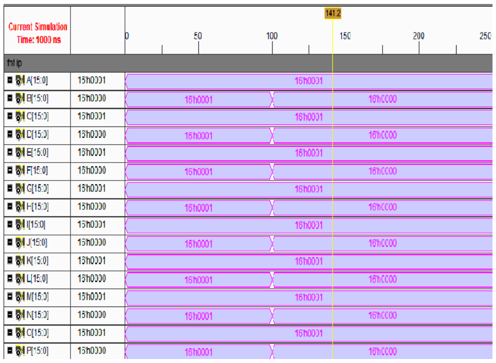
From the definition of FNT and IFNT in section 2, the only difference between the FNT and the IFNT is the normalization factor 1/N and the sign of the phase factor α_N. If ignoring the normalization factor 1/N, the above formula is the same as that given in the FNT except that all transform coefficients $\alpha_N^{<ik>}$ used for the FNT need to be replaced by $\alpha_N^{-<ik>}$ for the IFNT computation. The proposed FNT structure can be used to complete the IFNT as well with little modification as shown in Fig. 4(b). After the IFNT of N-point bit reversed input data, the interim results are multiplied by 1/N in the finite field or ring. Then x[j] and x [j+N/2] (j=1, 2,… N/2-1) exchange their positions to produce the final results of the IFNT in natural order.

Our architecture for the cyclic convolution gives a good speed performance without requiring a complicated control. Furthermore, it is very suitable for implementation of the overlap-save and overlap add techniques which are used to reduce a long linear convolution to a series of short cyclic convolutions.

V. COMPARISION AND RESULTS

In this section, we compare the proposed parallel architecture for the cyclic convolution against that by using FFT. The structure of the FNT is identical to that of the discrete Fourier transform for power of two lengths the same algorithm can be used for the classical radix-2 FFT and the radix-2 fast FNT, respectively. The only difference is the substitution of the complex multiplication by complex N-th roots of unity in the Fourier transform case by simple bit shifting operations in the case of the Fermat number transform. In the structure of FNT, it doesn't need multipliers when doing transform, because multiply by 2^K can be accomplished by only left shift. Thus the FNT seems to provide a more efficient transform domain convolution than utilizing FFT.Thus our architecture is faster and more efficient than the existing one. The number theoretic transforms are defined over a finite ring (as the definition mentioned) of integers and are operated in modulo arithmetic. In this ring, using transforms similar to the DFT, the cyclic convolution can be performed very efficiently and without any round off errors.

Fnt input

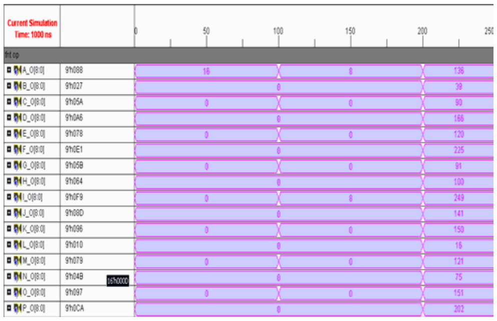


Current Simulation Time: 1000 ns

fnt ip

A[15:0]	16'h0001	16'h0001	
B[15:0]	16'h0000	16'h0001	16'h0000
C[15:0]	16'h0001	16'h0001	
D[15:0]	16'h0000	16'h0001	16'h0000
E[15:0]	16'h0001	16'h0001	
F[15:0]	16'h0000	16'h0001	16'h0000
G[15:0]	16'h0001	16'h0001	
H[15:0]	16'h0000	16'h0001	16'h0000
I[15:0]	16'h0001	16'h0001	
J[15:0]	16'h0000	16'h0001	16'h0000
K[15:0]	16'h0001	16'h0001	
L[15:0]	16'h0000	16'h0001	16'h0000
M[15:0]	16'h0001	16'h0001	
N[15:0]	16'h0000	16'h0001	16'h0000
C[15:0]	16'h0001	16'h0001	
P[15:0]	16'h0000	16'h0001	16'h0000

Fnt output

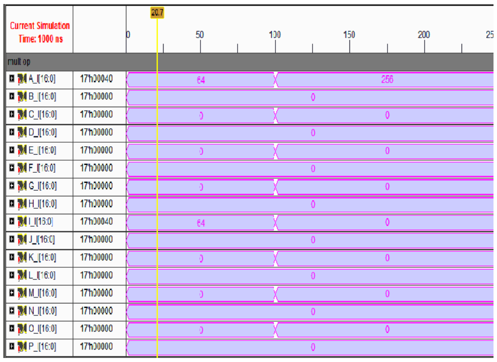


Current Simulation Time: 1000 ns

fnt op

A_O[8:0]	9'h088	16	8	136
B_O[8:0]	9'h027	0		39
C_O[8:0]	9'h05A	0	0	90
D_O[8:0]	9'h0A6	0		166
E_O[8:0]	9'h078	0	0	120
F_O[8:0]	9'h0E1	0		225
G_O[8:0]	9'h05B	0	0	91
H_O[8:0]	9'h064	0		100
I_O[8:0]	9'h0F9	0	8	249
J_O[8:0]	9'h08D	0		141
K_O[8:0]	9'h096	0	0	150
L_O[8:0]	9'h010	0		16
M_O[8:0]	9'h079	0	0	121
N_O[8:0]	9'h04B	0		75
O_O[8:0]	9'h097	0	0	151
P_O[8:0]	9'h0CA	0		202

Multiplier output

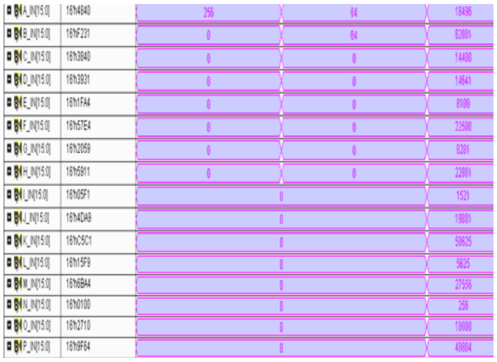


Current Simulation Time: 1000 ns

mult op

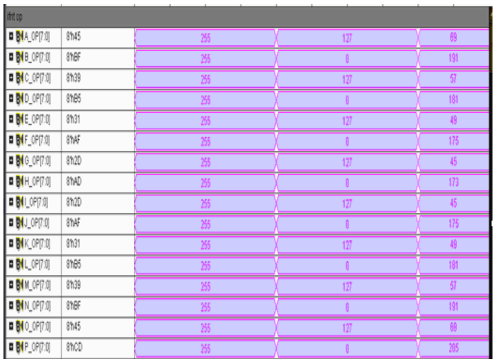
A_I[16:0]	17'h00040	64	256
B_I[16:0]	17'h00000	0	
C_I[16:0]	17'h00000	0	0
D_I[16:0]	17'h00000	0	
E_I[16:0]	17'h00000	0	0
F_I[16:0]	17'h00000	0	
G_I[16:0]	17'h00000	0	0
H_I[16:0]	17'h00000	0	
I_I[16:0]	17'h00040	64	0
J_I[16:0]	17'h00000	0	
K_I[16:0]	17'h00000	0	0
L_I[16:0]	17'h00000	0	
M_I[16:0]	17'h00000	0	0
N_I[16:0]	17'h00000	0	
O_I[16:0]	17'h00000	0	0
P_I[16:0]	17'h00000	0	

Ifnt input



A_IN[15:0]	16'h4840	256	64	18496
B_IN[15:0]	16'hF231	0	64	62001
C_IN[15:0]	16'h3840	0	0	14400
D_IN[15:0]	16'h3931	0	0	14641
E_IN[15:0]	16'h1FA4	0	0	8100
F_IN[15:0]	16'h57E4	0	0	22500
G_IN[15:0]	16'h2059	0	0	8281
H_IN[15:0]	16'h5911	0	0	22801
I_IN[15:0]	16'h05F1	0		1521
J_IN[15:0]	16'h4DA9	0		19881
K_IN[15:0]	16'hC5C1	0		50625
L_IN[15:0]	16'h15F9	0		5625
M_IN[15:0]	16'h6BA4	0		27556
N_IN[15:0]	16'h0100	0		256
O_IN[15:0]	16'h2710	0		10000
P_IN[15:0]	16'h9F64	0		40804

Ifnt output



ifnt op

A_OP[7:0]	8'h45	255	127	69
B_OP[7:0]	8'hBF	255	0	191
C_OP[7:0]	8'h39	255	127	57
D_OP[7:0]	8'hB5	255	0	181
E_OP[7:0]	8'h31	255	127	49
F_OP[7:0]	8'hAF	255	0	175
G_OP[7:0]	8'h2D	255	127	45
H_OP[7:0]	8'hAD	255	0	173
I_OP[7:0]	8'h2D	255	127	45
J_OP[7:0]	8'hAF	255	0	175
K_OP[7:0]	8'h31	255	127	49
L_OP[7:0]	8'hB5	255	0	181
M_OP[7:0]	8'h39	255	127	57
N_OP[7:0]	8'hBF	255	0	191
O_OP[7:0]	8'h45	255	127	69
P_OP[7:0]	8'hCD	255	0	205

TABLE 1

Operation	Device utilization	Delay
Using FNT	49% (2314 out of 4656)	47.029 ns
Using FFT (For FFT of one sequence only)	64% (3016 out of 4656)	61.291 ns

Delay and the area utilization of cyclic convolution using FNT and FFT are given in Table1 indicating that the proposed architecture comprising the CCWA and the BOWA require less delay and area than FFT.

VI. CONCLUSION

A novel parallel architecture for the cyclic convolution based on FNT is proposed in the case the principle root of unity is equal to 2 or its integer power. The FNT and the IFNT are accomplished by the CCWA and the BOWA mainly. The point wise multiplication is performed by the modulo 2^n+1 partial product multiplier. Thus there are very little modulo 2^n+1 carry-propagation addition. Implementations using synthesis, FPGA prototype the proposed parallel architecture can attain lower area and delay than that of the existing one.

REFERENCES

[1] C. Cheng, K.K. Parhi, “Hardware efficient fast DCT based on novel cyclic convolution structures”, *IEEE Trans. Signal processing*, 2006, 54(11), pp. 4419- 4434

[2] H.C. Chen, J.I. Guo, T.S. Chang, *et al.*, “ A memory efficient realization of cyclic convolution and its application to discrete cosine transform”, *IEEE Trans. Circuit and system for video technology*, 2005, 15(3), pp. 445-453.

[3] R. Conway, “Modified Overlap Technique Using Fermat and Mersenne Transforms”, *IEEE Trans. Circuits and Systems II: Express Briefs*, 2006, 53(8), pp.632 – 636

[4] A. B. O'Donnell, C. J. Bleakley, “Area efficient fault tolerant convolution using RRNS with NTTs and WSCA”, *Electronics Letters*, 2008, 44(10), pp.648-649

[5] H. H. Alaeddine, E. H. Baghious and G. Madre *et al.*, “Realization of multi-delay filter using Fermat number transforms”, *IEICE Trans. Fundamentals*,2008, E91A(9), pp. 2571-2577

[6] N. S. Rubanov, E. I. Bovbel, P. D. Kukharchik, V. J. Bodrov, “Modified number theoretic transform over the direct sum of finite fields to compute the linear convolution”, *IEEE Trans. Signal Processing*, 1998, 46(3), pp. 813-817

[7] T. Toivonen, J. Heikkila, “Video filtering with fermat number theoretic transforms using residue number system”, *IEEE Trans. circuits and systems for video technology*, 2006, 16(1), pp. 92-101

[8] L. M. Leibowitz, “A simplified binary arithmetic for the Fermat number transform,” *IEEE Trans. Acoustics Speech and Signal Processing*, 1976, 24(5):356-359

[9] H. T. Vergos, C. Efstathiou, D. Nikolos, “Diminished one modulo 2^n + 1 adder design”, *IEEE Trans. Computers*, 2002, 51(12), pp. 1389-1399

[10] M. Nagamatsu, S. Tanaka, J. Mori, *et. al.* “15-ns 32 × 32-b CMOS multiplier with an improved parallel structure”, *IEEE Journal of Solid-State Circuits*, 1990,25(2), pp.494-497

[11] C. Efstathiou, H. Vergos, G. Dimitrakopoulos, et al., “Efficient diminished-1 modulo 2^n+1 multipliers”, *IEEE Trans. Computers*, 2005, 54(4), pp. 491-496

[12] J. G. Proakis and D. G. Manolakis, *Digital signal processing: principles, algorithms, and applications*, Prentice Hall, New Jersey, 2007.

Location Based Tracking System through location id sms ---“poor man’s GPS”

Pavankumar D Smt V.Geetha
Department of Electronics & communication University of B.D.T college of engineering
Davangere, Kuvempu University

ABSTRACT

With the world becoming ever smaller through technology, hiding is increasingly difficult. Cameras peer down on us at red lights, in our workplace, in stores and even at home. Now, those cameras are being augmented by new technologies that track our cars, cell phones and possibly any product we buy. Cell phones will take on a new role in 1998, beginning a slow transition to becoming user tracking devices. The outcome of this shift reassures some, but has others calling for restrictions on how cell-locating information can be used.

The present work helps where there exists a situation of emergency in which we are not in a position to explain where we are but need an emergency help. We need to transfer own location based details continuously to the server side cell phone and not leading to any suspicious behavior. Just load out application to your cell phone and enter the phone number of the cell phone to which you would like to pass the location data.

1. INTRODUCTION

Location based service could be used for tracking systems. It sends the location string automatically and continuously to the server and the server in-turn reads the message through commands and displays the same on computer. The project can be used in any investigation related services or anywhere where knowing the whereabouts of a person becomes very important. The marketing executive’s movements can be monitored by using this Location based service tracking system. It sends the location string automatically and continuously to the server and the server in-turn reads the message displays the same on computer.

This paper explains about the tracking the location of the mobile host and send the name of the location to the server. In turn it is to be displayed on computer for large audience. This project is mainly developed to monitor the sales person or any kind of marketing executives, by retrieving the location string which is the location of the tower provided by the service provider.

The present work can be used to monitor working of marketing executives and also helps to identify whether employees working or not. To maintain a better employee profile. Helps to improve the working condition of all the employees. It’s from reference (i) To get the location detail and Real time tracking of the mobile host. Cost effective GSM technology is deployed in the application.

1.1 OBJECTIVES OF PROJECT

- To get the location detail and Real time tracking of the mobile host.
- Cost effective GSM technology is deployed in our application.
- Can help tracking people in distress without much expectation from the other end.
- To display the location on LCD for large audience.

2. Location Based Tracking System (LBTS)

The present work helps where there exists a situation of emergency in which we are not in a position to explain where we are, but need an emergency help. We need to transfer our location based details continuously to the police department without touching the cell phone and leading to any suspicious behavior. Just load the application to the cell phone and enter the phone number of the cell phone to which we would like to pass the location data. The application continuously reads our location and transfers the location information to the destination PC without expecting any button press. Sending of SMS is automatic.

This paper gives overview gives an idea about overall working of the important modules which constitutes the LBTS. This project is mainly aimed on tracing and retrieving the location of the client.

There are two main two modules in this project.

(i) Tracing Module.
(ii) Displaying on monitor module.

Tracing Module:

This module contains the following sub-modules that are required during the mobile tracking condition. The following are the list and explanation of the sub-modules:

(i) Invoking the Application:
The application has to be invoked manually as soon as it is installed into the mobile phone.

(ii) Creating the message:
Created the message to be sent to the predefined number to which the message has to be sent.

(iii) Searching and Retrieving Location information:
The application fetches the location string from mobile application and inserts it into the body of the newly created message. Retrieving the location information is nothing but the location string information stored in the mobile phone. Some of the standard built in functions helps us to retrieve this location string.

(iv) Sending the Message:
This module is similar to any of the text messaging modules. The message created by the application is sent to the number which resides on server side.

Displaying on Monitor module:

This module comes into existence by getting the location of the client to print on to the monitor. The sub-modules prevailing under this module are:

(i). Fetching the Message:

It fetches the newly arrived message which contains location string of client.

(ii). Insert into Database:

The location string and the phone number of client are inserted into the database.

(iii). Displaying on monitor:

The location string is displayed onto the monitor through the interface of parallel port.

Software Requirements:

The software tools used are:

(i) Symbian software.
(ii) Visual c++ v 6.0.
(iii) Series 60 v 2.0 toolkit.
(iv) Symbian emulator.
(v) My Phone Explorer.

Any one of the operating systems WIN 98/2000/NT/XP.

Hardware Requirements:

These are the minimum hardware configuration required to run and install above mentioned software's. The hard requirements are:

(i) 128MB RAM.
(ii) 600MB of Hard Disk space.
(iii) USB Bluetooth dongle or Memory Card Reader.
(iv) Data cable to connect cell phone to PC.
(v) Cell Phone with PC connectivity option and built-in modem.
(vi) Smart phone with Symbian OS (Any 60 series cell phone).

3. SYSTEM DESIGN

System design is the process of art of defining the hardware and software architecture, components, modules, interface and data for a computer system to satisfy the specified requirements. One could see it as the application of system theory to computing.

System design is phase of software development where abstract model of the complete project is converted into the actual implementation model such as state diagrams, algorithm and actual coding in any programming language. At the end of design phase we will be having a raw product i.e. a complete coded project which has to be tested to prove its efficiency, reliability etc.

User interface for the system must be developed in system design phase. Good user interface helps user to understand the entire system with minimum effort.

Server-side: The server PC is connected to a cell phone with built-in modem via data cable. Also, the server is connected to the monitor display unit via parallel port.

Client-side: At the client end, we need a Symbian OS cell phone. Symbian C++ is the language used to develop the client side project. Once the project is tested with the emulator, .sis file is created and loaded to the smart phone. Once the application is started in the cell phone, give the phone number of the cell, which is connected to the server machine and let the application be in running state.

The client-side cell phone continuously reads the location string i.e. the tower that transmits signal to it and auto sends the same as SMS to the server side cell phone. The server PC auto read the incoming message and displays the same on monitor.

Project Cycle:

- Server-side: The server PC is connected to a cell phone with built-in modem via data cable. Also, the server is connected the display unit.
- Client-side: At the client end, you need a Symbian OS cell phone. Symbian C++ is the language used to develop the client side project. Once the project is tested with the emulator, .sis file is created and loaded to the smart phone. Once the application is started in the cell phone, give the phone number of the cell, which is connected to the server machine and let the application be in running state.
- The client-side cell phone continuously reads the location string i.e. the tower that transmits signal to it and auto sends the same as SMS to the server side cell phone. The server PC read the incoming message and displays the same on Monitor.

Fig shows the project flow diagram, when we start the application module client-side cell phone continuously reads the location string after that we retrieve location string, next we proceed to transfer the location string to server side if error=0, otherwise it will return loop back. Then we receive location string at the server from database and we display the same on Monitor.

Database Design:

The database we are using in this project is MS-Access. The snapshot of the database table is shown in figure 4.2. In this database design it include 3 attributes like place, phone number, date of tracking which are used to create a table.

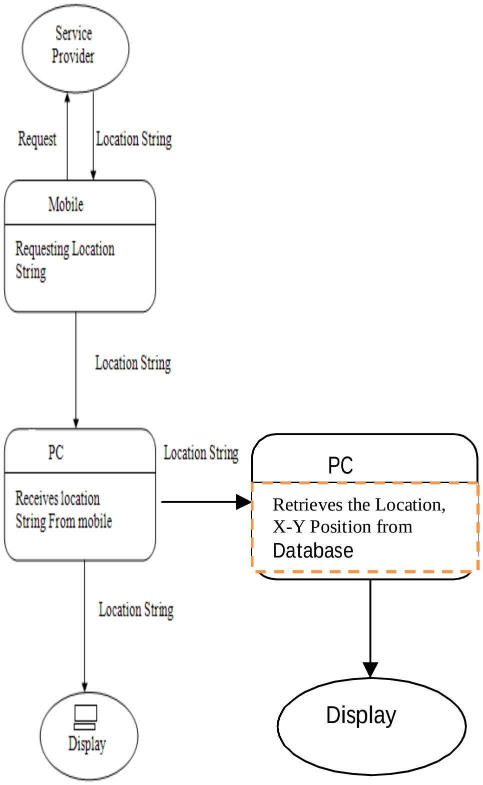

Service Provider
Request
Location String
Mobile
Requesting Location String
Location String
PC
Receives location String From mobile
Location String
PC
Retrieves the Location, X-Y Position from Database
Location String
Display
Display

Fig shows how the data flows between different modules of the project. Here we can observe the input which flows to modules and the output we will obtain from the modules.

Programming Techniques/Skills:

JDBC (Java Database Connectivity)

The JDBC (Java Database Connectivity) API defines interfaces and classes for writing database applications in Java by making database connections. Using JDBC you can send SQL, PL/SQL statements to almost any relational database. JDBC is a Java API for executing SQL statements and supports basic SQL functionality. It provides RDBMS access by allowing you to embed SQL inside Java code. Because Java can run on a thin client, applets embedded in Web pages can contain downloadable JDBC code to enable remote database access. You will learn how to create a table, insert values into it, query the table, retrieve results, and update the table with the help of a JDBC

Java SMS Library

SMS is an extensively used service and almost everyone is able to use it. This has lead to an increasing number of applications using SMS as the interface to the user. This project contains an SMS framework, written in Java. The framework is designed for robustness in order to serve as a gateway between users with their mobile phone, and the application. The implemented application is an SMS-to-OSC gateway, sending all incoming SMS' to an OSC capable host.

Description of SMS

Short Message System (SMS) is a part of the Global System for Mobile Communications (GSM) specification. In the first phase of the GSM networks, SMS reception was mandatory for a mobile phone (called Mobile Station (MS) in the specification), but sending SMS was an optional feature. From phase 2 of the GSM networks (1997), sending SMS became mandatory. The current networks are all phase 2 or later.

Serial Communication in Java

In order to use SMS in a computer application, an interface between the computer and the GSM network is needed. This interface is typically an MS connected to a serial port on the computer. In this project the serial connection to an MS is the interface. In Java, serial communication is specified in the package javax.comm, which is a part of the extended specification (the javax). When an application is communicating through a serial interface, it can basically do this in two ways: It can poll the serial port for data (synchronously), and it can receive data events (something like interrupts) when data is received (asynchronously). The javax.comm allows you to do both. Synchronous communication is done by using the read methods on the Input Stream of the serial port.

Front End Form Design:

The front end of the project is developed using swings.

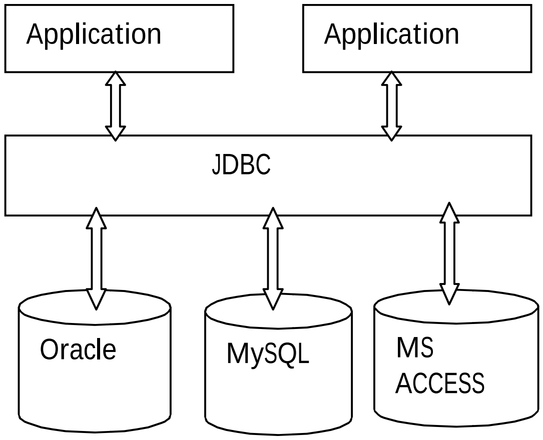
Swings

Javax.Swing is a package used to create platform Independent GUI which will have good Look and Feel. Look and Feel will remain same even if you run the program on other platform which cannot be achieved through AWT.Swings are lightweight component,AWT are heavy weight component.

The Structure of JDBC

JDBC accomplishes its goals through a set of Java interfaces, each implemented differently by individual vendors. The set of classes that implement the JDBC interfaces for a particular database engine is called a JDBC driver. In building a database application, you do not have to think about the implementation of these underlying classes at all; the whole point of JDBC is to hide the specifics of each database and let you worry about just your application. Figure illustrates the JDBC architecture. Figure shows illustrate the steps to mention the data source name to connect database to java program.

The JDBC Architecture



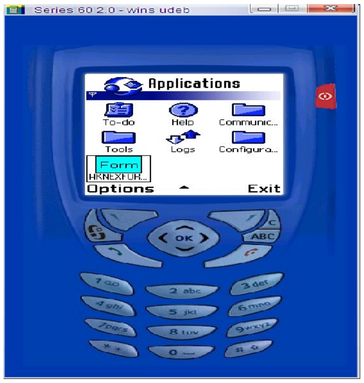
4. RESULTS:

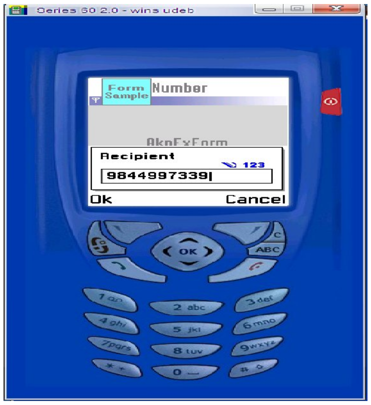
To execute the project Administration side

To execute the project first go to command prompt type command then press enter, after that set the path and Class path. The path is directing to bin folder and class path is directing to lib folder after that compilation and execution of java files to be carried out.

Go to the command prompt as shown below:
Start>run>write cmd & press ‘Ok’ button
C:\>cd C:\java\jdk1.5\bin
Now, Set the path & class path as shown below:
Set path= C:\java\jdk1.5.0\bin;.;

Client Side:





Recipient
9844997339

Server Side:

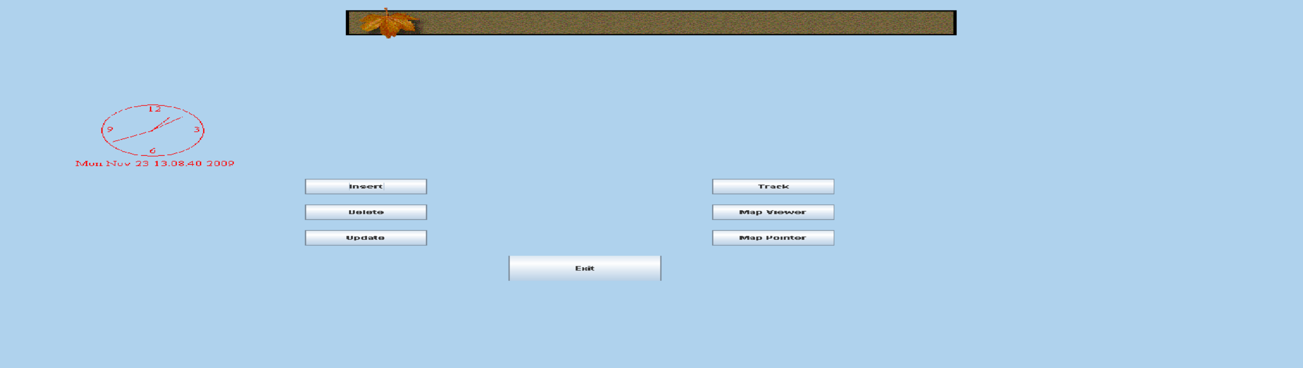


Fig Snapshot of Front end screen

Figure shows the front end screen of the project. When we execute project seven fields will appear. In that first field is to insert the place and X-Y position, second is to delete records, third is to update record, fourth field is to track the mobile, fifth is for viewing map, next is to point the map using map pointer and finally exit field.

Client Side:

Figure shows the snaps of how to enter the destination number to which the location string is to be sent. That mobile which receives the location string will be connected to server and the location information will be updated to database in the server.

Advantages and Disadvantages:

Advantages:

- Companies can track their employees.
- Can help tracking people in distress without much expectation from the other end
- Helps the marketing people in moving their products to the market faster, and to monitor assets and prevent inventory loss.

Disadvantages:

- Our project is intended to trace only when the mobile is switched on and has network.
- Manually starting of application at the client side.
- The client side Mobile when its application switched on the other application will be struck mobile is only for organization use.
- Whenever if any unwanted messages came manager to be manually deleted.

5. CONCLUSION

With the available facilities and infrastructure provided, we are successful in completing the project in the stipulated time. The project is not very expensive and as we have already mentioned it can be called as “poor man’s GPS” .The project can be very helpful to a person in trouble. The project can be used in any investigation related services or anywhere where knowing the whereabouts of a person becomes very important. But a major disadvantage of the project is intrusion into one’s private life.

Though this project, we learnt about Symbian operating system and java language and also concepts related to java language.

6. FUTURE ENHANCEMENT:

In future, application can be developed on running the application in hidden mode. In case of a company tracking its employee, the employee will not come to know that he is being tracked.

Technology can be developed to auto start the application. We can also tracking implement Location tracking system using GPS/GPRS technology which uses real time tracking system and Internet TCP/IP technology which can help for online personal tracking.

References:

i) Conference paper ;----> Location Based Services; Debarshi Bandyopadhyay; School of Electrical & Electronic Engineering; Nayang Technological University; Nanyang Avenue, Singapore 639798

ii) IEEE 802.16m-08/016, ----> “Call for Contributions on Project 802.16m System Description Document (SDD)”. – Comments on P802.16m SDD

iii)GEOLOCATION UPDATING SCHEMES FOR LOCATION AWARE SERVICES IN WIRELESS NETWORKS Amer Catovic Sirin Tekinay Dept. of Electrical and Computer Engineering New Jersey Institute of Technology Newark, New Jersey

iv)Support for Location Based Services Mei-Dai Chen, I-Kang Fu, Yih-Shen Chen, Kelvin Chou, and Paul ChengMediaTek Inc.Chung-Hsien Hsu and Kai-Ten Feng

v) Location Management in Mobile Computing Riky Subrata Albert Y. Zomaya

Parallel Computing Research Lab, Dept of Electrical and Electronic Engineering, The University of Western Australia, Western Australia 6907, I subra-r,zomaya

vi) Location Area Planning in Cellular Networks Using Simulated Annealing 1NETLAB, Department of Computer Engineering BUSIM Lab., Department of Electrical and Electronics Engineering Bogazici University, Bebek 80815 Istanbul, Turkey

vii) Intelligent Handover and Location Updating Control for a Third Generation Mobile Network Kuo-Hsing Chiang, Nirmala Shenoy, John Asenstorfer Mobile Communications Research Centre Institute for Telecommunications Research University of South Australia

viii) Leigh Edwards, Richard Barker, Developing Series 60 Applications.

ix) Patrick Naughton, The java handbook, 11th edition, TMH Publications.

x) Herbert Schildt, The Complete reference, 4th edition, TMH Publications.

Future Location-Based Experiences Professor Steve Benford School of Computer Science & IT The University of Nottingham

Bulk Transfers in Research Networks using Advance Reservations and Scheduling

Jagadish R.M.
M.Tech. 4th sem (CNE), BITM Bellary.
E-mail: rm.jagadish@gmail.com

T.R. Muhibur Rahman
Professor in Dept. of CSE. BITM, Bellary

Abstract— Research networks often require the transfer of large files with predictable performance. To meet the need, design admission control (AC) and scheduling algorithms for bulk transfer in e-science. Due to their small sizes, the research networks can afford a centralized resource management platform, specifies a start time and an end time for each bulk transfer job request. If admitted, the network guarantees to complete the transfer before the end time. However, there is flexibility in how the actual transfer is carried out, that is, in the bandwidth assignment on each allowed path of the on each time interval, and it is up to the scheduling algorithm to decide this. To improve the network resource utilization or lower the job rejection ratio, the network solves opmization problems in making AC and scheduling decisions. This design combines following elements into a optimization-based framework: advance reservations, multipath routing, and bandwidth reassignment via periodic reoptimization.

Index Terms—Admission control, advance reservation, bulk data transfer, e-science, grid computing, scheduling, multipath routing, network flow.

1 INTRODUCTION

The advance of communication and networking technologies, together with the computing and storage technologies is dramatically changing the ways how scientific research is conducted. A new term, e-science, has emerged to describe the “large-scale science carried out through distributed global collaborations enabled by networks, requiring access to very large scale data collections, computing resources, and high-performance visualization” [1]. Well quoted e-science (and the related grid computing [2]) examples include high-energy nuclear physics (HEP), radio astronomy, geoscience, and climate studies.

The need for transporting large volume of data in e-science has been well argued [3], [4]. For instance, the HEP data is expected to grow from the current petabytes (PB) (10^{15}) to exabytes (10^{18}) by 2012 to 2015. In particular, the Large Hadron Collider facility at CERN is expected to generate PB of experimental data every year, for each experiment. E-scientists routinely request schedulable high-bandwidth low-latency connectivity with known and knowable characteristics. Instead of relying on the public Internet, which has unpredictable service performance, national governments are sponsoring a new generation of optical networks to support e-science. Examples of such research and education networks include the Internet2-related National Lambda Rail [6] and Abilene [7] networks in the US, and CA*net4 [8] in Canada.

To meet the need of e-science, this design studies *admission control (AC)* and *scheduling algorithms* for high bandwidth data transfers (also known as jobs) in research networks. The results will not only advance the knowledge and techniques in that area but also compliment the protocol, architecture, and infrastructure projects currently underway in support of e-science and grid computing [9], [10], [11], by providing more efficient network resource reservation and management algorithms. AC and scheduling algorithms handle two classes of jobs, bulk data transfer and those that require a minimum bandwidth guarantee (MBG). Bulk transfer is not sensitive to the network delay but may be sensitive to the delivery deadline. It is useful for distributing high volumes of scientific data, which currently often relies on ground transportation of the storage media. The MBG class is useful for real-time rendering or visualization of data remotely. In this framework, the algorithms for handling bulk transfer also contain the main ingredients of those for handling the MBG class. For this reason, only focus on bulk transfer.

One distinguishing feature in this study is that each job request can be made in advance and can specify a start time and an end time. The reservation-based approach gives the network users more predictability and control over their work schedule and is deemed very useful by the e-science community [12]. If a job is admitted, as determined by the AC algorithm, the network guarantees that it will finish the data transfer for the job before the requested end time. The challenge is how to provide this guarantee while maintaining efficient utilization of the network resources and keeping the request rejection ratio low. (If a request is rejected, there are many possible follow-up scenarios depending on the design. The simplest is that the user of the request may modify the end time and resubmit the request. The resubmission process can be automated and repeated by the user-side software agent.)

The need for efficient network resource utilization is especially relevant in the context of advance reservations and large file sizes or long-lasting flows. As argued in [13], there is an undesirable phenomenon known as bandwidth fragmentation. The simplest example of bandwidth fragmentation occurs when the interval between the end time of one job and the beginning of another job is not long enough for any other job request. Then, the network or relevant links will be idle on that interval. If there are too many of these unusable intervals or if their durations are long, the job rejection ratio is likely to be high while the network utilization remains low. Over-provisioning the network capacity may not be the right solution due to the high cost, time delay, or other practical constraints.

The solution advocated in this design for reducing the job rejection ratio and increasing the network utilization efficiency is to bring in more flexibilities in how the data are transferred. The process of determining the manner of data transfer is known as scheduling. For instance, one can take advantage of the elastic nature of bulk data and have the network transferring the data at time-varying bandwidth instead of a constant bandwidth. Another example is to use multiple paths for each job. In order to achieve the greatest flexibilities, this design formulates the AC/scheduling problems as optimization problems. A centralized network controller is used to administer AC and scheduling, including solving the optimization problems. Different from the public Internet, research networks typically have far less than 1,000 core nodes in the backbone. Hence, it possible to use a centralized network controller for making AC and scheduling decisions, setting up network paths, and reserving the allocated bandwidth or optical circuits. One advantage of the centralized approach is that resource reservation and allocation decisions are based on a global view of the network and on all the job requests. It is possible to manage the network resources as a whole and make tradeoffs among all the jobs in the network. The result is greatly improved efficiency in network resource utilization.

Recently, some authors have begun to study AC and scheduling for bulk transfer with advance reservations [14], [15], [16], [17], [18], [19], [13], [20], [21]. Compared with these earlier studies, this work distinguishes itself for its comprehensiveness in bringing several important ingredients together under a single optimization framework with well-defined objectives. These include

1. Periodic AC for handling continuous arrivals of job requests rather than one-shot AC,
2. AC and scheduling for the whole network rather than for each link separately,
3. multipath routing,
4. time-varying bandwidth assignment for each job,
5. dynamic bandwidth reassignment at each AC/scheduling instance, which leaves more room to accept new requests, and
6. a novel time discretization scheme (i.e., the congruent time-slice structures) that allows the admission of new requests and bandwidth reallocation to existing jobs while not violating the end-time requirements of the existing jobs.

The rest of this design is organized as follows: The main technical contribution of this design is to describe a suite of algorithms for AC and scheduling (Section 2) . A highlight of our scheme is the introduction of non uniform time slices (Section 3), which can dramatically shorten the execution time of the AC and scheduling algorithms. The conclusion is drawn in Section 4.

2 ADMISSION CONTROL AND SCHEDULING ALGORITHMS

2.1 The Setup

The network is represented as a (directed) graph G = (V,E), where V is the set of nodes and E is the set of edges. The capacity of a link (edge) e ε E is denoted by Ce. Job requests arrive at the network following a random process. Each bulk transfer request i is a 6-tuple [Ai, si, di,Di, Si, Ei], where Ai is the arrival time of the request, si and di are the source and destination nodes, respectively, Di is the size of the file, and Si and Ei are the requested start time and end time, where Ai ≤ Si ≤Ei. In words, request i, which is made at time t=Ai, asks the network to transfer a file of size Di from node si to node di on the time interval [Si, Ei]. A bulk transfer request may optionally specify a minimum bandwidth and/or a maximum bandwidth. In practice, even more parameters can be added if needed, such as an estimated range for the demand size or for the end times when the precise information is unknown [22]. For ease of presentation, ignore these options. But, they usually can be incorporated into this optimization-based AC/scheduling framework by modifying the formulations of the optimization problems. The approach of using a centralized network controller has an advantage here for an evolving system, since, to accommodate new types of parameters or functions, the only necessary changes are at the central controller's software. The user-side software will be updated only if the user needs the new parameters or functions.

In the basic scheme, AC and scheduling are done periodically after every T time units, where T is a positive number. More specifically, at time instances kT, k = 1, 2. . ., the controller collects all the new requests that arrived on the interval [(k – 1)T, kT], makes the AC decision first, and then, schedules the transfer of all jobs. Both AC and scheduling must take into account the old jobs, i.e., those jobs that were admitted earlier but remain unfinished. The admission of new jobs is formulated as a feasibility problem subject to the constraint that the old jobs must retain their performance guarantee. To increase the admission rate, this step takes into account the possibility that the bandwidth of each old job on different routes can be reassigned. In the second step, scheduling, the network controller assigns the actual

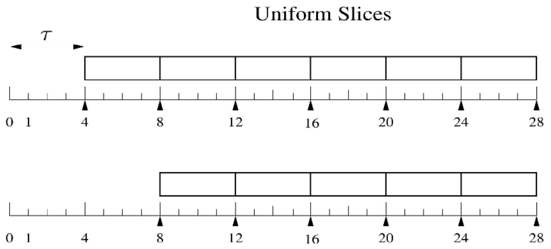


Fig. 1. Uniform time-slice structure.

bandwidth to all jobs in the system, including the old jobs, on the allowed paths so as to optimize a performance objective. Examples that are consider in this design are to minimize the worst case link utilization or to minimize an objective that encourages earlier completion of the jobs. The bandwidth assignment is time varying. The value of T should be small enough so that new job requests can be checked for admission and scheduled as early as possible. However, should be greater than the computation time required for AC and scheduling.

2.1.1 The Time-Slice Structure

At each scheduling instance, t = kT, the timeline from t onward is partitioned into time slices, i.e., closed intervals on the timeline, which are not necessarily uniform in size. The significance of the time slice is that the bandwidth (rate) assignment to each job is done at the slice level. That is, the bandwidth assigned to a particular path of a job remains constant for the entire time slice, but it may change from slice to slice.

A set of time slices, Gk, is said to be anchored at t = kT if all slices in Gk are mutually disjoint and their union forms an interval [t,t'] for some t'. The set {Gk} k=1..∞ is called a slice structure if each Gk is a set of slices anchored at t =kT, for t=1,…,∞.

Definition 1. A slice structure {Gk} $_{k=1..\infty}$ is said to be congruent if the following property is satisfied for every pair of positive integers, k and k', where k'>k ≥ 1. For any slice s'ε Gk' ,if s' overlaps in time with a slice s, s ε Gk, then s'⊆s.

In words, any slice in a later anchored slice collection must be completely contained in a slice of any earlier collection, if it overlaps in time with the earlier collection. Alternatively speaking, if slice s ε Gk' overlaps in time with Gk', then either s ε Gk' or s is partitioned into multiple slices all belonging to Gk' .

One example of a congruent slice structure is the uniform slices (US), where the timeline is divided into equal-sized time slices of duration T (coinciding with the AC/scheduling interval length). The set of slices anchored at any t = kT is all the slices after t. Fig. 1 shows the US at two time instances t =T and t = 2T. In this example, T = 4 time units. The arrows point to the scheduling instances. The two collections of rectangles are the time slices anchored at t =T and t = 2T, respectively. It is easy to check the congruent property of this slice structure.

Nearly all prior works that discretize the timeline use the US. The motivation for defining the more general concept of the congruent slice structure is as follows. Although easy to
understand, the US is not necessarily an ideal slice structure to use because, in our linear programming formulation of the AC and scheduling problems, the number of time slices is positively related to the number of variables, and in turn to the execution time of our algorithms. Here face a problem of covering a long enough segment of the timeline for advance reservations with a small number of slices, say 100. In this design, advocate a congruent slice structure with non-US sizes, the nested slices (NS). The NS contains different classes of time slices with exponentially (geometrically) increasing sizes. Suppose the current time t = kT is a scheduling instance. The timeline near t is divided into fine slices. The timeline away from t is divided into increasingly larger slices. Later, as time progresses, say to k'T, some coarse time slices will become close to the new current time, k'T, and will be divided into fine slices, which will belong to Gk'. The NS can cover a large portion of the timeline using a small number of slices without sacrificing performance (e.g., the job rejection ratio).

The AC and scheduling algorithms introduced in this design apply to any congruent slice structure. When a non- US structure is used, the congruent property is the key to the existence of algorithms that allow the network to keep the commitment to the old jobs admitted earlier while admitting new jobs. The reason is that, in solving the AC problem, the bandwidth allocation (on each allowed path of each job) on each time slice is assumed to be constant over the time slice. After a time slice is divided into finer slices at a later time, each old job is still admissible since one can assign to it the same constant bandwidth on the finer slices as before the division. This will be further explained in Section 3.

2.2 Admission Control

For each pair of nodes s and d, let the collection of allowable paths from s to d be denoted by Pk(s, d). In general, the set may vary with k. For each job i, let the remaining demand at time t = kT be denoted by Rk(i), which is equal to the total demand Di minus the amount of data transferred until time t.

At t = kT, let J ⊆ Jk be a subset of the jobs in the systems. Let fi(p,j) be the total flow (total data transfer) allocated to job i on path p, where p ε Pk(si,di), on time slice j, where j ε Lk. As part of the AC algorithm, the solution to the following feasibility problem is used to determine whether the jobs in J can all be admitted:

$$\mathbf{AC}(k, J)$$

$$\sum_{j=1}^{M_k} \sum_{p \in P_k(s_i, d_i)} f_i(p,j) = R_k(i), \quad \forall i \in J, \tag{3}$$

$$\sum_{i \in J} \sum_{\substack{p \in P_k(s_i, d_i) \\ p \ni e}} f_i(p,j) \leq C_e(j) LEN_k(j), \quad \forall e \in E, \forall j \in \mathcal{L}_k, \tag{4}$$

$$f_i(p,j) = 0, \quad j \leq I_k(\hat{S}_i) \text{ or } j > I_k(\hat{E}_i), \quad \forall i \in J, \forall p \in P_k(s_i, d_i), \tag{5}$$

$$f_i(p,j) \geq 0, \quad \forall i \in J, \forall j \in \mathcal{L}_k, \forall p \in P_k(s_i, d_i). \tag{6}$$

Equation (3) says that, for every job, the sum of all the flows assigned on all time slices for all paths must be equal its remaining demand. Equation (4) says that the capacity constraints must be satisfied for all edges on every time slice. Note that the allocated rate on path p for job i on slice j is fi(p, j)/LENk(j), where LENk(j) is the length of slice j. The rate is assumed to be constant on the entire slice. Here, Ce(j) is the remaining link capacity of link e on slice j. equation (5) is the start and end time constraint for every job on every path. The flow must be zero before the rounded start time and after the rounded end time.

Recall that assuming every job to be a bulk transfer for simplicity. If job i is of the MBG class and requests a minimum bandwidth Bi between the start and end times, then the remaining capacity constraint (3) will be replaced by the following MBG condition:

$$\sum_{p\in P_k(s_i,d_i)} f_i(p,j) \geq B_i, \quad \forall j \in \mathcal{L}_k. \qquad (7)$$

The AC/scheduling algorithms are triggered every T time units with the AC part before the scheduling part. AC examines the newly arrived jobs and determines their admissibility. In doing so, need to ensure that the earlier commitments to the old jobs are not broken. This can be achieved by adopting one of the following AC procedures:

1. Subtract-Resource (SR). An updated (remaining) net- work is obtained by subtracting the bandwidth assigned to old jobs on future time slices, from the link capacity. Then, determine a subset of the new jobs that can be accommodated in this remaining network. This method is helpful to perform quick admission tests. However, it runs the risk of rejecting new jobs that can actually be accommodated by reassigning the flows to the old jobs on different paths and time slices.
2. Reassign-Resource (RR). This method attempts to reassign flows to the old jobs. First, we cancel the existing flow assignment to the old jobs on the future time slices and restore the network to its original capacity. Then, determine a subset of the new jobs that can be admitted along with all the old jobs under the original network capacity. This method is expected to have a better acceptance ratio than SR. However, it is computationally more expensive because the flow assignment is computed for all the jobs in the system, both the old and the new.

2.3 Scheduling Algorithm

Given the set of admitted jobs, $J^a{}_k$, which always includes the old jobs, the scheduling algorithm allocates flows to these jobs to optimize a certain objective. Here consider two objectives, Quick-Finish (QF) and Load-Balancing (LB). Given a set of admissible jobs J, the problem associated with the former is

$$\begin{aligned} &\textbf{Quick-Finish}(k, J) \\ &\min \quad \sum_{j\in\mathcal{L}_k} \gamma(j) \sum_{i\in J} \sum_{p\in P_k(s_i,d_i)} f_i(p,j) \qquad (8) \\ &\text{subject to (3)-(6).} \end{aligned}$$

In the above, γ(j)is a weight function increasing in j, which is chosen to be γ(j)=j+1 in our experiments. One may use other weight functions, which will have different consequences. In this problem, the cost increases as time increases. The intention is to finish a job earlier rather than later, when it is possible. The solution tends to pack more flows in earlier slices but leaves the load light in later slices. The problem associated with the LB objective is

$$\begin{aligned} &\textbf{Load-Balancing}(k, J), \\ &\max \quad Z \qquad (9) \\ &\text{subject to} \sum_{j=1}^{M_k} \sum_{p\in P_k(s_i,d_i)} f_i(p,j) = Z R_k(i), \quad \forall i \in J \qquad (10) \\ &(4)\text{-}(6). \end{aligned}$$

Let the optimal solution be Z* and fi*(p,j) for all i, j, and p. The actual flows assigned are fi*(p,j)/Z*. Note that (10) ensures that fi*(p,j)/Z* satisfies (3). Also, Z*≥1must be true since J is admissible. Hence, fi*(P,J)/z*'s are a feasible solution to the AC(k, J) problem. The Load-Balancing (k, J) problem above is written in the maximizing concurrent throughput form. It reveals its LB nature when written in the equivalent minimizing congestion form. For that, make a substitution of variables, fi(p, j)<= fi(p, j)/Z, and let μ=1/Z.

We have

$$\begin{aligned} &\textbf{Load-Balancing-1}(k, J), \\ &\min \quad \mu \qquad (11) \\ &\text{subject to} \sum_{i\in J} \sum_{\substack{p\in P_k(s_i,d_i) \\ p\ni e \in p}} f_i(p,j) \leq \mu C_e(j) LEN_k(j), \\ &\qquad \forall e \in E, \forall j \in \mathcal{L}_k \qquad (12) \\ &(3), (5), \text{ and } (6). \end{aligned}$$

Hence, the solution minimizes the worst link congestion across all time slices in Lk.

The scheduling algorithm is to apply J = $J^a{}_k$ Quick-Finish(k,J) or Load-Balancing (k,J). This determines an optimal flow assignment to all jobs on all allowed paths and on all time slices. Given the flow assignment fi(p, j), the allocated rate on each time slice is denoted by xi(k, j)=fi(p, j)/LENk(j) for all j €Lk. The remaining capacity of each link on each time slice is given by

$$C_e(j) = \begin{cases} C_e - \sum_{i\in J_k^a} \sum_{\substack{p\in P_k(s_i,d_i) \\ p\ni e \in p}} x_i(p,j) & \text{if SR,} \\ C_e & \text{if RR.} \end{cases} \qquad (13)$$

2.4 Putting It Together: The AC and Scheduling Algorithm

In this section, integrate various algorithmic components and present the complete AC and scheduling algorithm, which is listed as Algorithm 1. In line 6, Algorithm 1 calls the AC algorithm, listed as Algorithm 2. It also summarizes the complete algorithm in the following.

Recall that the AC and scheduling algorithm runs at a discrete set of time instances, called the scheduling instances, kT , for k = 1, 2,.., On the interval ((k -1)T, kT], the system keeps track of the new requests arriving on that interval. It also keeps track of the status of the old jobs. If an old job is completed, it is removed from the system. If an old job is serviced on the interval, the amount of data transferred for that job is recorded. At t = kT, the steps described in Algorithm 1 are taken.

In line 1, the system discretizes the future timeline into time slices starting at t = kT. Gk is the collection of these future time slices. Let J_k, J^o_k, and J^n_k denote the collection of all jobs, the collection of old jobs, and the collection of new jobs known by the system at time t, respectively (line 2). These sets can be easily constructed by the system based on the new job requests that have arrived on the interval ((k -1)T, kT] and the record of the jobs already in the system. Since an old job may have already been serviced on ((k -1)T, kT] ,the amount of remaining data to be transferred for the job is computed (line 3). The start and end times of the jobs are rounded to be aligned on time-slice boundaries (lines 3 and 4). Line 5 determines the finite collection of future time slices that the AC and scheduling linear programs need to consider; none of the jobs currently known by the system has an end time falling beyond these slices. In line 6, the AC algorithm (Algorithm 2) is called, which in turn calls the AC (k, J) linear program. This linear program checks all links in the network and all affected future time slices whether there is enough bandwidth to accept some new jobs. Details are given in Algorithm 2. Once the set of admitted jobs is determined from theca step, one of the scheduling linear programs is called (line 7), depending whether the QF or LB objective is preferred. This linear program actually assigns the flow amount fi(p, j) for each admitted job i € Jka , over all paths for job i, and all time slices j € Lk. This may result in changing the bandwidth assignment made in earlier scheduling instances.

Algorithm 1. AC and Scheduling

```
1: Construct the anchored slice set at t = kT, Gk.
```

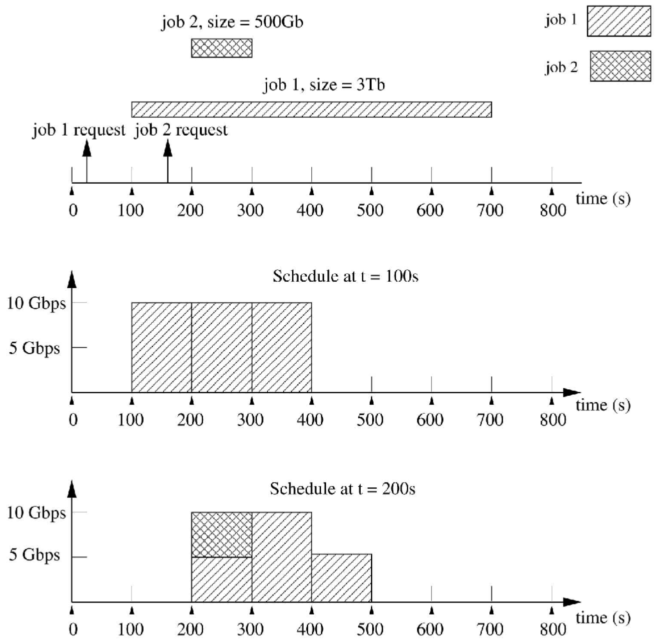


Fig. 2. An AC and scheduling example for a network with one link with a capacity 10 Gbps

```
2: Construct the job sets $J_k$, $J^o_k$ and $J^n_k$, which are the collection of all jobs, the collection of old jobs, and the collection of new jobs in the system, respectively.
3: For each old job i, update the remaining demand Rk(i) by subtracting from it the amount of data transferred for i on the interval ((k – 1)T, kT]. Round the start times as Si ^ =t.
4: For each new job l, let Rk ( l)=Dl. Round the requested start and end time according to (1) and (2), depending on whether the stringent or relaxed rounding policy is used.
// This produces the rounded start and end times, Sl ^and El^.
5: Derive Mk = Ik($\max_{i\ \varepsilon\ Jk}$ Ei^). This determines the finite collection of slices Lk={1; 2; . . .,Mk}, the first Mk slices of Gk.
6: Perform admission control as in Algorithm 2. This produces the list of admitted jobs $J^a_{k.}$
7: Schedule the admitted jobs as follows. If Quick-Finish (k, Jk a) is preferred, then solve Quick-Finish(k; $J^a_k$); else solve Load-Balancing (k, $J^a_k$ ).
// This yields the flow amount fi(p, j) for each admitted job i € $J^a_k$, over all paths for job i, and all time slices j € Lk.
8: Compute the remaining network capacity by (13).
```

Algorithm 2. AC—Step 6 of Algorithm 1

```
1: if Subtract-Resource is used then
2: Sequence the new jobs ($J^n_k$) in the system. Denote the sequence by (1, 2, . . .,m).
3: Apply binary search to the sequence of new jobs (1, 2,. . .,m). Find the last job j in the sequence so that the set of jobs J = (1, 2 . . . , j) is admissible by AC(k,J), under the remaining network capacity.
4: else if Reassign-Resource is used then
```

5: Sequence all the jobs (J k) in the system, so that the old jobs ($J^o{}_k$) are ahead of the new jobs ($J^n{}_k$).Denote the sequence of jobs by (1,2,. . . , l,l +1, . . .,m), where the first l jobs are the old jobs, followed by the new jobs.
6: Apply binary search to the subsequence of new jobs (l + 1,l +2, . . .,m). Find the last job j in the subsequence so that the set of jobs J =(1, 2, . . . , j)is admissible by AC(k,J), under the original network capacity.
7: end if
8: Return the admissible set, $J^a{}_k$ = J.

Finally, in Fig. 2 show a very simple example of theca and scheduling algorithms at work. The network has only one link with a capacity of 10 Gbps. The US is used and the AC/scheduling interval length is T=100 seconds. QF is used for scheduling. Fig. 2a shows two job requests, represented as two up-arrows at approximately 20 seconds and 160 seconds. The sizes of jobs 1 and 2 are 3 terabits and 500 gigabits, respectively. The requested start and end times are 100 seconds and 700 seconds for job 1 and 200 seconds and 300 seconds for job 2. In this case, job 1 is admitted at t = 100 seconds. Fig. 2b shows the schedule at t = 100 seconds. At t =200 seconds, job 2 is also admitted. Fig. 2c shows the schedule at t = 200 seconds. Note that, by t = 200 seconds, 1 terabits of data have already been transferred for job 1. Note also how the bandwidth assignment for job 1 is changed t =200 seconds, when compared to that at t = 100 seconds. This is in response to the admission of job 2, which has a stringent end time requirement. Furthermore, it can be seen that the bandwidth assignment for job 1 is time varying. network resources.

3 NON-UNIFORM SLICE STRUCTURE

As discussed in Sections 2.1.1 and 2.2, the number of time slices directly affects the number of variables in AC and scheduling linear programs, and in turn the execution time of our algorithms. It face a problem of covering a large enough segment of the timeline for advance reservations with a small number of slices, say about 100. In this section, design a new slice structure with non-US sizes. They contain a geometrically (exponentially) increasing subsequence, and therefore, are able to cover a large timeline with a small number of slices. The key is that, as time progresses, coarse time slices will be further divided into finer slices. The challenge is that the slice structure must remain congruent.

Recall that the congruent property means that, if a slice in an earlier anchored slice set overlaps in time with a later anchored slice set, it either remains as a slice, or is partitioned into smaller slices in the later slice set. The definition is motivated by the need for maintaining consistency in bandwidth assignment across time. As an example, suppose at time (k-1)T, a job is assigned a bandwidth x on a path on the slice jk-1.At the next scheduling instance t = kT, suppose the slice jk-1 is partitioned into two slices. Then understand that a bandwidth x has been assigned on both slices. Without the congruent property, it is likely that a slice, say jk, in the slice set anchored at kT cuts across several slices in the slice set anchored at (k -1)T. If the bandwidth assignments at (k-1)T are different for these latter slices, the bandwidth assignment for slice jk is not well defined just before the AC/scheduling run at time kT.

variants. They allow the coverage of a long segment of time for advance reservations with a small number of slices without compromising performance. They lead to reduced execution time of the AC/scheduling algorithms, thereby making it practical.

3.1 Nested Slice Structure

In the NS structure, there are l types of slices, known as level-i slices, i =1,2, . . . , l. Each level-i slice has a duration Δi, with the property that Δi=kiΔi+1, where ki > 1 is an integer, for i =1, . . . , l-1. Hence, the slice size increases at least geometrically as i decreases. For practical applications, a small number of levels suffice. here also require that, for such that Δi+1 ≤T<Δi, T is an integer multiple of Δi+1 and Δi is an integer multiple of T. This ensures that each scheduling interval contains an integral number of slices and that the sequence of scheduling instances does not skip any level-j slice boundaries, for 1 ≤ j ≤ i.

The NS structure can be defined by construction. At t = 0, the timeline is partitioned into level-1 slices. The first j1 level-1 slices, where j1 ≥ 1, are each partitioned into level-2 slices. This removes j1 level-1 slices but adds j1k1 level-2 slices. Next, the first j2 level-2 slices, where j2 ≤j1k1, are each partitioned into level-3 slices. This removes j2 level-2 slices but adds j2k2 level-3 slices. This process continues until, in the last step, the first jl-1 level-(l – 1) slices are partitioned into jl-1kl-1 level-l slices. Then, the first jl-1 level-(l – 1) slices are removed and jl-1k-1 level-l slices are added at the beginning. In the end, the collection of slices at t = 0 contains σ1≜jl-1kl-1 (≜means "defined as") level-l slices,…σl-1≜j1-2 kl-2-jl-1 level-(l -1) slices,. . . σ2≜j1k1- j2 level-2 slices, and followed by an infinite number of level-1 slices. The sequence of ji's must satisfy j2 ≤ j1k1, j3 ≤ j2k2, . . . ,jl-1 ≤ jl-2kl-2. This collection of slices is denoted by G0.

As an example, to cover a maximum of 30-day period, take Δ1=1 day, Δ2=1 hour, and Δ3=10 minutes. Hence, _1 ¼ 24 and _2 ¼ 6. The first two days are first divided into a total 48 1-hour slices, out of which the first
8 hours are further divided into 48 10-minute slices. The final slice structure has 48 level-3 (10-minute) slices, 40 level-2 (1-hour) slices, and as many level-1 (1-day) slices as needed, in this case, 28. The total number of slices is 116.

4 CONCLUSION

This study aims at contributing to the management and resource allocation of research networks for data-intensive e-science collaborations. The need for large file transfer and high-bandwidth, low-latency network paths is among the main requirements posed in such environments. The opportunities lie in the fact that research networks are generally much smaller in size than the public Internet, and hence afford a centralized resource management platform. This design combines the following novel elements into a cohesive framework of AC and flow scheduling: advance reservations for bulk transfer and minimum bandwidth guaranteed traffic, multipath routing, and bandwidth reassignment via periodic reoptimization.

To handle the start and end time requirement of advance reservations, as well as the advancement of time, we identify a suitable family of discrete time-slice structures, namely, the congruent slice structures. With such a structure, avoid the combinatorial nature of the problem and are able to formulate several linear programs as the core of AC and scheduling algorithms. Moreover, develop simple algorithms that can retain the performance guarantee for the existing jobs in the system while admitting new jobs. These algorithms apply to all congruent slice structures, which are fairly rich. In particular, describe the design of the NS structure and its variants. They allow the coverage of a long segment of time for advance reservations with a small number of slices without compromising performance. They lead to reduced execution time of the AC/scheduling algorithms, thereby making it practical.

References

[1] The U.K. Research Councils, http://www.research councils.ac.uk/ escience/, Feb. 2008.

[2] I. Foster and C. Kesselman, The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999.

[3] H.B. Newman, M.H. Ellisman, and J.A. Orcutt, "Data-Intensive e-Science Frontier Research," Comm. ACM, vol. 46, no. 11, pp. 68-77, Nov. 2003.

[4] J. Bunn and H. Newman, "Data-Intensive Grids for High-Energy Physics," Grid Computing: Making the Global Infrastructure a Reality, F. Berman, G. Fox, and T. Hey, eds., John Wiley & Sons, 2003.

[5] T. DeFanti, C.d. Laat, J. Mambretti, K. Neggers, and B. Arnaud, "TransLight: A Global-Scale LambdaGrid for e-Science," Comm. ACM, vol. 46, no. 11, pp. 34-41, Nov. 2003.

[6] National Lambda Rail, http://www.nlr.net, Feb. 2008.

[7] Abilene Network, http://www.internet2.edu/network/, Feb. 2008.

[8] CA_net4, http://www.canarie.ca/canet4/index.html, Feb. 2008.

[9] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, and A. Roy, "A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation," Proc. IFIP Seventh Int'l Workshop Quality of Service (IWQoS), 1999.

[10] The Globus Alliance, http://www.globus.org/, Feb. 2008.

[11] T. Lehman, J. Sobieski, and B. Jabbari, "DRAGON: A Framework for Service Provisioning in Heterogeneous Grid Networks," IEEE Comm. Magazine, Mar. 2006.

[12] T. Ferrari, Grid Network Services Use Cases from the e-Science Community, The Open Grid Forum, http://www.ogf.org/Public_ Comment_Docs/Documents/Jul-2007/draft-ggf-ghpn-net services-usecases-2-12.pdf, site last visited on Feb. 18, 2008. Dec. 2007.

[13] S. Naiksatam and S. Figueira, "Elastic Reservations for Efficient Bandwidth Utilization in LambdaGrids," The Int'l J. Grid Computing, vol. 23, no. 1, pp. 1-22, Jan. 2007.

[14] B.B. Chen and P.V.-B. Primet, "Scheduling Deadline-Constrained Bulk Data Transfers to Minimize Network Congestion," Proc. Seventh IEEE Int'l Symp. Cluster Computing and the Grid (CCGRID '07), May 2007.

[15] L. Marchal, P. Primet, Y. Robert, and J. Zeng, "Optimal Bandwidth Sharing in Grid Environment," Proc. IEEE High Performance Distributed Computing (HPDC '06), June 2006.

[16] K. Rajah, S. Ranka, and Y. Xia, "Scheduling Bulk File Transfers with Start and End Times," Computer Networks, vol. 52, no. 5, pp. 1105-1122, Apr. 2008.

[17] K. Munir, S. Javed, M. Welzl, and M. Junaid, "Using an Event Based Priority Queue for Reliable and Opportunistic Scheduling of Bulk Data Transfers in Grid Networks," Proc. 11th IEEE Int'l Multitopic Conf. (INMIC '07), Dec. 2007.

[18] K. Munir, S. Javed, M. Welzl, H. Ehsan, and T. Javed, "An Endto- End QoS Mechanism for Grid Bulk Data Transfer for Supporting Virtualization," Proc. IEEE/IFIP Int'l Workshop Endto- End Virtualization and Grid Management (EVGM '07), Oct. 2007.

[19] K. Munir, S. Javed, and M. Welzl, "A Reliable and Realistic Approach of Advance Network Reservations with Guaranteed Completion Time for Bulk Data Transfers in Grids," Proc. ACM Int'l Conf. Networks for Grid Applications (GridNets '07), Oct. 2007.

[20] R. Guerin and A. Orda, "Networks with Advance Reservations: The Routing Perspective," Proc. IEEE INFOCOM, 1999.

[21] L.-O. Burchard and H.-U. Heiss, "Performance Issues of
Bandwidth Reservation for Grid Computing," Proc. 15th Symp. Computer Architecture and High Performance Computing (SBACPAD), 2003.

[22] E. He, X. Wang, V. Vishwanath, and J. Leigh, "AR-PIN/ PDC: Flexible Advance Reservation of Intradomain and Interdomain Lightpaths," Proc. IEEE GLOBECOM, 2006.

OFDM and its Applications to 4G Cellular Network

Prajna K.S[1], Pooja K Gurukar[2]

7th Semester, Bachelor of Engineering
[*]Department of Electronics & Communication Engineering,
G.S.S.S.I.E.T.W, Mysore, Karnataka

[1]prajnasetty@gmail.com [2]poojagurukar@gmail.com

Abstract **- This paper investigates the effectiveness of Orthogonal Frequency Division Multiplexing (OFDM) as a modulation technique for wireless radio applications. The main aim is to assess the suitability of OFDM as a modulation technique for a wireless phone system. Several of the main factors affecting the performance of a wireless channel are shown . How OFDM effectively combats these is also shown.**

***Keywords*-** **OFDM, 4G, FFT-IFFT**

1. INTRODUCTION

With the rapid progress in telecommunications, more and more services are provided on the basis of broadband communications, such as video services and high-speed Internet. However, the basic problem of wireless access is that the available spectrum is too limited compared to the almost unlimited service requirement, just like cars jammed in crowded narrow paths.

Traditional phone networks (2G cellular networks) such as GSM, used mainly for voice transmission, are essentially circuit-switched. 2.5G networks, such as GPRS, are an extension of 2G networks, in that they use circuit switching for voice and packet switching for data transmission. Circuit switched technology requires that the user be billed by airtime rather than the amount of data transmitted since that bandwidth is reserved for the user. Packet switched technology utilizes bandwidth much more efficiently, allowing each user's packets to compete for available bandwidth, and billing users for the amount of data transmitted. Thus a move towards using packet-switched, and therefore IP networks, is natural. 3G networks were proposed to eliminate many problems faced by 2G and 2.5G networks, like low speeds and incompatible technologies (TDMA/CDMA) in different countries. Expectations for 3G included increased bandwidth: 128 Kbps in a car, and 2Mbps in fixed applications. In theory, 3G would work over North American as well as European and Asian wireless air interfaces. In reality, the outlook for 3G is neither clear nor certain. Most 3G mobile phone systems are proposing to use Code Division Multiple Access (CDMA) as their modulation technique. CDMA was found to perform poorly in a single cellular system, with each cell only allowing 7-16 simultaneous users in a cell. This low cell capacity of CDMA was attributed to the use of non-orthogonal codes used in the reverse transmission link, leading to a high level of inter-user interference. There is a concern that in many countries, 3G will never be deployed. This concern is grounded, in part, in the growing attraction of 4G wireless technologies.

4G characteristics

The defining features of 4G networks are listed below:

- **High Speed** –4G systems should offer a peak speed of more than 100Mbits per second in stationary mode with an average of 20Mbits per second when travelling.
- **High Network capacity** - Should be at least 10 times that of 3G systems. This will quicken the download time of a 10-Mbyte file to one second on4G, from 200 seconds on 3G, enabling high-definition video to stream to phones and create a virtual reality experience on high-resolution handset screens.
- **Fast/Seamless handover across multiple networks** – 4G wireless networks should support global roaming across multiple wireless and mobile networks.
- **Next-generation multimedia support -** The underlying network for 4G must be able to

support fast speed and large volume data transmission at a lower cost than today.

OFDM for Mobile Communication

OFDM represents a different system design approach. It can be thought of as combination of modulation and multiple access schemes that segment a communications channel in such a way that many users share it. Whereas TDMA segments are according to time and CDMA segments are according to spreading codes, OFDM segments are according to frequency. OFDM is the most suitable candidate to be used in 4G in order to achieve the above mentioned characteristics. In the next sections OFDM is discussed in detail and also how it eliminates most of the problems faced by present generation wireless communications.

2. PROPAGATION CHARACTERISTICS OF MOBILE RADIO-CHANNELS

Attenuation: It is the drop in the signal power when transmitting from one point to another. It can be caused by the transmission path length, obstructions in the signal path, and multipath effects. Any objects that obstruct the line of sight signal from the transmitter to the receiver can cause attenuation. Shadowing of the signal can occur whenever there is an obstruction between the transmitter and receiver. It is generally caused by buildings and hills, and is the most important environmental attenuation factor.

Multipath effects:

A) Rayleigh fading: In a radio link, the RF signal from the transmitter maybe reflected from objects such as hills, vehicles etc, this gives rise to multiple transmission paths at the receiver. Figure 1 show some of the possible ways in which multipath signals can occur.

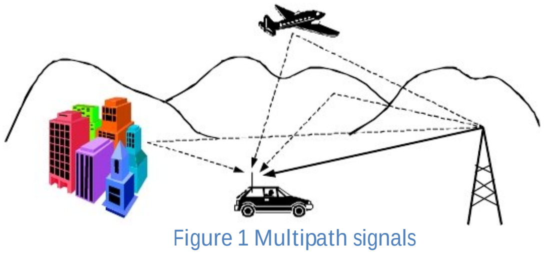


Figure 1 Multipath signals

Relative phase of the multiple reflected signals can cause instructive or destructive interference at the receiver. This is experienced over very short distances (typically at half wavelength distances), thus is given the term fast fading. These variations can vary from 10-30dB over a short distance. Figure 2 shows the level of attenuation that can occur due to the fading.

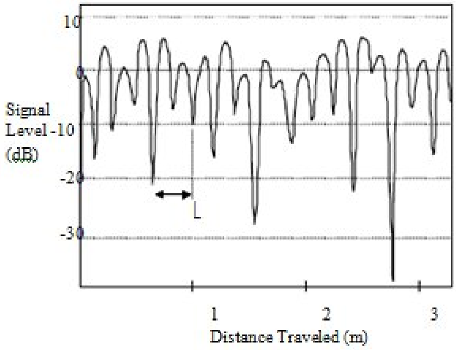


Figure 2 Typical rayleigh fading while the mobile unit is moving (for at 900 MHz)

B) Frequency selective fading: As the carrier frequency of a signal is varied, the magnitude of the change in amplitude will vary. The coherence bandwidth measures the separation in frequency after which two signals will experience uncorrelated fading.

- In **flat fading**, the coherence bandwidth of the channel is larger than the bandwidth of the signal. Therefore, all frequency components of the signal will experience the same magnitude of fading.
- In **frequency-selective fading**, the coherence bandwidth of the channel is smaller than the bandwidth of the signal. Different frequency components of the signal therefore experience decorrelated fading.

Since different frequency components of the signal are affected independently, it is highly unlikely that all parts of the signal will be simultaneously affected by a deep fade. Modulation schemes such as OFDM and CDMA are well-suited to employing frequency diversity to provide robustness to fading. OFDM divides the wideband signal into many slowly modulated narrowband sub carriers, each exposed to flat fading rather than frequency selective fading. This can be combated by means of error coding, simple equalization or adaptive bit loading.

C) Intersymbol interference: ISI is usually caused by multipath propagation or the inherent non linear frequency response of a channel causing successive symbols to blur together. The presence of ISI in the system introduces errors in the decision device at the receiver output. Inter-symbol interference is avoided by introducing a guard interval between the symbols.ISI is usually studied with the help of eye pattern which shows the interference pattern of the signals.

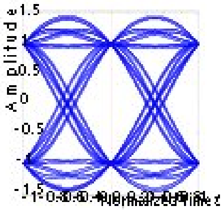


Figure 3 Eye pattern showing effects of ISI

D) Delay spread: Delay spread is the time spread between the arrival of the first and last significant multipath signal seen by the receiver. In a digital system, the delay spread can lead to inter-symbol interference. This is due to the delayed multipath signal overlapping with the following symbols. This can cause significant errors in high bit rate systems, especially when using time division multiplexing (TDMA). Figure 4 shows the effect of inter-symbol interference due to delay spread on the received signal. As the transmitted bit rate is increased the amount of inter-symbol interference also increases.

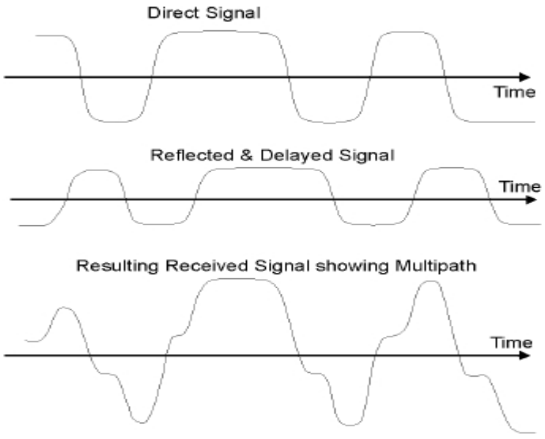

Direct Signal
Time
Reflected & Delayed Signal
Time
Resulting Received Signal showing Multipath
Time

Figure 4 multipath Delay spread

Doppler Effect: When a wave source and a receiver are moving relative to one another the frequency of the received signal will not be the same as the source. When they are moving toward each other the frequency of the received signal is higher then the source, and when they are approaching each other the frequency decreases. This is called the Doppler Effect. An example of this is the change of pitch in a car's horn as it approaches then passes by. This effect becomes important when developing mobile radio systems. The amount the frequency changes due to the Doppler Effect depends on the relative motion between the source and receiver and on the speed of propagation of the wave.

3. HISTORY OF OFDM

In a classical parallel data system, the total signal frequency band is divided into N non overlapping frequency sub channels. Each sub-channel is modulated with a separate symbol and then the N sub channels are frequency-multiplexed. It seems good to avoid spectral overlap of channels to eliminate inter-channel inter-ference. However, this leads to inefficient use of the available spectrum. To cope with the inefficiency, the ideas proposed from the mid-1960s were to use parallel data and FDM with overlapping sub-channels, in which, each carrying a signaling rate b is spaced b apart in frequency to avoid the use of high-speed equalization and to combat impulsive noise and multipath distortion, as well as to fully use the available bandwidth.

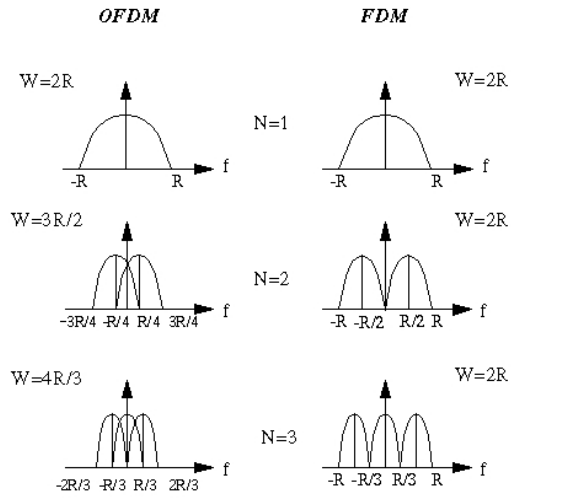


Figure 5 Concept of OFDM signal: Orthogonal multicarrier technique versus conventional multicarrier technique

Figure 5 illustrates the difference between the conventional non-overlapping multicarrier technique and the overlapping multicarrier modulation technique. As shown in Figure 5, by using the overlapping multicarrier modulation technique, we save almost 50% of bandwidth. To realize the overlapping multicarrier technique, however we need to reduce crosstalk between subcarriers, which means that we want orthogonality between the different modulated carriers.

It is possible, to arrange the carriers in an OFDM signal so that the sidebands of the individual carriers overlap and the signals are still received without adjacent carrier interference. To do this, the carriers must be mathematically orthogonal. The receiver acts as a bank of demodulators, translating each carrier down to DC, with the resulting signal integrated over a symbol period to recover the raw data. If the other carriers all beat down the frequencies that, in the time domain, have a whole number of cycles in the symbol period T, then the integration process results in zero contribution from all these other carriers. Thus, the carriers are linearly independent (i.e., orthogonal) if the carrier spacing is a multiple of 1/T.

Qualitative description of OFDM

In multimedia communication, a demand emerges for high-speed, high-quality digital mobile portable reception and transmission. A receiver has to cope with a signal that is often weaker than desirable and that contains many echoes. Simple digital systems do not work well in the multipath environment.

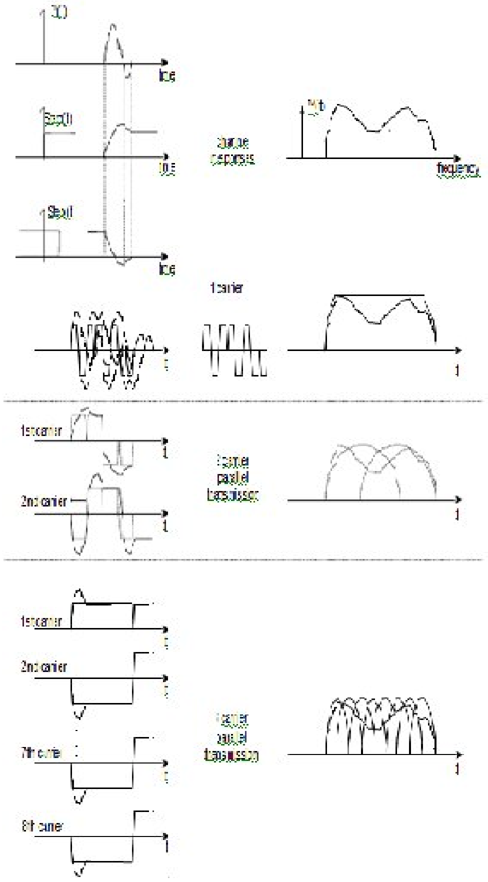


Figure 6

The effect of adopting a multicarrier system. For a given overall data rate, increasing the number of carriers reduces the data rate that each individual carrier must convey, and hence (for a given modulation system) lengthens the symbol period. This means that the inter-symbol interference affects a smaller percentage of each symbol as the number of carriers and hence the symbol period increases. For example, on the picture is shown an 8 bit long part of a data sequence. For a single carrier system, the responses of individual bits are overlapping, thus creating ISI. Multicarrier system is robust against these physical effects.

In OFDM, the data is divided among large number of closely spaced carriers. The entire bandwidth is filled from a single source of data. Instead of transmitting in serial way, data is transferred in a parallel way. Only a small amount of the data is carried on each carrier, and by this lowering of the bitrate per carrier (not the total bitrate), the influence of inter-symbol interference is significantly reduced.

It is an important part of the OFDM system design that the bandwidth occupied is greater than the correlation bandwidth of the fading channel. A good understanding of the propagation statistics is needed to ensure that this condition is met. Then, although some of the carriers are degraded by multipath fading, the majority of the carriers should still be adequately received. OFDM can effectively randomize burst errors caused by Rayleigh fading, which comes from interleaving due to parallelization. So, instead of several adjacent symbols being completely destroyed, many symbols are only slightly distorted. Because of dividing an entire channel bandwidth into many narrow subbands, the frequency response over each individual subband is relatively flat. Since each sub-channel covers only a small fraction of the original bandwidth, equalization is potentially simpler than in a serial data system. A simple equalization algorithm can minimize mean-square distortion on each sub-channel, and the implementation of differential encoding may make it possible to avoid equalization altogether. This allows the precise reconstruction of majority of them, even without forward error correction (FEC). In addition, by using a guard interval the sensitivity of the system to delay spread can be reduced.

4. OFDM IMPLEMENTATION

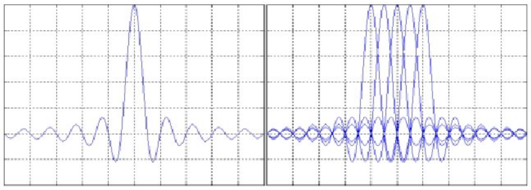


Figure 6 spectra of (a) an OFDM subchannel and (b) an OFDM signal

Much of the research focuses on the high efficient multicarrier transmission scheme based on "orthogonal frequency" carriers. In 1971, Weinstein and Ebert applied the discrete Fourier transform (DFT) to parallel data transmission systems as part of the modulation and demodulation process. Figure 6(a) shows the spectrum of the individual data of the subchannel. The OFDM signal, multiplexed in the individual spectra with a frequency spacing b equal to the transmission speed of each subcarrier, is shown in Figure 6(b). Figure 6 shows that at the center frequency of each subcarrier, there is no crosstalk from other channels. Therefore, if we use DFT at the receiver and calculate correlation values with the center of frequency of each subcarrier, we recover the transmitted data with no crosstalk. In addition, using the DFT-based multicarrier technique, frequency-division multiplex is achieved not by band pass filtering but by baseband processing. Moreover, to eliminate the banks of subcarrier oscillators and coherent demodulators required by frequency-division multiplex, completely digital implementations could be built around special-purpose hardware per-forming the fast Fourier transform (FFT), which is an efficient implementation of the DFT. Recent advances in very-large-scale integration (VLSI) technology make high-speed, large-size FFT chips commercially affordable. Using this method, both transmitter and receiver are implemented using efficient FFT techniques that reduce the number of operations from N2 in DFT down to N log N. In the 1980s, OFDM was studied for high-speed modems, digital mobile communications, and high-density recording. One of the systems realized the OFDM techniques for multiplexed QAM using DFT and by using pilot tone, stabilizing carrier and clock frequency control and implementing trellis coding are also implemented. Moreover, various-speed modems were developed for telephone networks. In the 1990s, OFDM was exploited for wideband data communications over mobile radio FM channels, high-bit-rate digital subscriber lines (HDSL; 1.6 Mbps), asymmetric digital subscriber lines (ADSL; up to 6 Mbps), very-high-speed digital subscriber lines (VDSL; 100 Mbps), digital audio broadcasting (DAB), and high-definition television (HDTV) terrestrial broadcasting.

General Structure

The basic principle of OFDM is to split a high-rate DataStream into a number of lower rate streams that are transmitted simultaneously over a number of subcarriers. The relative amount of dispersion in time caused by multipath delay spread is decreased because the symbol duration increases for lower rate parallel subcarriers. The other problem to solve is the intersymbol interference, which is eliminated almost completely by introducing a guard time in every OFDM symbol. This means that in the guard time, the OFDM symbol is cyclically extended to avoid intercarrier interference. An OFDM signal is a sum of subcarriers that are individually modulated by using phase shift keying (PSK) or quadrature amplitude modulation (QAM). The symbol can be written as:

$$s(t) = \mathrm{Re}\left\{ \sum_{i=-\frac{N_S}{2}}^{\frac{N_S}{2}-1} d_{i+N_S/2} \exp(j2\pi(f_c - \frac{i+0.5}{T})(t-t_s)) \right\}, t_S \le t \le t_S + T$$

$$s(t) = 0, t < t_S \ and \ t > t_S + T$$

Where:
NS is the number of subcarriers
T is the symbol duration
fc is the carrier frequency

The equivalent complex baseband notation is given by:

$$s(t) = \sum_{i=-\frac{N_S}{2}}^{\frac{N_S}{2}-1} d_{i+N_S/2} \exp(j2\pi \frac{i}{T}(t-t_s)) \quad , t_S \le t \le t_S + T \qquad (1)$$

$$s(t) = 0, t < t_S \ and \ t > t_S + T$$

In this case, the real and imaginary parts correspond to the in-phase and quadrature parts of the OFDM signal. They have to be multiplied by a cosine and sine of the desired frequency to produce the final OFDM signal. Figure 7 shows the block diagram for the OFDM modulator.

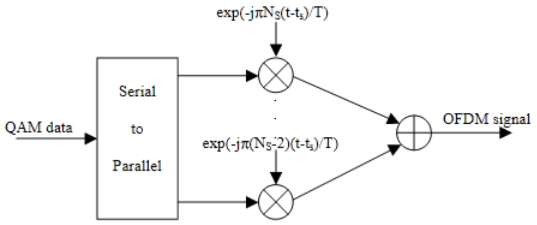


Figure 7 OFDM Modulator

The complex baseband OFDM signal defined the equation (1) is the inverse Fourier transform of Ns QAM input symbols. The time discrete case is the inverse discrete Fourier transform. In practice, this transform can be implemented very efficiently by the inverse fast Fourier transform (IFFT). The IFFT drastically reduces the amount of calculations by exploiting the regularity of the operations in the IDFT.

Basic block diagram of OFDM

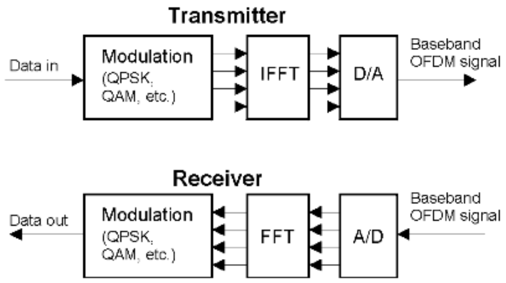


Figure 8 basic FFT, OFDM Transmitter and receiver

Adding a guard period to OFDM

One of the most important properties of OFDM transmissions is its high level of robustness against multipath delay spread. This is a result of the long symbol period used, which minimizes the inter-symbol interference.

The level of multipath robustness can be further increased by the addition of a guard period between transmitted symbols. The guard period allows time for multipath signals from the pervious symbol to die away before the information from the current symbol is gathered. The most effective guard period to use is a cyclic extension of the symbol. If a mirror in time, of the end of the symbol waveform is put at the start of the symbol as the guard period, this effectively extends the length of the symbol, while maintaining the orthogonality of the waveform. Using this cyclic extended symbol the samples required for performing the FFT (to decode the symbol), can be taken anywhere over the length of the symbol. This provides multipath immunity as well as symbol time synchronization tolerance.

5. KEY BENEFITS OF OFDM

Bandwidth efficiency: A key aspect of all high speed communications lies in the bandwidth efficiency. This is especially important for wireless communications where all current future devices are expected to share an already crowded range of carrier Frequencies. For wireless networks to remain profitable, it is necessary to achieve maximum bandwidth efficiency. Figure 9 shows the bandwidth efficiency of OFDM compared to FDM.

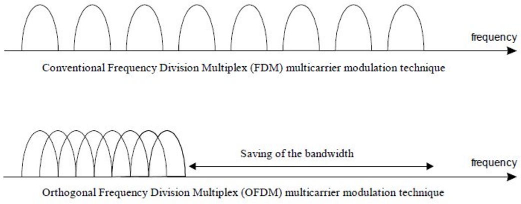

frequency
Conventional Frequency Division Multiplex (FDM) multicarrier modulation technique
Saving of the bandwidth
frequency
Orthogonal Frequency Division Multiplex (OFDM) multicarrier modulation technique

Figure 1 Comparison of bandwidth utilization by FDM and OFDM

Multipath fading

Each subcarrier in an OFDM signal has a very narrow bandwidth, thus the symbol rate is very low. This results in the signal having high tolerance to multi path delay spread, reducing any significant intersymbol interference.

RF interference

To combat the effects of random signal noise, which can prevent the receiver from fully recovering the signal, a spreading forward error correcting code is applied to the signal before transmission this has the effect of spreading the symbols over many frequencies, white maintaining the ability to recover the symbols even if some carriers are subjected to noise.

LEADING – EDGE MOBILE OFDM TECHNOLOGIES

Unlike most existing forms of wireless access, including 3G technologies, conventional wireless systems have been designed primarily at the physical layer. To address the unique demands posed by mobile users of high speed data applications, new air interface must be designed and optimized across all the layers of the protocol stack, including networking layers. A prime example of this is flash-OFDM. It is a system level technology that exploits the unique physical properties of OFDM, enabling significant higher layer advantages that contribute to very efficient packet transmission in a cellular network.

Packet switched Air interface: The telephone network, designed basically for voice is an example of circuit switched systems. Circuit switched systems exist only at the physical layer that uses the channel resource to create a bit pipe. Circuit switched systems are very inefficient for burst data traffic. Packet switched systems on the other hand, are very efficient for data traffic but require control layers in addition to the physical layer that creates the bit pipe. The internet is the best example for packet switched interface network. Because all conventional cellular wireless systems are designed for circuit switched voice, they are designed and optimized at the physical layer flash OFDM is a packet-switched designed for data and is optimized across the physical MAC, Link and network layers.

Disadvantages of OFDM

- OFDM signal is contaminated by non-linear distortion of transmitter power amplifier, because it is a combined amplitude-frequency modulation (it is necessary to maintain linearity)

- OFDM is very sensitive to carrier frequency offset caused by the jitter of carrier wave and Doppler Effect caused by moving of the mobile terminal.

- At the receiver, it is very difficult to decide the starting time of the FFT symbol

6. Conclusion

OFDM has long been studied and implemented to combat transmission channel impairments. Its applications have been extended from high frequency radio communications to telephone networks, digital audio broadcasting and terrestrial broadcasting of digital television. The advantages of OFDM, especially in the multipath propagation, interference and fading environment, make the technology a promising alternative in digital communications including mobile multi-media. Communications research and current development of OFDM around the world will certainly provide us with valuable findings in theory and implementation. Further studies should be conducted on the synchronization of OFDM signal, power demand, counter-measures against frequency offset, fading and multiple access.

REFERENCES

[1] Dušan Matiæ “*OFDM as a possible modulation technique for multimedia applications in the range of mm waves*”2nd edition, October ’98.

[2] Eric Lawrey “*The suitability of OFDM as a modulation technique for wireless telecommunications, with a CDMA comparison*” .2nd edition, October 97.

[3] Dr. Jayakumari.J”*MIMO-OFDM for 4G Wireless Systems*” International Journal of Engineering Science and Technology, vol2 (7), 2010

[4] Anibal Luis intini ”*Orthogonal frequency division multiplexing for wireless communications*” December 2000.

[5] William Stallings ”*Introduction to wireless telecommunications*” Pearson Education

[6] Simon Haykin ”*Digital Communications*” Wiley India Edition 2009.

[7] William Stallings ”*Wireless Communications and Networks*” Pearson Education 2006.

Load Balancing Technique for NGEO Satellite IP Networks

A. Vinay
M. Tech Student, BITM, Bellary
E-Mail: vinayarvi2007@rediffmail.com

Prof. Phani Ram Prasad
Dept. of CSE, BITM, Bellary
E-Mail: vinayarvi2007@rediffmail.com

ABSTRACT: Non-geostationary (NGEO) satellite communication systems offer an array of advantages over their terrestrial and geostationary counterparts. Due to several geographical and climatic constraints, some Inter-Satellite Links (ISLs) are expected to be heavily loaded with data packets while others remain underutilized. Such scenario obviously leads to congestion of the heavily loaded links. It ultimately results in buffer overflows, higher queuing delays, and significant packet drops. To guarantee a better distribution of traffic among satellite constellation. By using the proposed scheme is "Explicit Load Balancing" (ELB) scheme, in this an explicit exchange of information on congestion status among neighboring satellites. In turn, a satellite notifies its congestion status to its neighboring satellites. When it is about to get congested, it requests its neighboring satellites to decrease their data forwarding rates by sending them a self status notification signaling message. In response, the neighboring satellites search for less congested paths that do not include the satellite in question and communicate a portion of data, primarily destined to the satellite, via the retrieved paths. This operation avoids both congestion and packet drops at the satellite, it also ensures a better distribution of traffic over the entire satellite constellation.

1. INTRODUCTION

During the recent advances in terrestrial communication technologies, the ever growing community of Internet users poses serious challenges to current terrestrial networks. Terrestrial networks are expected to provide a plethora of bandwidth intensive services, with different Quality of Service (QoS), to a potential number of users, dispersed over extensively wide areas and requiring different degrees of mobility. Network technicians and telecommunication operators have envisaged optical-fiber networks and have considered temporary solutions such as Asynchronous Digital Subscriber Line (ADSL) and High-rate DSL (HDSL) technologies. However, as the demand for advanced

multimedia services is growing in terms of both the number of

users and the services to be supported, applying such solutions to bridge the last mile between local service providers and end-terminals will require an immense investment in terms of time, infra-structure, and human resources. The efficiency of satellite-based broadband services is strongly remarkable in remote zones and low-density population areas.

The key technologies required to support broadband communications over satellite systems have been already developed. That is satellite return channels and onboard processing technologies, satellites are now able to provide full two-way services to and from earth terminals. Additionally, several techniques are on-demand onboard switching has been proposed to make efficient use of satellites capacity. Unlimited connectivity can be accordingly guaranteed. Based on these advancements and on-going enhancements in satellite communications, it is now possible to design and implement satellite based communication systems for high bit rate services.

Satellite communication systems exhibit unique features and offer an array of advantages over traditional terrestrial networks. In inherent multicast capabilities and flexible deployment features, they are able to provide coverage to extensive geographic areas and to interconnect among remote terrestrial networks (e.g., islands). They can be also used as an efficient alternative to damaged terrestrial networks to recover from natural disasters. A significant number of satellite communication constellations have been thus proposed using Geostationary (GEO), Medium Earth Orbit (MEO) or Low Earth Orbit (LEO) satellites.

In long propagation delays, GEO system cause mobile terminals in high latitude regions to experience frequent cut-offs of propagation signals by tall buildings, trees or mountains. To provide global communication with reasonable latency and low terminal power requirements, constellations made of multi Non-Geostationary (NGEO) satellites (e.g. LEO and MEO). Due to geographical and climatic constraints, the community of future NGEO satellite users will exhibit a significant variance in its density over the globe. In satellites covering urban areas which dense with users will be more congested than satellites serving rural regions this density variance, along with the highly dynamic feature of NGEO constellations, will yield a scenario where some satellite links are get congested while others are underutilized. The absence of an efficient routing algorithm that takes into account the traffic distribution, this unfair distribution of network traffic will lead to significant queuing delays and large number of packet drops at the congested satellites. Obviously such performance will lead to poor throughput and will ultimately affect the Quality of Service (QoS) credibility of the entire system.

In ELB, satellite continuously monitors its queue size to determine its state which may be free, fairly-busy, or busy. A change in the state of a satellite is immediately notified to its neighboring satellites via a Self-State Advertisement packet. As a consequence, the cost of the links between the busy satellite and its neighbors is then increased. To avoid an imminent congestion, a satellite with high traffic load requests its neighboring satellites of forward a portion of data, originally destined to travel through the satellite, via alternative paths that do not involve the satellite. The ELB scheme therefore alters the traffic sending rate of neighboring nodes of the satellite in Question before it gets congested. Since minimum cost links are preferred, packets will be routed on the least loaded links and busy links will therefore have less packets in the queues.

The remainder of this paper is organized as follows. Section II

presents a detailed survey on the state-of-art in the context of routing protocols for multi-hop NGEO satellite constellations.

The key design philosophy and distinct features that were incorporated in the proposed scheme are described in Section III. The dynamic settings of its parameters are also discussed in the section. The performance of ELB is evaluated and compared to other schemes in Section IV. The paper concludes in Section V with a summary recapping the main advantages and achievements of the proposed architecture.

2. RELEATED WORK

Inter-Satellite Links (ISLs) in multi-hop NGEO constellations provides more flexibility, it leads to complex dynamic routing. The routing complexity becomes more substantial as NGEO satellites change their coverage areas on the Earth surface due to their continuous motion, and accordingly have to transmit different amounts of traffic load. This ultimately results in an unbalanced distribution of the total traffic cover the entire constellation.

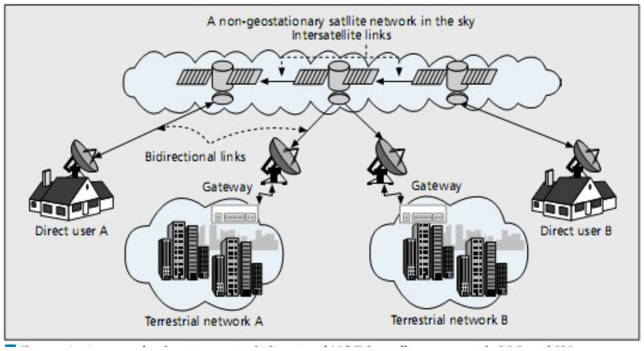


Figure 2.1. A non-geostationary satellite network in the sky with ISL

In order to route traffic over dynamic satellite constellations, several strategies have been proposed.

Dynamic Virtual Topology Routing (DVTR) and Virtual Node (VN) are the best known concepts.

In general, a communication delay consists of both propagation and queuing delays. In that propagation delay is the dominating factor in the communication delay. Thus focused used on developing routing mechanisms that find minimum propagation delay paths with minimal hop count for communication. For maximizing throughput in LEO satellite networks is proposed. The proposed paths strategy consists of an algorithm that finds the minimum hop path using Dijkstra's algorithm and a scheduling mechanism that favors packets destined to near by destinations. While the scheduling mechanism maximizes the throughput, it yields poor fairness against packets destined to distant destinations.

On board distributed routing protocol that selects the next hop based on minimization of the remaining geographic distance to the destination. In other words, depending on the geographic information embedded in the addresses, each satellite forwards the packet to its neighbor that most reduces the distance to the destination. Another vision for path minimization consists in favoring ISLs with higher life time to reduce the additional delays that may be caused by ISL handovers. For better QoS conditions, the routing algorithm should distribute the traffic in a balanced way over appropriate ISLs between end-terminals. This operation can be performed in either a central or a distributed manner. They also introduce extra signaling delays as the gathered information takes significant time, due to high propagation delays, till it is distributed in the constellation. Therefore, it does not accurately reflect the actual condition of the network.

Then a Minimum Flow Maximum Residual (MFMR) routing protocol is proposes where the minimum-hop path with the minimum number of flows is selected. One of the main drawbacks of the protocol consists in the fact that it implies knowledge of the flows over the constellation and does not consider the case where the flows count increases along the selected path. This would lead to the congestion of the chosen MFMR paths and ultimately results in unfavorable performance. A Probabilistic Routing Protocol (PRP) is proposed. The PRP scheme uses a cost metric as a function of time and traffic load. The traffic load is assumed to be location homogeneous. The major pitfall of the protocol consists in this assumption as it is far away from being realistic. In turn, newly coming traffic can easily congest the chosen PRP path and leave other resources underutilized. In Compact Explicit Multipath Routing (CEMR) algorithm based on a cost metric that involves both propagation and queuing delays. At a given satellite, the queuing delay is predicted by monitoring the number of packets in the outgoing queue of the satellite over a time interval. It is assumed that the network state over each time interval is updated be for a routing calculation is carried out. While the used cost metric gives a good insight about the queuing delay that may be experienced by a packet at a given satellite, it does not reflect the congestion state of the next hop, nor does it estimate the queuing delay a packet may experience there. Taking these remarks in to account, we are developed a routing strategy where packet drops are avoided and traffic burden is efficiently and fairly distributed among all participating satellites on the working of TCP protocol with ELB scheme is used.

3. OPERATIONAL OVERVIEW OF THE ELB SCHEME

This section presents a detailed description of the proposed ELB scheme, the rationale behind the setting of its parameters and its interactions with service differentiating mechanisms Adequate measures to cope with the issue of packet reordering In connection-oriented protocols, such as TCP, are also portrayed. For the sake of simplicity, let consider the case of a single traffic class.

The envisioned multi-hop NGEO satellite constellation satellites with on-board processing capabilities consists of S satellite with on board processing capabilities, uniformly distributed over N orbits, forming a mesh network topology .Each satellite is able to setup a maximum of with its neighboring satellites. These links are called Inter satellite Links (ISLs). Satellites are assumed to be aware of neighboring satellites.

The state of a satellite is marked as Free State (FS) when the queue ratio of its current queue occupancy to the total queue size Qr, is inferior to a pre-defined threshold. The satellite is considered to be in a Fairly Busy State (FBS) when its queue ratio is between the threshold and another predetermined threshold .The satellite is considered to be in a Busy State (BS) if its queue ratio exceeds the threshold.

It should be stressed out that warning messages and BSA packets do not incur any significant over head, in terms of neither bandwidth consumption nor scalability, as they are broadcasted merely upon a state transition and only to the neighboring satellites(maximum M satellite)Not over the entire connection path. The next sub sections portray the setting procedure of the queue Ratio thresholds, and the TRR parameter

A. Setting of Queue Ratio Thresholds

The key philosophy behind an optimum setting of Is to reflect the packet discarding probability in these two parameters so as to avoid packet drops when a satellite is running under heavy loads. Let and denote the total input and output traffic rates at a given satellite, respectively. Let note the total length of its queue and the occupancy of its queue at time t, respectively. Assuming that the input and output traffic rates constant over a short period of time, the elapsing time till a packet drop occurs can be expressed as follows:

$$\delta_d = \frac{(Q_l - q(t)) \cdot P_{\text{avg}}}{I - O}$$

Where is the average packet size. If the satellite is assumed to monitor its queue occupancy every interval time, it needs a maximum of time to notify its neighboring satellites of a possible packet drop, where denotes the ISL delay. In this case, two scenarios can be envisioned:

$\delta_d \le \delta + d$Packet drops happen before neighboring satellites are notified and adequate measures are taken. In this case, the packet dropping probability is one

$\delta_d > \delta + d$: In this case, if the satellite keeps receiving and transmitting data at the same rates over a number of monitoring intervals, packet drops happen only once during times of monitoring operations. The packet dropping probability is thus in both cases, the packet dropping probability can be expressed as

$$p = \text{Min}\left(1, \frac{\delta + d}{\delta_d}\right).$$

To refect the packet dropping probability in the setting of, we set to

$$\beta = 1 - p.$$

The rationale behind this setting is that when traffic load gets heavy and gets higher values, should be set to small values so as the satellite would quickly transit to the busy state and neighboring satellites would be promptly requested to reduce their sending rates to avoid possible congestion and packet drops. In this regard, it should be noted that setting the monitoring interval to high values may lead to significant packet drops. Indeed, in case of long monitoring intervals, by the time a satellite monitors its queue length, congestion may have already occurred and packet drops become then inevitable. In such case, the packet dropping probability will be equal to one consequently, will be always set to zero. As a remedy to this issue, the satellites are assumed to monitor their queues in a real is set to 1ms throughout this paper.

B. Setting of the Traffic Reduction Ratio

The main objective behind the setting of the TRR parameter is to allow satellites to return back to their free state and reside in this state for at least a predetermined period of time. Let and denote the total rate of traffic coming from terminals within the coverage area of a satellite and that of traffic coming from neighboring satellites, respectively(Fig.1).When the satellite shifts to the busy state, it requests neighboring satellite to reduce their sending rates. By the time the BSA signaling packet reaches the neighboring satellites, the queue occupancy of the satellite is

$$q(t_{\mathrm{BSA}}) = \mathrm{Min}\left(Q_l \cdot \beta + \frac{d \cdot (I_s + I_t - O)}{P_{\mathrm{avg}}}, Q_l\right).$$

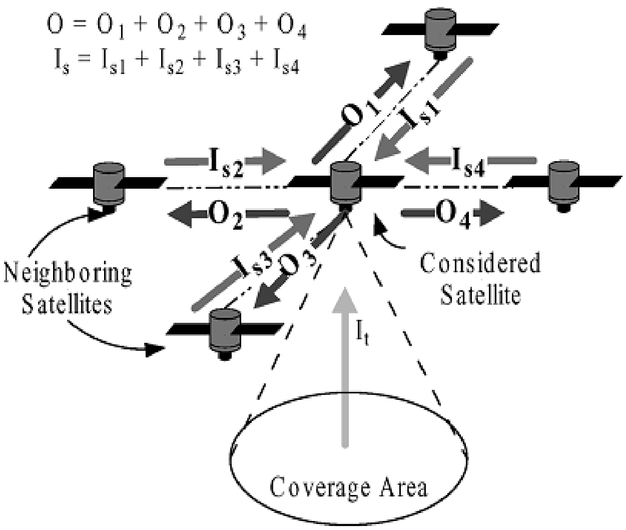


Fig.1. Rates of Traffic coming from neighboring satellites and terrestrial terminals.

So as that the satellite is ensured a prompt recovery and a time, the new residual time in the normal state for at least time, the new rate of traffic coming from neighboring satellites, should satisfy the following equation:

$$(I_s^{\mathrm{new}} + I_t) - O = \frac{P_{\mathrm{avg}} \cdot (q(t_{\mathrm{BSA}}) - Q_l \cdot \alpha)}{\theta}.$$

The TRR parameter can be accordingly computed as

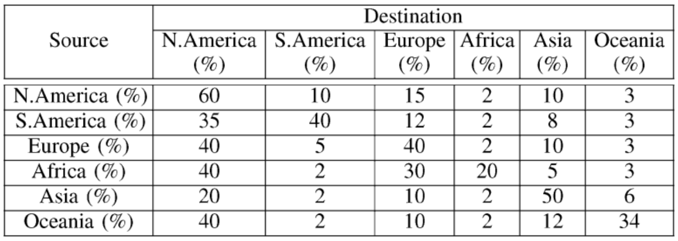
$$\chi = \mathrm{Min}\left(\mathrm{Max}\left(0, \frac{I_s^{\mathrm{new}}}{I_s}\right), 1\right).$$

4. PERFORMANCE VALUATION

A. Simulation Setup

In this section, we evaluate the performance of the EL scheme using the Network Simulator (NS) [35]. We consider an Iridium-like constellation. The constellation is formed of 66 satellites evenly and uniformly distributed over six orbits In the considered constellation, we do not consider the seams where two ISLs are switched off due to the motion in opposite directions. There by, it is assumed that at any time each Satellite maintains four ISLs with its neighboring satellites Uplinks ,downlinks, and ISLs are each given a capacity equal 25 Mbps(C=25 Mbps). In all conducted simulations all links are presumed to be error-free. The rationale beneath this assumption is to avoid any possible confusion between throughput degradation due to packet drops (due in turn buffer over flows at satellites) and that due to satellite channel errors. While such an assumption does not hold in real networks ,results of simulations conducted in environments wit channel errors demonstrated that link errors do not change any of the fundamental observations made about the proposes ELB scheme. The same thing applies to the performance ELB in environments with varying ISL delays .For this reason unless otherwise stated the delays of ISL links are all set to 20ms(d=20ms).With no specific purpose in mind ,the average packet size is set to 1KB. Drop-Tail based buffers of lengths equal to 200 packets are used. For traffic generation, we consider 600 non-persistent on-Of flows. The On/Off periods of the connections are derived from a Pareto distribution with a shape equal to 1.2. The average burst time and the average idle time are set to 200ms.

TABLEII DISTRIBUTION OF END-TERMINALS OVER THE SIX CONTINENTAL REGIONS



Source	Destination					
	N.America (%)	S.America (%)	Europe (%)	Africa (%)	Asia (%)	Oceania (%)
N.America (%)	60	10	15	2	10	3
S.America (%)	35	40	12	2	8	3
Europe (%)	40	5	40	2	10	3
Africa (%)	40	2	30	20	5	3
Asia (%)	20	2	10	2	50	6
Oceania (%)	40	2	10	2	12	34

The source and destination end-terminals are dispersed all over the Earth, divided into six continental regions, following a distribution identical to the traffic distribution used in[36],[37](TableII).The sources send data at constant rates from with in the range of 0.8 Mbps to 1.5Mbps.

B. Simulation Results

Single Traffic Class: First, we consider the case of a single traffic class. To investigate the abilities of the ELB scheme Supporting QoS, we evaluate its performance in terms of the achieved throughput and the total packet drops experienced by the simulated 600 connections. Fig.2 graphs the total number of packet drops experienced by all the connections during the entire simulation course and that is for different sending rates of the connections. For all the considered bit rates, the implementation of ELB over CEMR shows the best performance as it achieves the lowest packet drop rate. Note also that even the implementation of ELB over DSP avoids more packet drops than the other two routing protocols, DSP and CEMR. This indicates an important feature of the ELB scheme in avoiding Packet drops by alleviating congestion at satellites. The good performance of the ELB scheme in avoiding packet drops is also manifested in terms of the high throughput achieved by the ELB scheme. Fig.3 shows that implementing ELB over CEMR and DSP leads to be makeable increase in the total achieved throughput compared to the other two schemes, DSP and CEMR.

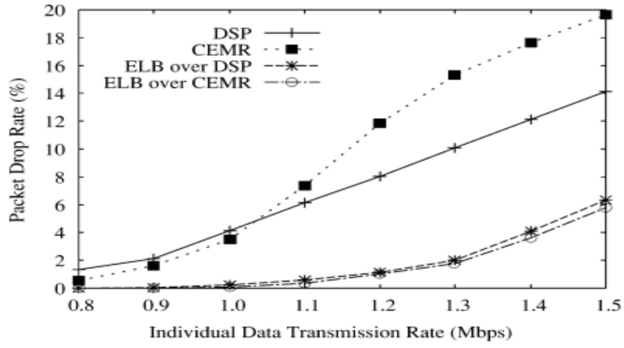


Fig.2. Packet drops for different individual sending rates

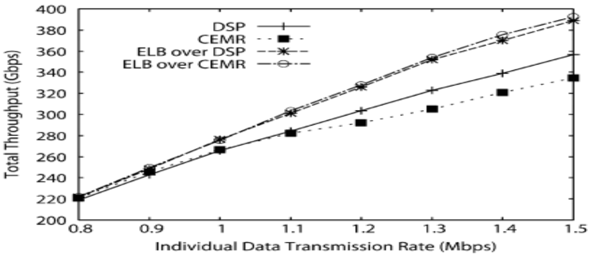


Fig.3. Total throughput for different individual sending rates.

V. CONCLUSION

In this paper, we proposed a cooperative routing strategy that enables neighboring satellites to explicitly exchange information their current congestion status. Satellites with queue occupancies exceeding a pre-determined threshold request their neighboring satellites to reduce their data forwarding rates. In response, the neighboring satellites transmit a predetermined portion of their data via less congested paths. The working of the proposed routing scheme is based on three metrics. A dynamic setting of these parameters is proposed based on easy-to-implement equations. The philosophy behind the parameters setting consists in reflecting the packet dropping probability in the parameters and guaranteeing a minimum level of stability for the satellites. To avoid the packet redistribution cascading issue, a routing cost metric, involving both the propagation delay and the queuing delay, is used. The targeted applications of the ELB scheme are preferably those that are delay insensitive and most importantly tolerant to a certain level of packet disorder or delay jitter. For this purpose, a class-based traffic detouring mechanism is added to the design of ELB. To cope with packet reordering issue in ELB and its impact on TCP, a TTL-based enhanced congestion control mechanism is also portrayed. The proposed ELB scheme is practical and can be accomplished without changing the routing protocol in use. A set of simulations is conducted to evaluate the performance of the ELB scheme. Two implementations are considered; one over a recently proposed scheme, CEMR, and the other over the most widely used Dijkstra algorithm. The obtained simulation results elucidate the better performance of the ELB scheme in avoiding congestion, reducing queue lengths, lowering packet drops, and increasing the total throughput while maintaining a more balanced distribution of traffic over the constellation. The performance of the scheme is also evaluated in terms of delays. Interestingly, encouraging results are obtained. Indeed, while individual flows suffer from a slight increase in their delays as their packets have to traverse additional hops, the aggregate performance of the ELB scheme, seen in terms of the cumulative distribution function of flow average delays, is fairly good. This result is attributable to the abilities of the ELB scheme in reducing queuing delays. Furthermore, considering the extra time that may be required for retransmitting dropped packets in case of connection-oriented transport protocols, this result would be seen more promising. Finally, it should be emphasized that the obtained results are critical for the implementation of Differentiated Services architectures over NGEO satellite constellations. The actual enhancements that the ELB scheme can indeed bring to such Diff-Serv-supporting NGEO satellite systems is an interesting area of research and forms the basis of our future research work.

References

[1] T.Taleb, N.Kato, and Y.Nemoto, “Recent trends in IP/NGEO satellite communication systems: Transport, routing, and mobility management”, IEEE Wireless Commun.Mag., vol.12, no.5, pp.63-69, Oct.2005.

[2] J.Farserotu and R.Prasad, “A survey of future broadband multimedia satellite systems, issues and trends”, IEEE Commun.Mag., vol.38, no.6, pp.128-133, 2000.

[3] J.Neale, R.Green, and J.Landovskis, “Interactive channel for multimedia satellite networks”, IEEE Commun.Mag., vol.39, no.3, pp.192-198, 2001.

[4] M.Wittig, “Satellite on board processing for multimedia applications “, IEEE Commun.Mag., vol.38, no.6, pp.134-140, Jun.2000.

[5] M.Wittig, ”Large capacity multimedia systems”, IEEE CommuMag., vol.35, no.8, pp.44-49, Aug.1997.

[6] T.Taleb, N.Kato, and Y.Nemoto, “REFWA: An efficient and fair congestion control scheme for Leo satellite networks”, IEEE/ACM Trans. Networking, vol.14, no.5, pp.1031-1044, Oct.2006.

[7] T.Taleb, D.Mashimo, A.Jamalipour, K.Hashimoto, Y.Nemoto, and N.Kato, “ELB: An explicit load balancing routing protocol for multi-hop NGEO satellite constellations”, in IEEE Globecom’06, San Francisco, CA, Nov.2006.

[8] L.Wood, A.Clerget, I.Andrikopoulos, G.Pavlou, and W.Dabbous, “IP routing issues in satellite constellation networks”, Int.J.Satellite Commun., vol.19, no.1, pp.69-92, Jan./Feb.2001.

[9] M.Werner, “A dynamic routing concept for ATM based satellite personal communication networks”, IEEE J. Sel .Areas Commun., vol.15 no.8, pp.1636-1648,Oct.1997.

[10] R.Mauger and C.Rosenberg,”QoS guarantees for multimedia service on a TDMA based satellite network”, IEEE Commun.Mag., vol.35, no7, pp.56-65,July1997.

Spectrum Sensing Algorithm for Cognitive Radio based on Covariance Matrix

Rohitha[1] U M, BITM, Bellary.

Dr. Siddarama R Patil[2], PDACE, Gulbarga

1. Email: rohitha_ujjini@rediffmail.com

2. Email: pdapatil@gmail.com

Abstract **-** In a cognitive radio network, the spectrum that is allocated to primary users can be used by secondary users if the spectrum is not being used by the primary user at the current time and location. The only consideration is that the secondary users have to vacate the channel within a certain amount of time whenever the primary user becomes active. Thus, the cognitive radio faces the difficult challenge of detecting (sensing) the presence of the primary user. Since the statistical covariances of received signal and noise are usually different, they can be used to differentiate the case where the primary user's signal is present from the case where there is only noise. In this paper, spectrum sensing algorithms are proposed based on the sample covariance matrix calculated from a limited number of received signal samples. Two test statistics are then extracted from the sample covariance matrix. A decision on the signal presence is made by comparing the two test statistics. Theoretical analysis for the proposed algorithms is given. Detection probability and associated threshold are found based on statistical theory. The methods do not need any information of the signal, the channel and noise power *a priori*. Also, no synchronization is needed.

1. Introduction

The recent rapid growth of wireless communications has made the problem of spectrum utilization ever more critical. On one hand, the increasing diversity and demand of high quality-of-service applications have resulted in overcrowding of the allocated spectrum bands, leading to significantly reduced levels of user satisfaction. There are many holes in the radio spectrum that could be exploited by the secondary users. The secondary user must exploit these spectrums opportunities without causing harmful degradation to the primary user.

Recent studies by the Federal Communications Commission (FCC) shows that the spectrum utilization in the 0 to 6 GHz band varies from 15% to 85%. This has prompted the FCC to propose the opening of licensed bands to unlicensed users and given birth to cognitive radio. Cognitive users need to monitor the spectrum activities continuously to find a suitable spectrum band for possible utilization and to avoid possible interference to the licensed users. Since the primary users have the priority of service, spectrum sensing by cognitive users includes detection of possible collision when a primary user becomes active in the spectrum momentarily occupied by a cognitive user and relocation of the communication channels.

Cognitive radio, a new and novel way of thinking about wireless communications, has the potential to become the solution to the spectrum underutilization problem. Building on spectrum sensing and other basic tasks, the ultimate objective of a cognitive radio network is twofold:

- Provide highly reliable communication for all users of the network, wherever and whenever needed;
- Facilitate efficient utilization of the radio spectrum in a fair-minded and cost-effective manner.

2. Spectrum Sensing

Spectrum sensing is a fundamental task for cognitive radio. However, there are several factors which make spectrum sensing practically challenging. First, the signal to noise ratio (SNR) of the primary users may be very low. For example, the wireless microphones operating in TV bands only transmit signals with a power of about 50mW and a bandwidth of 200 kHz. If the secondary users are several hundred meters away from the microphone devices, the received SNR may be well below −20dB. Secondly, multipath fading and time dispersion of the wireless channels make the sensing problem more difficult. Multipath fading may cause the signal power fluctuates as large as 20 − 30dB. On the other hand, coherent detection may not be possible when the time dispersed channel is unknown, especially when the primary users are legacy systems which do not cooperate with the secondary users. Thirdly, the noise/interference level may change with time, which yields noise uncertainty. There are two types of noise uncertainty: receiver device noise uncertainty and environment noise uncertainty. The receiver device noise uncertainty comes from non-linearity of components and time-varying thermal noise in the components. The environment noise uncertainty may be caused by transmissions of other users, either unintentionally or intentionally. Because of noise uncertainty, in practice, it is very difficult to obtain the accurate noise power.

3. Spectrum Sensing Methods

There have been several sensing methods, including the likelihood ratio test (LRT), energy detection method, matched filtering (MF)-based method and cyclostationary detection method, each of which has different requirements and advantages/disadvantages. Although LRT is proved to be optimal, it is very difficult to use it in practice, because it requires exact channel information, and distributions of source signal and noise. In order to use LRT for detection, we need to obtain the channels and signal and noise distributions first, which are practically intractable. MF-based method requires perfect knowledge of the channel responses from the primary user to the receiver and accurate synchronization (otherwise its performance will be reduced dramatically). As mentioned earlier, this may not be possible if the primary users do not cooperate with the secondary users. Cyclostationary detection method needs to know the cyclic frequencies of the primary users, which may not be realistic for many of the spectrum reuse applications. Furthermore, this method demands excessive analog to digital converter (ADC) requirement and signal processing capabilities. Energy detection, unlike the other two methods, does not need any information of the signal to be detected and is robust to unknown dispersed channel and fading. However, energy detection requires perfect knowledge of noise power. Wrong estimation of the noise power leads to SNR wall and high probability of false alarm. As pointed out earlier, the estimated noise power could be quite inaccurate due to noise uncertainty. Thus, the main drawback for the energy detection is its sensitiveness to noise uncertainty. Furthermore, while energy detection is optimal for detecting independent and identically distributed (iid) signal, it is not optimal for detecting correlated signal, which is the case for most practical applications.

In this paper, to overcome the shortcoming of energy detection, we propose new methods based on statistical covariances or auto-correlations of the received signal. The statistical covariance matrices or auto-correlations of signal and noise are generally different. Thus this difference is used in the proposed methods to differentiate the signal component from background noise. In practice, there are only limited numbers of signal samples. Hence, the detection methods are based on the sample covariance matrix. The steps of the proposed methods are as follows. First, the sample covariance matrix of the received signal is computed based on received signal samples. Then two test statistics are extracted from the sample covariance matrix. Finally, a decision on the presence of the signal is made by comparing the ratio of two test statistics with a threshold. Theoretical analysis for the proposed algorithms is given. Detection probability and associated threshold for decision are found based on statistical theory. The methods do not need any information of the signal, the channel and noise power *a priori*. Also, no synchronization is needed.

4. Covariance Based Detections

Let x(t) = s(t) + v(t) be the continuous-time received signal, where s(t) is the possible primary user's signal and v(t) is the noise. n(t) is assumed to be a stationary process satisfying E{n(t)} = 0, E{v^2(t)} = σ_v^2 and E{(v(t) v (t+p)} = 0 for any p ≠ 0. Assume that we are interested in the frequency band with central frequency fc and bandwidth W. We sample the received signal at a sampling rate fs, where fs ≥ W. Let Ts = 1/fs be the sampling period. For notation simplicity, we define

x(n) =, x(nTs), s(n) = s(nTs) and v(n) = v(nTs). There are two hypothesizes: H0, the signal does not exist; and H1, the signal exists. The received signal samples under the two hypothesize are given respectively as follows:

$$H_0: \; x(n) = v(n)$$

$$H_1: \; x(n) = s(n) + v(n)$$

where s(n) is the transmitted signal samples passed through a wireless channel consisting of path loss, multipath fading and time dispersion effects, and v (n) is the white noise which is independent and identically distributed (iid), and with mean zero and variance σ_v^2. Note that s(n) can be the superposition of the received signals from multiple primary users. Two probabilities are of interest for spectrum sensing: probability of detection, Pd, which defines, at the hypothesis H1, the probability of the sensing algorithm having detected the presence of the primary signal; and probability of false alarm, Pfa, which defines, at the hypothesis H0, the probability of the sensing algorithm claiming the presence of the primary signal.

5. Covariance Absolute Value (CAV) Detection

Let us consider L consecutive samples and define the following vectors:

$$x(n) = [\, x(n), x(n-1), \cdots . \; x(n-L+1) \,]^T$$

$$s(n) = [\, s(n), s(n-1), . \cdots \; s(n-L+1) \,]^T$$

$$v(n) = [\, v(n), v(n-1), \cdots \; v(n-L+1) \,]^T$$

The parameter L is called smoothing factor in the following. Considering the statistical covariance matrices of the signal and noise defined as

$$Rx = E\{\, x(n)\, x^T(n) \,\}$$

$$Rs = E\{ s(n)\, s^T(n) \,\}$$

we can verify that

$$Rx = Rs + \sigma_v^2 \, I_L$$

If the signal s(n) is not present, Rs = 0. Hence the off diagonal elements of Rx are all zeros. If there is signal and the signal samples are correlated, Rs is not a diagonal matrix. Hence, some of the off-diagonal elements of Rx should be non-zeros. Denote r_{nm} as the element of matrix Rx at the n^{th} row and m^{th} column, and let

$$T_1 = \frac{1}{L} \sum_{n=1}^{L} \sum_{m=1}^{L} |\, rnm|$$

$$T_2 = \frac{1}{L} \sum_{n=1}^{L} |\, rnn|$$

Then, if there is no signal, T1/T2 = 1. If the signal is present, T1/T2 > 1. Hence, the ratio T1/T2 can be used to detect the presence of the signal. In practice, the statistical covariance matrix can only be calculated using a limited number of signal samples. Define the sample auto-correlations of the received signal as

$$Y(\,k) = \frac{1}{Ns} \sum_{m=1}^{Ns-1} x(m)x(m-k)$$

where Ns is the number of available samples. The statistical covariance matrix Rx can be approximated by the sample covariance matrix defined as

$$\mathbf{R}^{\wedge}x(Ns) = \begin{bmatrix} Y(0) & Y(1) \cdots & Y(L-1) \\ \vdots & \vdots & \vdots \\ Y(L-1) & Y(L-2) & Y(0) \end{bmatrix}$$

The sample covariance matrix is symmetric and Toeplitz. Based on the sample covariance matrix, we propose the following signal detection method.

Algorithm: The covariance absolute value (CAV) detection algorithm

Step 1. Sample the received signal

Step 2. Choose a smoothing factor L and a threshold α_1, where α_1 should be chosen to meet the requirement for the probability of false alarm.

Step 3. Compute the auto-correlations of the received signal Y(k), k = 0, 1, · · · ,L − 1, and form the sample covariance matrix.

Step 4. Compute

$$T_1(Ns) = \frac{1}{L} \sum_{n=1}^{L} \sum_{m=1}^{L} | \mathrm{rnm}\,(Ns)|$$

$$T_2\,(Ns) = \frac{1}{L} \sum_{n=1}^{L} | \mathrm{rnn}(Ns)|$$

where $\mathrm{rnm}\,(Ns)$ are the elements of the sample covariance matrix $\mathbf{R}^{\wedge}x\ (N_s)$

Step 5. Determine the presence of the signal based on $T_1(Ns)$, $T_2(Ns)$ and the threshold α_1, i.e., if $T_1(Ns)/T_2(Ns) > 1$, signal exists; otherwise, signal does not exist. The statistics in the algorithm can be calculated directly from the auto-correlations Y(k).

6. Theoretical Analysis for CAV Algorithm

The proposed method only uses the received signal samples. It does not need any information of the signal, the channel and noise power as a priori. Also, no synchronization is needed. The validity of the proposed CAV algorithm relies on the assumption that the signal samples are correlated, that is, Rs is not a diagonal matrix (some of the off-diagonal elements of Rs should be non-zeros). Obviously, if the signal samples s(n) are iid, then Rs = $\sigma_s^2\ I_L$. In this case, the assumption is invalid and the algorithm cannot detect the presence of the signal. However, usually the signal samples should be correlated due to the following reasons.

The signal is oversampled. Let T_0 be the Nyquist sampling period of the signal s(t) and $s(nT_0)$ be the sampled signal based on the Nyquist sampling rate. Based on the sampling theorem, the signal s (t) can be expressed as

$$S(t) = \sum_{n=-\infty}^{\infty} s(nT0)g\ (t - nT0\ )$$

Where g(t) is an interpolation function. Hence, the signal samples s(n) = s(nTs) are only related to $s(nT_0)$. If the sampling rate at the receiver fs > 1/ T_0, that is, Ts < T_0, then s(n) = s(nTs) must be correlated. An example of this is the narrowband signal such as the wireless microphone signal. In a 6 MHz bandwidth TV band, a wireless microphone signal only occupies about 200 KHz. When we sample the received signal with sampling rate not lower than 6 MHz, the wireless microphone signal is actually over-sampled and therefore highly correlated.

The original signal is correlated. In this case, even if the channel is a flat fading channel and no oversampling, the received signal samples are correlated. Another assumption for the algorithm is that the noise samples are iid. This is usually true if no filtering is used. However, if a narrowband filter is used at the receiver, sometimes the noise samples will be correlated. To deal with this case, we need to pre-whiten the noise samples or pre-transform the covariance matrix.

The computational complexity of the algorithm is as follows. Computing the auto-correlations of the received signal requires about LNs multiplications and additions. Computing $T_1(Ns)$ and $T_2(Ns)$ requires about L^2 additions.

Therefore, the total number of multiplications and additions is about LNs + L^2.

7. Simulation Results:

First, we simulate the probabilities of false alarm (Pfa) because Pfa is not related to signal (at H0, there is no signal at all). We set the target Pfa = 0.1, and choose L = 10 and Ns = 50000. We then obtain the thresholds based on the Pfa, L and Ns. Table 1 gives the simulation results for various cases, where and in the following “EG-x dB” means the energy detection with x-dB noise uncertainty. The Pfa for the proposed method and energy detection without noise uncertainty meet the target, but the Pfa for the energy detection with noise uncertainty (even as low as 0.5 dB) far exceeds the limit. This means that the energy detection is very unreliable in practical situations with noise uncertainty.

method	EG-2dB	EG-1.5dB	EG-1dB	EG-0.5dB	EG-0dB	CAV
Pfa	0.497	0.496	0.490	0.470	0.102	0.099

Table 1: Probabilities of false alaram

Figure 1 gives the simulation results (the corresponding Pfa is shown in Table 1). Due to some approximations, the theoretical result does not exactly match the simulated results. The CAV is better than ideal energy detection (without noise uncertainty).

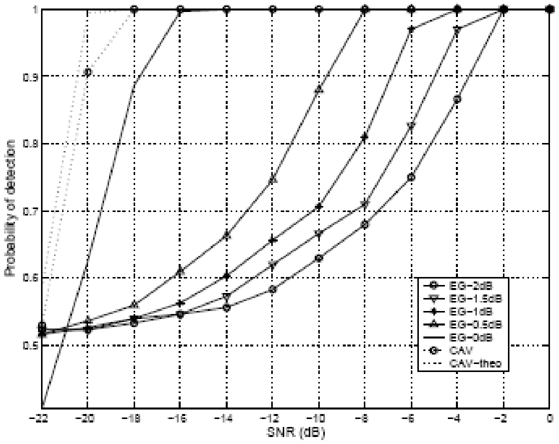


Figure 1: Probability of detection for wireless microphone signal: Ns = 50000

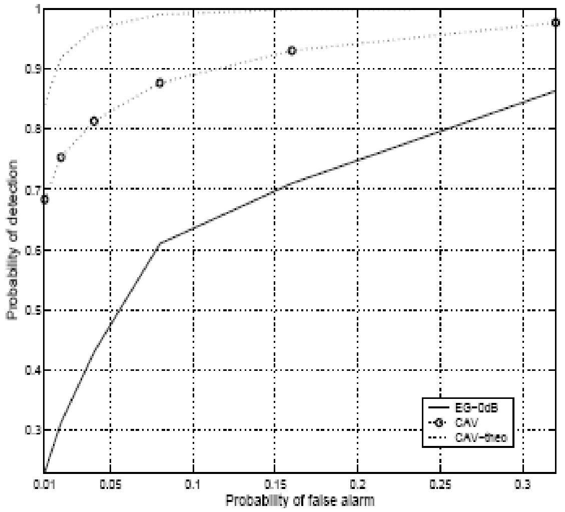


Figure 2: P_d versus P_{fa} for wireless Microphone Signal: N_s=50000, SNR= -20db

As shown in the figure 2, if there is noise uncertainty, the Pd of the energy detection is much worse than that of the proposed method. Figure 2 gives the Receiver Operating Characteristics (ROC) curve (Pd versus Pfa) at fixed SNR = −20dB. It is clear that CAV is always better than the ideal energy detection.

Conclusions

In this paper, sensing algorithms based on the sample covariance matrix of the received signal have been proposed. Statistical theories have been used to set the thresholds and obtain the probabilities of detection. The methods can be used for various signal detection applications without knowledge of the signal, the channel and noise power. Simulations based on the narrowband signals, captured DTV signals and multiple antenna signals have been carried out to evaluate the performance of the proposed methods. It is shown that the proposed methods are in general better than the energy detector when noise uncertainty is present. Furthermore, when the received signals are highly correlated, the proposed method is better than the energy detector even the noise power is perfectly known

References:

[1] J. Mitola and G. Q.Maguire, “Cognitive radios: making software radios more personal,” *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[2] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Trans. Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] D. Cabric, S. M. Mishra, D.Willkomm, R. Brodersen, and A. Wolisz, “A cognitive radio approach for usage of virtual unlicensed spectrum,” in *14th IST Mobile and Wireless Communications Summit*, June 2005.

[4] S. M. Mishra, A. Sahai, and R. W. Brodensen, “Cooperative sensing among cognitive radios,” in *IEEE International Conference on Communications (ICC)*, (Istanbul, Turkey), June 2006.

[5] Y.-C. Liang, Y. H. Zeng, E. Peh, and A. T. Hoang, “Sensing-throughput tradeoff for cognitive radio networks,” *IEEE Trans on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, 2008.

[6] A. Sahai and D. Cabric, “Spectrum sensing: fundamental limits and practical challenges,” in *Proc.IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, (Baltimore, MD), Nov. 2005.

[7] R. Tandra and A. Sahai, “Fundamental limits on detection in low SNR under noise uncertainty,” in*WirelessCom 2005*, (Maui, HI), June 2005.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2. Prentice Hall, 1998.

[9] H. Urkowitz, “Energy detection of unkown deterministic signals,” *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[10] H.-S. Chen, W. Gao, and D. G. Daut, “Signature based spectrum sensing algorithms for IEEE 802.22 WRAN,” in *IEEE Intern. Conf. Comm. (ICC)*, June 2007.

[11] C. Clanton, M. Kenkel, and Y. Tang, “Wireless microphone signal simulation method,” in *IEEE 802.22- 07/0124r0*, March 2007.

[12] Monson H. Hayes, “Statistical Digital Signal Processing and Modeling,”

.

NEUROCOMPUTING BASED SECURITY SYSTEM

K.S.Shivakumar & D.Aradhana
Dept. of Computer Science Engineering
Bellary Engineering College, Bellary.

Abstract

Security is a broad topic and covers many issues. Malicious people trying to gain some benefit, attention, or to harm someone intentionally cause most security problems. As, Complete security not possible in real life, Transition will be long in coming, Cryptographic methods and have their own problems, The stricter the mechanisms the lower the efficiency are the limitations related to security. Intrusion Detection System (IDS) is an emerging new technology, being informed is the best weapon in the security analyst's arsenal. "An ounce of prevention is worth a pound of detection". An Intrusion Detection System detects attacks as soon as possible and takes appropriate action. Security is a compulsory need for data operation today. The authentication process or commerce exchanges need security and reliability. The area in which this algorithm developed using ANN can be used anywhere where Security is a must. This paper along with a test result shows that the possibility of guessing keys is extremely weaker than using any other standard method. The presented results are obtained through MATLAB 7.0 software.

KEY WORDS: Security, BackPropagation ANN, Neural Networks Toolbox.

1. INTRODUCTION: The problem of protecting information has existed since information has been managed. However, as technology advances and information management systems become more and more powerful, the problem of enforcing information security also becomes more critical. The massive use of the communication networks for various purposes in the past few years has posed new serious security threats and increased the potential damage that violations may cause. As organizations are increasing their reliance on computer network environments, they are becoming more vulnerable to security breaches. Private and public sectors more than ever today depend on the information they manage. A violation to the security of the information may jeopardize the whole system working and cause serious damages. Advances in artificial neural networks (ANNs) provide effective solutions to this problem.

The security problem is considered here as the problem of keeping communications over the network private. In other words, a secure network allows only the intended recipient to intercept and read a message addressed to her/him. Thus, protection of information is required against possible violations that can compromise its secrecy (or confidentiality). Secrecy is compromised if information is disclosed to users not authorized to access it. While the encryption scheme used in this work is based on Boolean algebra, the decryption scheme here is based on a neural network technique that uses Feedforward and also Backpropagation learning algorithm.

1.1 Artificial Neural Networks (ANNs): A neural network is a massively parallel-distributed processor made up from simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. The use of neural network offers the Input-Output Mapping property and capability The ANNs learning algorithms can be divided into two main groups that are supervised (or Associative learning) and unsupervised (Self-Organization) learning. Supervised learning learns based on the target value or the desired outputs. During training the network tries to match the outputs with the desired target values. It is presented with an example picked at random from the set and the synaptic weights of the network are modified to minimize the difference between the desired response and the actual response of the network produced by the input signal in accordance with an appropriate statistical criterion. The training of the network is repeated for many examples in the set until the network reaches a steady state, where there are no further significant changes in the synaptic weights. The previously applied training example may be reapplied during the training session but in a difference order. Thus the network learns from the examples by constructing an input-output mapping for the problem at hand

Unsupervised learning method is not given any target value. A desired output of the network is unknown. During training the network performs some kind of data compression such as dimensionality reduction or clustering. The network learns the distribution of patterns and makes a classification of that pattern where, similar patterns are assigned to the same output cluster.

1.2 The backpropagation neural network: One of the most commonly used supervised ANN model is backpropagation network that uses backpropagation learning algorithm[2, 12, 13]. Backpropagation (or backprop) algorithm is one of the well-known algorithms in neural networks. The introduction of backprop algorithm has overcome the drawback of previous NN algorithm in 1970s where single layer perceptron fail to solve a simple XOR problem. The backpropagation neural network is essentially a network of simple processing elements working together to produce a complex output. These elements or nodes are arranged into different layers: input, middle and output. The output from a backpropagation neural network is computed using a procedure known as the forward pass.

* The input layer propagates a particular input vector's components to each node in the middle layer.

* Middle layer nodes compute output values, which become inputs to the nodes of the output layer.

* The output layer nodes compute the network output for the particular input vector.

The forward pass produces an output vector for a given input vector based on the current state of the network weights. Since the network weights are initialized to random values, it is unlikely that reasonable outputs will result before training. The weights are adjusted to reduce the error by propagating the output error backward through the network. This process is where the backpropagation neural network gets its name and is known as the backward pass:

* Compute error values for each node in the output layer. This can be computed because the desired output for each node is known.

* Compute the error for the middle layer nodes. This is done by attributing a portion of the error at each output layer node to the middle layer node, which feed that output node. The amount of error due to each middle layer node depends on the size of the weight assigned to the connection between the two nodes.

* Adjust the weight values to improve network performance using the Delta rule.

* Compute the overall error to test network performance. The training set is repeatedly presented to the network and the weight values are adjusted until the overall error is below a predetermined tolerance.

2. IMPLEMENTATION AND DESIGN

- To implement the function of ANN in this Security System like a **one-way HASH function**.

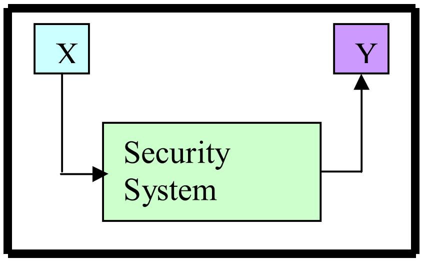


Fig. 1

X_i → User Identity
Y_i → Output of processing environment

- Implement unique value corresponding to X_i
- The Reverse Transformation should not be possible (i.e., X_i cannot be recovered from Y_i)
- To design the size of the Ann Architecture which will depend on the length of the user identity.
- The design of Feed forward Architecture and Back Propagation Algorithm (steepest descent) learning rule is used.
- Multilayer Architecture of Security such as Personally Identification Protection, System Identification Protection
- Multivariable are being used to provide security such as time of intrusion, identification inserted by intruder, time taken in inserting identity, number of trails taken by intruder before unauthentication declared.
- Resetting of all identity are allowed
- The main key of this work is Analysis/ Detection of intrusion enhancing the security service.

2.1 Multi-Parameter Description in Security System

The features of the Security System implemented in the project for Resource Protection is discussed as follows:

Multi-Variable Parameters and their Hierarchy in Security System

The multivariable parameters introduced in the project are

- Time of intrusion (Year, Month, Date, Hour, Minutes, Seconds)
- Identification inserted by intruder (Length check)
- Time taken in inserting the Identification
- No. of trails taken by intruder before unauthentication declared.

2.2 Multi-Layer Parameters and their Hierarchy in Security System

The multi-layer parameters introduced in the project are

- P.I.P → Personal Identification Protection
- S.I.P → System Identification Protection
- R → Resource being Protected

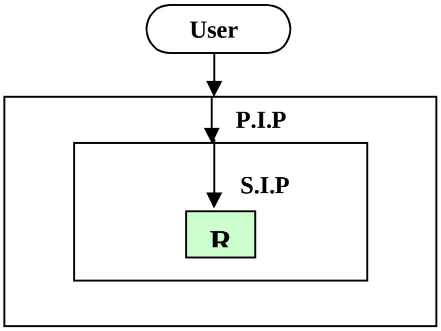


Fig. 2

Each P.I.P and S.I.P will be supported by protecting service provide by multivariable parameters.

To implement our Neural Network we used the Neural Network Toolbox in MATLAB 7.0

At the beginning of the learning process, the weight matrices between input and hidden layer and between hidden and output layer are initialized with the random values in the interval. Vectors for hidden neuron biases and output neuron biases are also initialized with random values. In the hidden and output layers, the linear activation functions have been used. After several iterations, when the difference between the calculated output and the desired output is less than the threshold value, the iteration is stopped.

3. RESULTS: The system has been implemented, trained and experimented using different data sets in the MATLAB environment. The system has been trained in several different ways and tested with different sets of test data, as illustrated in figures

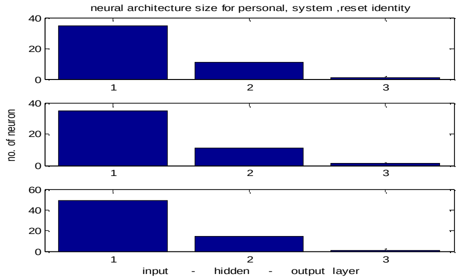


Fig. 3

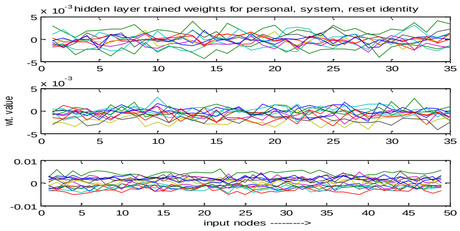


Fig. 4

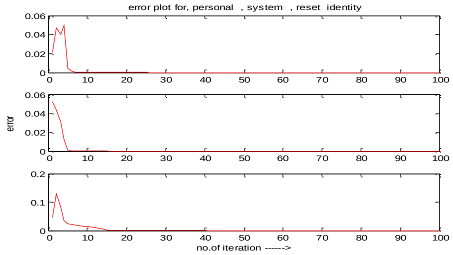


Fig. 5

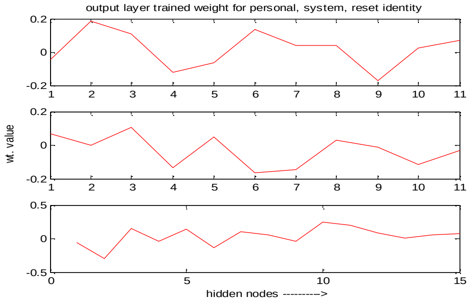


Fig. 6

4. CONCLUSION: Artificial neural networks are inspired by the learning processes that take place in biological systems. Neural networks represent a new computing paradigm based on the parallel architecture of the brain. They can be "trained " to produce an accurate output for a given input. Network posses the advantages of simple computations, fault tolerance, parallel processing, robust with respect to node failure. In this project, the concept of Error Back-Propagation Learning algorithm has made a break through in supervised learning of layered neural network. This project proves that the Protection of Resource is not only is secure but it has been secured. In this project, security and intrusion detection developed using multilevel, multivariable parameters, the advantage of building the architecture to the user desire level, and hiding learning phase from the intruder significantly increases the performance of the network. The only limitation focused in this project is training is slow, may converge to a local, not global, minimum of error.

Finally, I conclude that project 'Neuro Computing Based Security Systems' overcomes all the limitations stated in the literature survey, and has been proven successfully the ability of the network to protect information and system resources with respect to confidentiality and integrity.

REFERENCES:

[1] Khalil Shihab "A Backpropagation Neural Network for Computer Network Security", Journal of Computer Science 2 (9): 710-715, 2006 ISSN 1549-3636 © 2006 Science Publications

[2] JIAN LI, GUO-YIN ZHANG, GUO-CHANG GU "The Research And Implementation Of Intelligent Intrusion Etection System Based On Artificial Neural Network", Proceedings of the Third International Conference on Machine Laming and Cybernetics, Shanghai, 26-29 August 2004

[3] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan College Press, New York, 1994.

AN ARM EMBEDDED SOLUTION FOR LIQUID DIELECTRIC CONSTANT MEASURMENT SYSTEM

AUTHORS: Prabhakar.K[1], Prof. Sadyojatha. K.M[2], Dr.Nagabhushan Katte[3]

[1] prabhakark1978@gmail.com [2] saddukm@gmail.com [3] nagkatte@gmail.com

[1,2,3]Ballari Institute of Technology and Management, Bellary

ABSTRACT

The project describes an ARM based solution for dielectric constant measurement System, which adapts ARM7 processor as a single chip embedded solution for the dielectric constant measurement. An ARM based solution for dielectric constant measurement system is based on the principle of measurement of frequency, which is the function of RC component. The change in liquid dielectric medium of the dielectric cell, acts as a capacitor in the RC time constant causes to change the frequency output of an XR-2206 function generator. Hence the change in frequency causes change in dielectric constant. The output of XR-2206 is given to F/V converter and is further acquired by ARM through its on-chip A/D converter. Temperature of the liquid, whose dielectric constant is to be measured, is also monitored simultaneously, as dielectric constant is temperature dependent parameter. A 16x2 line LCD module is interfaced to the ARM to have display of the dielectric constant of a liquid as well as temperature. The necessary software has been developed in embedded 'C' in the Kiel IDE environment. The instrument covers wide range of dielectric constant measurements for liquids at various concentrations and temperatures.

1. INTRODUCTION

Dielectric constant:

The dielectric constant is the ratio of the permittivity of the substance to the permittivity of the free space. It is an expression of the extent to which the material concentrates electric flux.

As the dielectric constant increases, the electric flux increases, if all other factors remain unchanged. This enables objects of a given size, such as sets of metal plates to hold their electric charge for a long period of time. Materials with high dielectric constants are useful in the manufacture of high value capacitors.

A high dielectric constant of a substance is not necessarily desirable, because materials with high dielectric constant breakdown more easily when subjected to intense electric fields than that of materials with low dielectric constant.

For example, dry air has a low dielectric constant, but it makes an excellent dielectric material for capacitors used in high-power radio frequency (RF) transmitters.

Permittivity:

Permittivity, also called electric permittivity, is a constant of proportionality that exists between displacement and electric field intensity. This constant is approximately equal to $8.854*10^{-12}$farad/meter in free space (vacuum). In other materials, it can be much different, often greater than the free space value, which is symbolized e_o.

In many applications, permittivity is often expressed in relative, rather than in absolute terms. If e_0 represents the permittivity of free space and e represents the permittivity of the substance, then the relative permittivity, also called dielectric constant e_r is given by

$$e_{r} = e/e_0 \qquad 1.1$$

Various substances have dielectric constants e_r greater than 1. These substances are generally called dielectric materials or simply dielectrics. Commonly used dielectrics include glass, paper, mica, various ceramics, polythene and certain metal oxides.

Dielectrics are used in capacitors and transmission lines in alternating current (AC), audio frequency (AF), and radio frequency (RF) applications.

Liquid dielectric constant measurement

The below method explains Dual micro strip resonator sensor for measurement of dielectric constant with high resolution and compact area of material under test.

Microwave resonator technology has wide applications for dielectric measurement; a non-destructive resonant method has been employed to determine the permittivity of the overlay plastic and liquid. The resonant circuit including a cavity, a coaxial, a strip line, and a micro strip line are example structures. To obtain a compact size for liquid material measurement, a planar transmission line such as micro strip line and a slot line is comfortable to use.

The dielectric material can be determined from an each mode of resonant frequency and relates to an effective dielectric constant as following equation

$$f_{r=nc/(leff\backslash\sqrt{}eff)} \qquad 1.2$$

Where n=resonant number of modes
C = speed of light
fr = resonant frequency
$l_{eff=}$ effective length of resonator

The conventional micro strip resonators are one wavelength ring and half wavelength straight line resonators.

The above equation is a general formula determining an effective dielectric constant which can be calculated at each mode number of single resonator.

Solids dielectric constant measurement system:

The reliability of a power transformer is largely determined by its insulation condition. However, transformer insulation deteriorates due to different stresses over a life time. Insulation degradation reduces the dielectric capability of a transformer to withstand electric stresses and increase the probability of failure. It is therefore important to develop a variety of dielectric diagnostic methods and give an early indication of the changes of the dielectric properties of insulating materials used in power transformer.

Among these dielectric diagnostic methods the most common industry practice-dissipation factor (tanδ) and capacitance measurements at power frequency. Apart from its merit of short measuring time, variation of tanδ provides useful information about the insulation quality. The main downside of this technique is that the measurement results obtained on a transformer represent the insulation losses of the combined transformer insulation systems-involving oil, cellulose paper and press board as insulating materials and their size and dimensions. The contribution of each insulation material to the end result is not well known which makes it difficult to use tanδ and C in estimating the degradation status of the whole insulation structure and each material

Gases dielectric constant measurement system:

The below explains one of the method for measurement of complex dielectric constant of gases at microwave frequencies using resonant cavity.

The gas whose complex dielectric properties are to be determined is contained in a resonant cavity which is a part of the microwave circuit.

With the introduction of the gas, the real part of the dielectric constant changes the resonant frequency of the cavity, while the imaginary part changes the amplitude and breadth of the cavity response curve.

By a rapid variation of the frequency across the cavity resonance, the real and imaginary parts can be conveniently and accurately determined from measurements on an oscilloscope.

The above method can be used for the measurement of the resonant dispersion and absorption of microwaves by gas molecules.

2. Scheduling of ARM embedded solution for dielectric constant measurement system

In this section we formulate the problem and state the constraints on the problem. We also define the objective function. Next, along with the flow chart, we present calibration, measurement procedure, results and conclusion.

3. Problem Definition

Several experimental techniques were developed for the measurement of capacitance and in turn the dielectric constant over a wide frequency range. In general, a bridge method or resonance technique is employed to determine the capacitance with or without the sample. Among them heterodyne beat method is popular at radio frequencies. All these methods suffer from some of the following drawbacks.

1. Manual control (human intervention) for adjusting a bridge or in attaining resonance is essential. Hence, some amount of time is required to reach the steady state and to sharply determine the capacitance. Scope to human error is not eliminated. Reproducibility of the results with higher precision is rather limited.
2. The apparatus usually consists of complex circuits and it is expensive.

Hence, it is proposed to design a technique using a slightly different principle with IC version function generator.

No doubt, several investigators developed the conventional dielectric constant measurement systems, both analog and digital. However, the attempts to design and develop the ARM based systems for the measurements of dielectric constant in liquids are rather scarce particularly in India though they offer many advantages. Hence, in present study, a humble attempt is made to design and construct ARM embedded system for the measurement of dielectric constant in liquids.

4. Principle of operation

The dielectric constant ε of a liquid is defined as the ratio of the electrical capacitance of a cell when the liquid /solution forms the dielectric medium c_s to the capacitance of the cell when air forms the dielectric medium c_o at the given temperature which is given by the following equation.

$$\varepsilon = c_s / c_o \qquad 4.1$$

The cell consists of two circular discs (25 mm diameter) of brass metal separated by a distance which acts as a capacitor. When it is immersed in a liquid, it acts as a dielectric and its capacitance changes. The cell has to first standardize to measure the dielectric constant of unknown solution. This is accomplished by considering a pure liquid such as benzene as reference liquid. The dielectric constant of unknown liquid ε_x can be determined by measuring the capacitance of the cell in air $c_{o,}$ the capacitance of cell in reference liquid, such as benzene c_r and the capacitance of the cell in liquid whose dielectric constant has to be measured c_x using relation

$$\varepsilon_x = 1 + [(c_o - c_x) / (c_o - c_r)] \times (\varepsilon_r - 1) \qquad 4.2$$

Where ε_r is the dielectric constant of the reference liquid (benzene in the present study). Dielectric Constant Measurement System is as shown in the figure 3.1.The IC XR- 2206 is a function generator chip. It acts as an RC oscillator. The frequency of oscillator depends on the values of timing resistor R and timing capacitor C. when the value of R is kept constant, the dielectric cell acts as a capacitor C which varies the dielectric medium. Consequently the frequency of the oscillator also changes. The measurement of frequency of the oscillator enables one to measure the value of the capacitance of the cell and thus the dielectric constant of the medium. In the present study, with suitable interface of the oscillator circuit, the frequency of the oscillator is measured and displayed on the LCD.

5. System working

The figure below shows System working of ARM Embedded Solution for liquid Dielectric Constant Measurement System.

The liquid whose dielectric constant to be measured is taken in a dielectric cell and intern connected to XR2206 function generator. It acts as an RC oscillator. The frequency of the oscillator depends on the values of timing resistor R and timing capacitor C. When the value of R is kept constant, the dielectric cell acts as a capacitor which varies the dielectric medium. Consequently the frequency of the oscillator also changes.

The measurement of frequency of the oscillator enables one to measure the value of the capacitance of the cell and thus the dielectric constant of the medium.

The generated frequency related to the dielectric constant is given to frequency to voltage converter (LM 2907) where the frequency is converted to the corresponding voltage. This output voltage of LM2907 is buffered and applied to one of the on chip of ARM processor .Further ARM processor processes the dielectric constant of the liquid to be measured by substituting the acquired voltage in a curve fitted equation and measured dielectric constant at particular temperature is displayed on the LCD module.

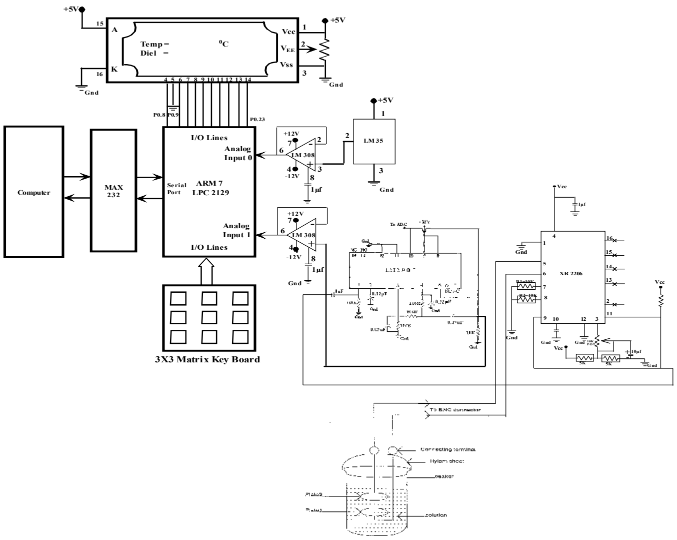
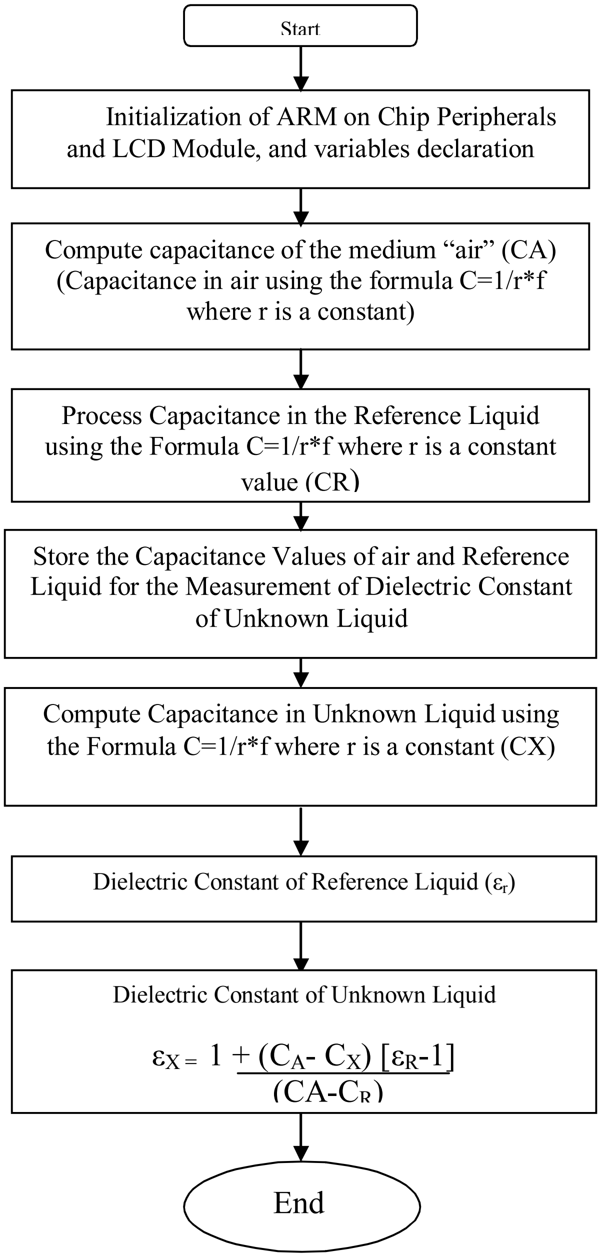


Fig1 Circuit Diagram of the Dielectric constant measurement system

6. Flow chart for development of software

The flow of the program has been depicted in fig.4.1. The flowchart is divided into three major parts; they are measurement of capacitance of air, measurement of capacitance of known liquid, and measurement of capacitance of unknown liquid. Numerical values of all above said capacitive measurements are substituted in an equation to compute the dielectric constant of an unknown liquid where dielectric cell has been placed.



Start
Initialization of ARM on Chip Peripherals and LCD Module, and variables declaration
Compute capacitance of the medium “air” (CA) (Capacitance in air using the formula C=1/r*f where r is a constant)
Process Capacitance in the Reference Liquid using the Formula C=1/r*f where r is a constant value (CR)
Store the Capacitance Values of air and Reference Liquid for the Measurement of Dielectric Constant of Unknown Liquid
Compute Capacitance in Unknown Liquid using the Formula C=1/r*f where r is a constant (CX)
Dielectric Constant of Reference Liquid (ε_r)
Dielectric Constant of Unknown Liquid
$\varepsilon_X = 1 + \frac{(C_A - C_X)\,[\varepsilon_R - 1]}{(CA - C_R)}$
End

7. Calibration and measurement procedure

Before using the system for measurement, it should be calibrated and standardized as per the following procedure

a) Clean the dielectric cell, dry it and keep it in the beaker containing air

b) Connect the cell to the circuit as shown in the fig

c) Switch on the system and activate the software.

d) The system measures and displays the frequency and in turn the capacitance of the cell. Note down the values.

e) Keep the reference liquid (benzene in the present study) in the cell and keep the cell in the temperature (30^{o}c in the present study).

f) Repeat steps a) to d).

g) Keep other liquid whose dielectric constant is to be measured in the cell and keep it in a temperature bath maintained at 30^{o}c

h) Repeat steps b) to d).

On execution of the program, the following will be displayed on the LCD

- DIELECTRIC CONSTANT
- TEMPERATURE

The calibration menu itself guides the user to calibrate and measure the dielectric constant which appears on the LCD

After making the appropriate adjustments both in the hardware and software and in calibration and measurement as per the procedure mentioned earlier, the instrument is tested for some liquids. The liquids are chosen such that they cover a wide range of dielectric constants. In the present investigation analytically reagent grade samples are used after necessary purification and distillation. Dielectric constants of pure liquids are measured using the equation.

$$\varepsilon_x = 1 + [(c_o - c_x)/(c_o - c_r)] \times (\varepsilon_r - 1)$$

Taking the value of dielectric constant of benzene (ε=2.264) as reference liquid. In the present work all the measurements are made at 30±0.01°c at 1MHz. the dielectric constants of pure liquids measured with this technique given in the table below along with the standard values from literature for comparison. The agreement between experimental and literature values is quite satisfactory with in experimental error. The measurements of the present study are accurate to ±0.1%.

8. Experimental Results

Liquid	Dielectric constant ε in Debye
Cyclohexane	17.96 present study ± 0.02 16.1 rubber hand book at 25 ºC 18.2 from website
Chlorobenzene	5.96 present study ± 0.01 5.91 weissberger at 30 ºC 5.8 rubber hand book at 25 ºC 5.9 from website.
Acetone	20.20 present study ± 0.02 21.01 rubber hand book at 25 ºC 20.34 International critical tables 20.35 From website.

9. Conclusion

The present project comes out with a systematic approach, design, development and fabrication of a low cost, compact and efficient system for liquid dielectric constant measurements.

The experimental results have been verified with the standard liquid dielectric constant measurement system which is available in the laboratory.

10 Future scopes

1. Intended to extend the work for measurement of Dielectric constant of solids and gasses.
2. Planned to improve the resolution of the system
3. Planning to provide SoC solution for dielectric constant measurement System

References

1. Curtis, A.J., in “Progress in Dielectrics”, (ed. J.B. Birks and J. Hart), Vol. 2, P.29, Heywood, London (1960).

2. Cole, R.H., in “Progress in Dielectrics”, (ed. J.B. Birks and J. Hart), Vol. 3, P.47, Heywood, London (1961).

3. McCrum.N.G., Read, B.E. and Williams, G.,”Anelastic and Dielectric effects in polymeric solids”, Wiley, London (1967

4. Roberts, S., Report No. 66-C333. General Electrical Research and Development Center, Schenectady, New York (1966).

5. Smyth, C.P., “dielectric Behavior and Structure”, McGraw-Hill, New York (1955).

6. Davidson, D.W., Auty, R.P. and Cole, R.H., Rev. Sci. Instrument, 22. 678 (1951).

Secure Routing In Mobile AdHoc Networks Using *SAODV* Protocol

Mallikarjuna A
Lecturer
BITM, Bellary mallismarty@rediffmail.com

Prof V C Patil
HOD ,Dept ECE
BITM, Bellary patilvc@rediffmail.com

Abstract

A mobile ad hoc network is a collection of nodes that is connected through a wireless medium forming rapidly changing topologies. The widely accepted existing routing protocols designed to accommodate the needs of such self-organized networks do not address possible threats aiming at the disruption of the protocol itself. In an Adhoc network nodes are always mobile, so routing becomes a challenge. Several protocols exist to address the dynamic nature of Adhoc networks. Adhoc On Demand Distance Vector (AODV) protocol is significantly accepted protocol which calculates the route dynamically in an efficient manner. Since routing is done in Adhoc network between source and destination with several intermediate nodes, security is a major challenge. So SAODV is one such protocol, which is used to provide security in Adhoc networks. The assumption of a trusted environment is not one that can be realistically expected; hence several efforts have been made towards the design of a secure and robust routing protocol for adhoc networks. So this paper addresses the major security problem in Adhoc network using SAODV.

I. Introduction.

During the last few years steady growth in the deployment of wireless and mobile communication networks have been witnessed. Mobile adhoc networks consist of nodes that are able to communicate through the use of wireless mediums and form dynamic topologies. Unfortunately all of the widely used adhoc routing protocols have no security considerations and trust all the participants to correctly forward routing and data traffic. This paper addresses the major security problem in Adhoc network using SAODV. Objectives of the work are to establish secure connection between two nodes using SAODV protocol and establishment of Connection through exchange of secure keys. This paper is organized into five sections.Section I gives introduction including objectives, Section II describes MANET, Section III briefs Adhoc network characteristics, and Section IV discusses various solutions Section V Describes the work done on SAODV protocol and concludes.

II.MANET

A MANET is considered a collection of wireless mobile nodes that are capable of communicating with each other without the use of a network infrastructure or any centralized administration. The mobile hosts are not bound to any centralized control like base stations or mobile switching centers.

Although this offers unrestricted mobility and connectivity to the users, the onus of network management is now entirely on the nodes that forms the network. Due to the limited transmission range of wireless network interfaces, multiple hops may be needed for one node to exchange data with another across the network. In such a network, each mobile node operates not only as a host but also as a router, forwarding packets for other mobile nodes in the network that may not be within direct wireless transmission range of each other.
Each node participates in an ad hoc routing protocol that allows it to discover multihop paths through the network to any other node.

The idea of MANET is also called *infrastructureless networking*, since the mobile nodes in the network dynamically establish routing among themselves to form their own network on the fly. It is formed instantaneously, and uses multihop routing to transmit information. MANET technology can provide an extremely flexible method of establishing communications in situations where geographical or terrestrial constraints demand a totally distributed network system without any fixed base station, such as battlefields, military applications, and other emergency and disaster situations. A sensor network, which consists of several thousand small low-powered nodes with sensing capabilities, is one of the futuristic applications of MANET. Figure 1 shows example applications of wireless MANETs. Obviously, security is a critical issue in such areas. Due to such charactertistics,the Wireless Adhoc networks are highly susceptible to malicious attacks.

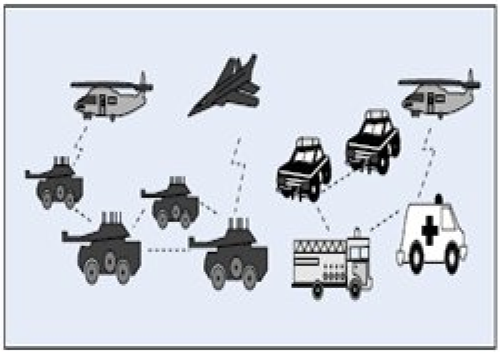


Fig 1: Applications of Wireless MANET

They need harder security than conventional wired and static internet. Irrespective of the number of Intrusion prevention schemes implemented in the Wireless Adhoc networks ,there will be vulnerable point in the network through which intruder can break in. Intrusion prevention measures such as encryption and authentication, at times fail to identify the attack, as these prevention measures cannot defend against compromised mobile nodes that carry private keys and easily authenticate themselves. Hence, to create a highly secured adhoc network, an intrusion detection system in the network to create another wall of defence is needed to be implemented.

The node after being part of the network could be compromised in such away that the incorrect and malicious behavior cannot be directly noted at all. Although the compromised nodes may appear to operate correctly, they make use of the flaws and inconsistencies in the routing protocol. Malicious nodes can also create new routing messages and advertise non existent links, provide incorrect link state information and flood the nodes with routing traffic. Such failures are severe especially because they may come from seemingly trusted nodes, whose malicious intentions have not yet been noted.Hence there are two types of attacks

External Attack: An attack caused by nodes that do not belong to the network.

Internal Attack: An attack from nodes that belong to the network due to them getting compromised or captured.

Ad hoc On-demand Distance Vector Routing (AODV) —

The AODV protocol uses route request (RREQ) messages flooded through the network in order to discover the paths required by a source node refer Fig 2. An intermediate node that receives a RREQ replies to it using a *route reply* message only if it has a route to the destination whose corresponding destination sequence number is greater or equal to the one contained in node replies to a RREQ only if it has a *fresh* enough route to the destination. Otherwise, an intermediate node broadcasts the RREQ packet to its neighbors until it reaches the destination. The destination unicasts a RREP back to the node that initiated the route discovery by transmitting it to the neighbor from which it received the RREQ.

As the RREP is propagated back to the source, all intermediate nodes set up forward route entries in their tables. The route maintenance process utilizes link-layer notifications, which are intercepted by nodes neighboring the one that caused the error. These nodes generate and forward route error (RERR) messages to their neighbors that have been using routes that include the broken link. Following the reception of a RERR message a node initiates a route discovery to replace the failed paths refer Fig 3.

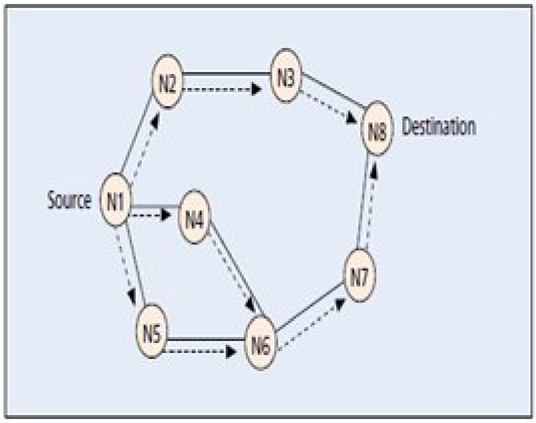


Fig 2: Propagation of RREQ

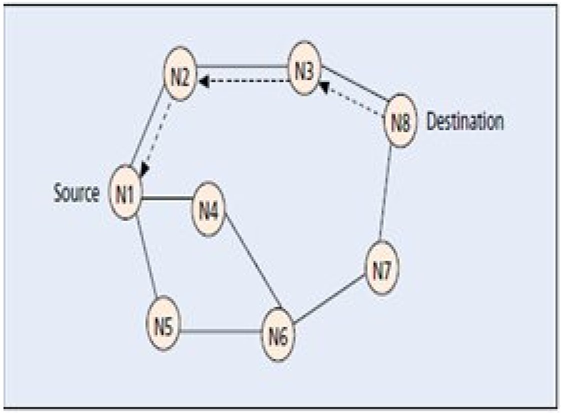


Fig 3: The path of routing reply

Analysis of the proposed secure adhoc routing protocols in a structured way is classified into two categories symmetric and asymmetric cryptographic solutions. This is the most common approach to digitally sign the immutable fields of routing messages in order to provide integrity and authentication, and also use hash chains to protect the hop count metric. Secure On-demand Distance Vector Routing (SAODV) is used to solve most of the issues present in SAODV.

III.Adhoc network characteristics

Mobility : The fact that the nodes can be rapidly repositioned and/or move is one of the essential feature of an adhoc network Rapid deployment in areas with no infrastructure often implies that the users must explore an area and perhaps from teams/groups that in turn coordinate among themselves to create a taskforce or a mission. The mobility model can have major impact on the selection of a routing scheme and can thus influence performance.

Multihopping: A multihop network is a network where the path from source to destination traverses several other nodes.Adhoc networks often exhibit multiple hops for obstacle negotiation, spectrum reuse, and energy conservation. Battlefield covert operations also favor a sequence of short hops to reduce detection by the enemy.

Self organization: The Adhoc network must autonomously determine its own configuration parameters including addressing,routing,clustering,position identification, power control, etc. In some special nodes (eg., mobile backbone nodes) can coordinate their motion and dynamically distribute in the geographic area to provide coverage of disconnected islands

Energy Conservation: Most Adhoc nodes(., laptops,PDAs,sensors,etc) have limited power supply and no capability to generate their own power(eg., solar panels). Energy efficient protocol design (eg., MAC,routing,resource discovery,etc) is critical for longevity of the mission

Scalability: For wireless infrastructure networks scalability is handled by a hierarchical construction. The limited mobility of infrastructure networks can be handled using mobile IP or local handoff techniques. Because of more extensive mobility and lack of fixed references pure adhoc networks do not tolerate mobile IP or fixed hierarchy structure. Thus mobility jointly with large scale is one of the most critical challenges in the adhoc design

Security : Adhoc networks are more vulnerable to attacks, active and passive attacks are possible. Due to complexity of Adhoc network protocols active attacks are more difficult to detect. Passive attacks are unique of adhoc networks and will be more insidious than the active ones. The active attacker is discovered and physically disabled. The passive attacker is never discovered by the network like a bug it is placed in a sensor field. Defense from passive attacks require powerful novel encryption techniques coupled with careful network protocol designs.

IV. Solutions for Secure Protocols

Secure Adhoc Routing

There exist several proposals that attempt to architect a secure routing protocol for ad hoc networks, in order to offer protection against the attacks mentioned in the previous section. These proposed solutions are either completely new stand-alone protocols, or in some cases incorporations of security mechanisms into existing ones. The design of these solutions focuses on providing countermeasures against specific attacks, or sets of attacks. Furthermore, a common design principle in all the examined proposals is the performance-security trade-off balance. Since routing is an essential function of ad hoc networks, the integrated security procedures should not hinder its operation. Another important part of the analysis is the examination of the assumptions and the requirements that each solution depends on. Although a protocol might be able to satisfy certain security constraints, its operational requirements might thwart its successful employment.

A.Asymmetric Cryptography Solutions

Protocols that use asymmetric cryptography to secure routing in mobile ad hoc networks require the existence of a universally trusted third party (TTP). The TTP issues certificates that bind a node's public key with a node's persistent identifier. Furthermore, the TTP can be either online or offline. In approaches that use an online TTP, revocation of the issued certificates is accomplished by broadcasting certificate revocation lists (CRLs) in the network. In offline systems revocation becomes a particularly complicated problem and usually involves the exchange of recommendations between the participating nodes.

B.Symmetric Cryptography Solutions

This category presents solutions that rely solely on symmetric cryptography to secure the function of routing in wireless ad hoc networks. The most commonly utilized mechanisms are *hash functions* and *hash chains*. A one-way hash function is a function that takes an input of arbitrary length and returns an output of fixed length Hash functions have the property of being computationally expensive to reverse i.e. if $h = f(m)$, it is hard to compute m such as
$f(m) = h$.

A hash chain can be generated by applying repeatedly a given hash function to a random number known as the *root* of the chain. Simply stated, in order to generate a hash chain of length n a hash function is applied n times to a random value p, and the final hash q that is obtained is called the *anchor* of the chain. In order to use a hash chain for authentication purposes an initial authenticated element of the chain is assumed, usually the anchor. Given this it is possible to verify the authenticity of the elements that come later in the sequence. Since hash functions are especially lightweight when compared to other symmetric and asymmetric cryptographic operations, they have been extensively used in the context of securing ad hoc routing and specifically in hop count authentication.

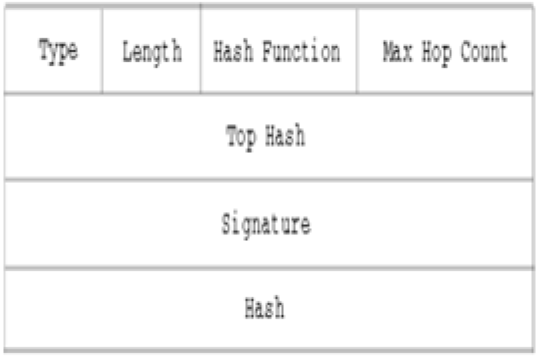
C.Hybrid Solutions

These secure routing protocols employ both symmetric and asymmetric cryptographic operations. The most common approach is to digitally sign the immutable fields of routing messages in order to provide integrity and authentication, and to use hash chains to protect the hop count metric.

V.Secure Ad hoc On-demand Distance Vector Routing (SAODV) .

Secure Ad hoc On-demand Distance Vector (SAODV) is a proposal for security extensions to the AODV protocol. The proposed extensions utilize digital signatures and hash chains in order to secure AODV packets. In particular, cryptographic signatures are used for authenticating the non-mutable fields of the messages, while a new one-way hash chain is created for every route discovery process to secure the hop-count field, which is the only mutable field of an AODV message. Since the protocol uses asymmetric cryptography for digital signatures it requires the existence of a key management mechanism that enables a node to acquire and verify the public key of other nodes that participate in the ad hoc network.

In order to facilitate the transmission of the information required for the security mechanisms, SAODV defines extensions to the standard AODV message format. These SAODV extensions consist of the following fields. The *hash function* field identifies the one-way hash function that is used. The field *max hop count* is a counter that specifies the maximum number of nodes a packet is allowed to go through. The *top hash* field is the result of the application of the hash function *max hop count* times to a randomly generated number, and finally the field *hash* is this random number, as shown in Figure 4



Type	Length	Hash Function	Max Hop Count
Top Hash			
Signature			
Hash			

Fig 4: SAODV protocol header

When a node transmits a route request or a route reply AODV packet it sets the *max hop count* field equal to the *time to live* (TTL) field from the IP header, generates a random number and sets the *hash* field equal to it, and applies the hash function specified by the corresponding field *max hop count* times to the random number, storing the calculated result to the *top hash* field. Moreover, the node digitally signs all fields of the message, except the *hop count* field from the AODV header and the *hash* field from the SAODV extension header. An intermediate node that receives a route request or a route reply must verify the integrity of the message and the *hop count* AODV field. The integrity requirement is accomplished by verifying the digital signature. The *hop count* field is verified by comparing the result of the application of the hash function *max hop count* minus *hop count* times to the *hash* field with the value of the *top hash* field. Before the packet is re-broadcasted by the intermediate node the value of the *hash* field is replaced by the result of the calculation of the one-way hash of the field itself in order to account for the new hop

The main problem with securing an on-demand protocol like AODV is that it allows intermediate nodes with *fresh* routes to reply to a route query since the reply has to be signed on behalf of the destination node. In order to overcome this problem the authors suggest two solutions. The first one is to forbid intermediate nodes to respond to route request messages since they cannot sign the message on behalf of the final destination. The second solution involves the addition of the signature that can be used by intermediate nodes to reply to a route request by the node that originally created the route request.

In SAODV route error messages (RERR) that are generated by nodes that inform their neighbors that they are not going to be able to route messages to specific destinations are secured using digital signatures. A node that generates or forwards a route error message cryptographically signs the whole message, except the destination sequence numbers as shown in figure 5.

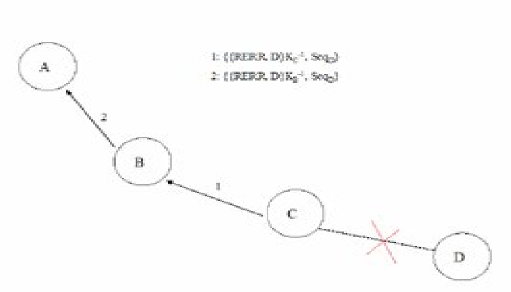

1: {(RERR, D)K_C^{-1}, Seq_D}
2: {(RERR, D)K_B^{-1}, Seq_D}
A
2
B
1
C
D

Fig 5: Route maintenance in SAODV

Since the destination does not authenticate the destination sequence number, the authors suggest that a node should never update the destination sequence numbers of the entries in its routing table based on route error messages. Route error messages are still useful in SAODV in order to allow a node to decide whether it should completely remove a route from its routing table or not.

SAODV is a set of security extensions to the AODV protocol. In SAODV every route discovery that is initiated by a node corresponds to a new one-way hash chain. The elements of the chain are used in order to secure the metric field in the route request packets.

Conclusion

This survey has presented the well known protocols for securing the routing function in mobile adhoc networks. The analysis of the different proposals has demonstrated that the inherent characteristics of adhoc networks, such as lack of infrastructure and rapidly changing topologies, introduce additional problems to the already complicated problem of secure routing.

Military applications of adhoc networks are probably that requires the most highly secure routing functionality. In this case the applications that run on top of the network are of critical importance, therefore the underlying routing process should provide a high level of protection, while possibly having less strict performance requirements On the other hand; commercial application scenarios of adhoc networking may have higher performance requirements of the underlying routing protocol. However security plays a important role since in commercial or domestic adhoc environments

the exchanged information is usually confidential,such as credit numbers,or of private nature.Therefore a flexible secure adhoc routing solution should be taken into account the performance-seurity trade-off associated with an application and dynamically achieve the required equilibrium.Research on Adhoc network security is still in its early stage.The future scope is explained below

References

[1].Patroklos g. Argyroudis and donal o'mahony, university of dublin, trinity college "Secure routing for Mobile ad hoc networks" third quarter 2005, volume 7

[2]. Zygmunt J. Haas, Jing Deng, Ben Liang, Panagiotis Papadimitratos, and S. Sajama "Wireless Ad Hoc Networks"2003

[3]. Hongmei Deng, Wei Li, and Dharma P. Agrawal, University of Cincinnati "Routing Security in Wireless Ad Hoc Networks" IEEE Communications Magazine • October 2002

[4].C.E.Perkins,E.M.Royer and S.Das, "Adhoc On-Demand Distance Vector(AODV),"RFC3561,july 2003.

Maskless Lithography: A Revolutionary Chip Making Technique

Naseeruddin
Ballari Institute of Technology & Management,Bellary,naseer_ece@rediffmail.com.

Abstract

The VLSI design industry is resource intense engineering discipline. Projects and products definition are economically motivated worldwide. The industry offers tremendous potentiality in both domains. The market window is often short due to both competition and changing consumer demands. With knowledge of dynamic nature of market the discussion starts with the basic information on semiconductor device fabrication process. The process is called as lithography. There are numerous methods of lithography available for IC fabrication and the choice among them is made by considering various tradeoff's in the industry. Since the advent of the semiconductor edge, optical lithography is widely used method for the fabrication. Lithography system uses masks to produce the desired pattern on Si-wafer. Utilization of masks raises the NRE cost because for a complete chip fabrication there is need of more the 25 sets of masks. In future as the transistor density increases on the chip, the cost incurred on generating mask will increase exponentially. This paper presents a revolutionary Lithography (chip making) method called maskless Lithography which totally avoids the usage of masks thus reducing the NRE cost incurred on the mask generation. In this approach masks are software based and micro -mirrors are used to transfer the desired pattern on the wafer, which is scalable which is still under development and would be boon for the ASIC industry in particular.

Index Terms: optical proximity-effect correction (OPC), Maskless optical Projection Lithography (MOPL),

I.INTRODUCTION

The first chip was invented by Jack kilby of Texas instrument in the year 1958. This marked The dawn of the VLSI design industry [7].Every semiconductor device fabrication has to pass through UNIT processes. The NRE cost invested in the device fabrication can be reduced if any positive changes can be made in the existing system of UNIT processes. One such change is possible in lithography process is presented in the paper called as Maskless lithography process. This paper is organized into six sections Section I: Gives introduction and importance of the cost effective changes in future.

Section II: Describes different lithography methods. Section III: Explains major steps in OPL and Crisis in lithography and information on the MASK industry.Section IV: Describes Maskless Lithography Process, types, advantages. Section V: Explains fabrication of micromirrors Section VI: Gives advantages, challenges of Maskless lithography, comparisons and conclusions.

Lithography literally means:*"printing"*. It is the creation of a wire patterns on a substrate to build IC which is done on micro scale through LASER or E-beam.

II.Types of Lithography [8]:

Semiconductor Lithography
Immersion Lithography
Optical projection Lithography
Nano Lithography
    E beam Lithography
    Scanning Probe Lithography
    X-ray Lithography.

This paper presents MOPL (Maskless optical Projection Lithography in detail.

IIIA.Major Steps in OPL[3]

A SI wafer is prepared for photolithography by coating it with a layer of Si nitride followed by a layer of Si dioxide and finally a layer of photo resist.

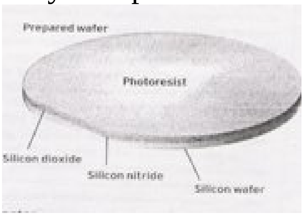


Fig: 1

Light from an illuminator is projected through the mask are reduced by a factor of four by focusing lens and projected onto the photoresist coated wafer. This step exposes the one chip on the wafer and the process is repeated for all chips on the wafer. Refer Fig 1

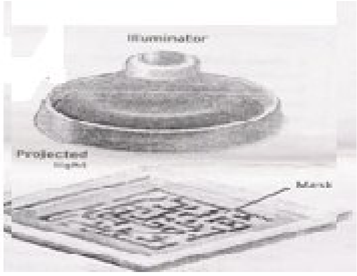


Fig:2

The photo resist that is exposed to the light become soluble and is rinsed away, leaving a miniature image of the mask pattern at each chip location.Refer Fig 2.

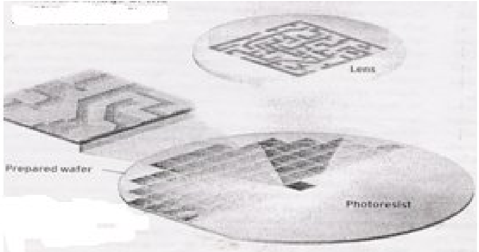


Fig:3

Regions unprotected by photoresist are etched away by gases removing the Si dioxide and Si nitride and exposing the Si Refer Fig 3

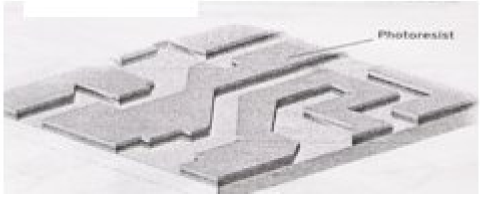


Fig:4

Impurities can be added to the etched areas, changing the electrical properties of the Si as needed t form the transistor Refer Fig 4.

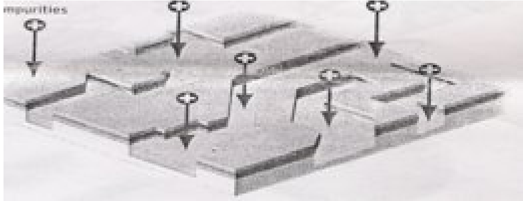


Fig:5

Step 1 to 5 are repeated for as many layers as required to form the transistor. Metal lines that connect transistor are made using photolithography as well. Refer Fig 5

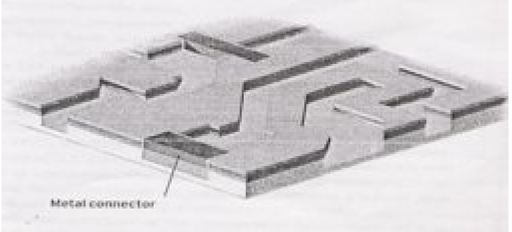


Fig:6:Final Pattern

IIIB.Optical photolithography [3]

Limitations of OPL:

Each reticle contains pattern for only part of the wafer.25-30 sets of reticle per chip (with 130 nm technology) required.

The crisis in lithography

In the semiconductor industry, the crisis includes: the difficulty, cost, and delay associated with designing the OPC features for a mask; the cost and long delays in making the masks; and the uncertainties associated with future scaling of transistor dimensions. The difficulty, time, and cost associated with designing and repairing a mask leads chip designers to be conservative, which stifles innovation.

Information on the MASK industry [12]

A modern mask shop - $200 to $500 million USD ! **Photronics** USA ($349mn),**Dai Nippon Printing Company** JAPAN($499mn) Giant Chipmakers –Intel , IBM , Samsung.

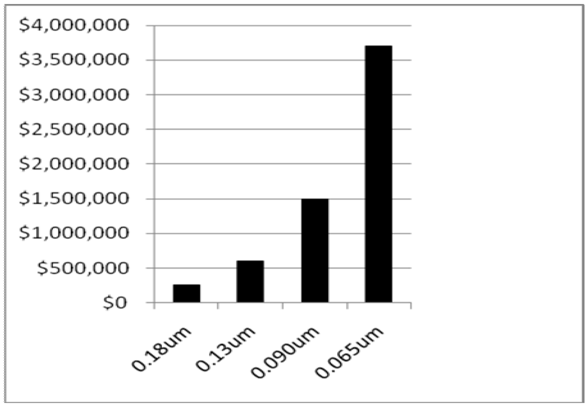


Fig 7: Mask Set cost Vs Technology Node showing the upcoming "Million Dollar Mask Set". [12]

The Big picture

25 – 30 sets of reticles per chip –a cost factor of *$500,000 USD* !
Next generation device at 90nm node – reticle set expected cost -> *$1 million USD* !!
Further at 65nm node, cost per set-*$3 million USD* !!!

MICRONIC LASER SYSTEMS AB (Sweden) & ASML (the Netherlands): The Solution

These two European companies teamed up to built a lithography system that replaces the reticle with arrays of millions of microscopic mirrors, which would direct the LASER light to the wafer according to patterns fed to them by a computer [2].

IV.MASKLESS LITHOGRAPHY

A.Forms of Maskless Lithography [9]

Scanning electron beam lithography (SEBL)
Focused ion beam lithography (FIB)
Multi –axis electron beam lithography (MABEL)
Interference lithography (IL)
Maskless optical projection lithography (MOPL)
Zone plate Array lithography (ZPAL)
Scanning probe lithography (SPL)
Dip pen lithography (DPL)

SEBL: Has been used widely in research for decades and to make mask for the semiconductor industry. The throughput depends on pattern resolution. This limits its application for smaller line widths.

MABEL: Has been proposed as a solution for this throughput problem. However its commercial availability is uncertain.

In general problem with SEBL or any other forms of electron or ion-lithography is pattern placement accuracy. Temperature gradients, stray magnetic fields, sample charging, vibrations and variety of other effects cause the electron beam to deviate from its intended position placement.

IL: Has many forms and throughput is quite high. However it is only applicable to periodic and quasi periodic patterns.

SPL and DPL: Have number of unique aspects which play an important role in nanostructure research.

MOPL: System has been proposed that replaces mask with programmable micromirror array.

B.Components of the system:[13]

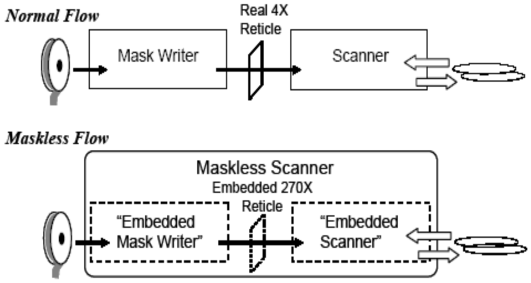


Fig 8: Shows the difference between normal flow and Maskless flow. Only the reticle is replaced by a programmable micromirror array but whole system has to be changed. It requires large amount of data and data rate should also be high to meet the current market needs

C.Lithography Equipment Using Micromirrors[11]

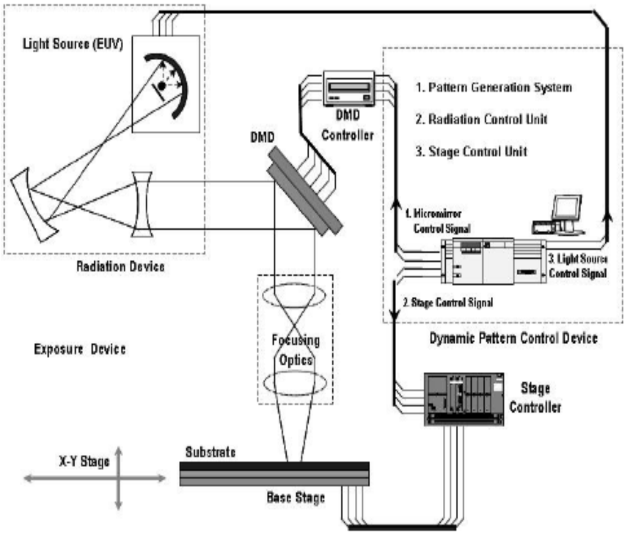


Fig 9: Prototype of the complete system of Lithography

The maskless lithography for the micromirror based pattern generation system consists of three major devices.

1. Radiation device: includes light sources: Optics

2. Exposure Device: Includes micromirror controller, micromirror, focusing optics, photo resistant coated glass substrate and the base stage assembly.

3. Dynamic pattern control device: Includes lithography pattern generation system radiation control unit and the stage control unit.

D.Pattern generation process:

The function of the first pattern procedure is loading the CAD data written in DXF through parsing of the CAD data.

Function of the second procedure is recognition of the pattern upon the substrate, and it is associated with two routines.The extraction of the pattern boundry is accomplished through the reconstruction process of geometric entities with open loops into closed loops. The construction of the pattern region is done by set operations on polygons upon computational geometry

The function of the third procedure is recognition of the pattern upon micromirror and it is associated with three routines. The confirmation of the micromirror dependent lithographic pattern region is performed in accordance with the micromirror configuration

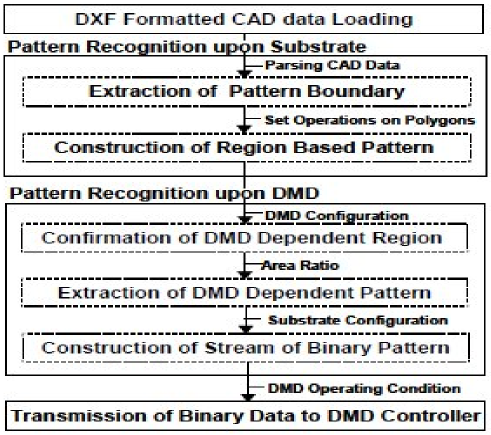

DXF Formatted CAD data Loading
Pattern Recognition upon Substrate
Parsing CAD Data
Extraction of Pattern Boundary
Set Operations on Polygons
Construction of Region Based Pattern
Pattern Recognition upon DMD
DMD Configuration
Confirmation of DMD Dependent Region
Area Ratio
Extraction of DMD Dependent Pattern
Substrate Configuration
Construction of Stream of Binary Pattern
DMD Operating Condition
Transmission of Binary Data to DMD Controller

Fig 10: Maskless pattern generation Flow

(Note: Where DMD- Digital Micromirror deviceDXF-Drwaing eXchange Format)

V.Fabrication of Micromirrors [14]

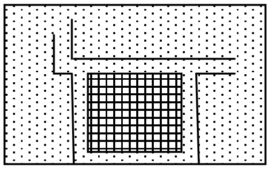


Fig 11:Micromirror Layout

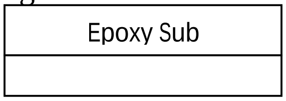


Fig 12:Step1

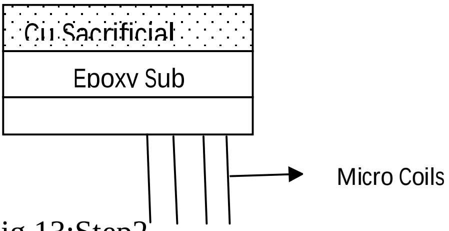
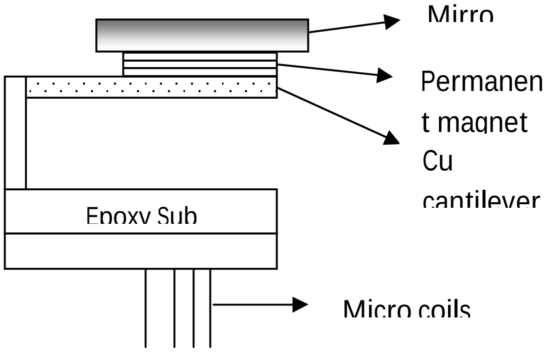


Fig 13:Step2



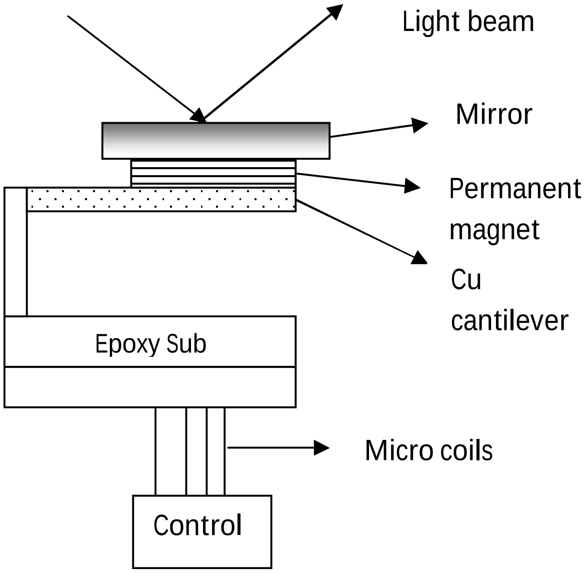


Fig 14: Step 4

First a Photoresist sacrificial layer is spin coated and patterned on the substrate then the Cr-Cu-Cr or Ti-Cu-Cr seed layer is deposited to perform the Cu Electroplating.

Second Photo resist layer is spun and patterned to serve as mould for the electroplating the 'Cu' based cantilever beam. The Cu cantilever beam is electroplated in the Cu-Sulfate based plating path. After the electroplating the Photoresist plating mould and the seed layer are removed releasing the cantilever beam structure. Depending on the permanent magnet used, the corresponding fabricated processes must be done before and after releasing the beam.

The permanent magnetic disk is positioned on the cantilever's free end. For example the polymer magnet can be “Screen” Printed after curing the epoxy magnet, the magnet is then magnetized by the external magnetic field.

Then the cantilever beam with fabricated mirror is released by removing the sacrificial layer using Acetone.

A.Comparison

Conventional	Maskless lithography
Very high NRE cost.	Very low.
More turn around time	Small.
Slow time to market	Quicker –boom for ASICs Industry.
Mask coated with chrome are used.	Masks are s/w based
High throughput –100 wafer/hour	Low - wafer/hour
Nominal BW required	Very high B W required

Table 1: Comparison between Conventional and Maskless Flow

Advantages of Maskless Lithography [10]

1. The physical carrier of the master pattern is eliminated. The master pattern resides in the computer memory in its purest form: information.
2. The master pattern is agile. Since it is resident in computer memory it can be easily upgraded or modified. Changes can be transmitted to all lithography systems in a factory and to all factories in the world.
3. Redundancy: the redundant pixel design is forgiving of single point defects so that pattern need not be free. This is a major cost saver.
4. No need for costly mask management outside the lithography system.
5. No physical transport of masks in/out of the stepper, and no scanning masks. This significantly reduces the complexity and cost of stage hardware in stepper systems.

6. Conventional mask manufacturing technology (pattern writing, inspection, repair, substrate materials etc..,) is eliminated, removing the major impediment to the continued performance growth of lithography system.

B.Challenges

1. Design of a mirror system
2. MEMS mirror if fails –more difficult to replace than reticle.
3. To achieve throughput of 1 wafer/minute per chip using 25nm pixel for 50nm feature requires data rate of 15 Tb/S !
4. Compression and decompression of layout data.
5. Powerful decoder circuitry to handle input at 400 Gb/S to output at 15 Tb/S

Solutions: reference [12] gives feasible solutions for challenges “3,4 and 5”.

Conclusion

A novel approach for ***Maskless lithography*** has been presented as a means to greatly reduce ***Non –Recurring Engineering Cost (NRE)*** in the fabrication of the future process generation. Though throughput is very low and don’t compete with the present trend but for makers of low volume product ***ASIC’s***, the reduced cost will make up for lower throughput and Time to market will be quicker.

References

[1]. B J Warlick and J Garett "Micromirror Array for maskless Lithography",Quantum electronics and LASER science conference ,2002.

[2].News analysis " A revolutionary chip making Technique ",IEEE spectrum Nov 2003,Page no:18.

[3]. Semiconductors ," A little light magic ",IEEE spectrum Sep 2003,Page no:34-39.

[4]. Craver Mead –Lynn Conway," Introduction to VLSI system's ,BS publication Edition :2003.

[5]. Neil H Weste & Kamran Eshraghain ,"Priciples of CMOS VLSI Design",Pearson education Edition:2005

[6].Arthut Besier"Concepts of Modern physics",Michelson-Morley Experiment

[7].Shivaprasad M Khened ," Tribute to Jack Kilby, A gentle giant of Miniaturization", History of science, Dream 2007,Nov:2005,Page:22-25.

[8]. Lithogarphy Challenges," The 2003 interbational conference on characterization & Metrology for ULSI technology",American institute of Physics,page:365.

[9]. Rajesh Menon et el..," Maskless lithography ", Material today,Feb 2005,page:26-33.

[10]:J A Folta, et el..," High density arrays of Micromirrors", A informal report ,Lawrenc Livemore National Laboratory (Under US department of Energy),Feb 1999.

[11]. Manseung Seo ,et el.., " Maskless Lithogarphy Pattern generation system upon Micromirrors" CAD and Applications ,Vol:3,1-4,2006.

[12]. Borivoje Nikolic,Ben Wild et.el..," Layout decompression chip for Maskless Lithogarphy ,Dept. of Electrical Engineering and Computer Science ,university of California.

[13]. David Lammers "Semi road map adds imprint,Maskless Lithography ",EE times Dec. 8,2008.

[14]. Sergey Edward Lyshevvski" Nano – and Micromechanical Systems ",2nd Edition ,CRC process 2005,Page :675-676.

Simple Calibration Technique for 8 Bit ADC Used in Voltmeter

P. A. Kadam, R. R. Mudholkar, S. S. Nirmale, M. S. Patil

Department of Electronics, Shivaji University, Kolhapur, 416004.

E-mail: pak_eln@unishivaji.ac.in, neuralfuzzy.lab@gmail.com

Abstract: In this Paper a very simple but effective method of calibrating 8-bit ADC for use as a 0–5V voltmeter is suggested. The design uses National Semiconductor's 8-bit microprocessor compatible A/D Converter ADC0804, the ATMEL AT89C2051 microcontroller and the 16*1 LCD display. The calibration introduces an additional algorithmic error of 1 LSB. But the added simplicity can make this design useful for many applications where great accuracy is not required.

1. INTRODUCTION:

The Use of data converters is confined to a very limited set of applications when used without calibration. Calibration is the most essential part of the data acquisition system. There are many methods of digital calibration of ADCs. The use of microcontroller adds the new dimensions to the digital calibration of ADCs.

2. THEORETICAL:

The ADC0804 is a CMOS 8-bit successive approximation A/D converter that uses a differential potentiometric ladder DAC. These converters are designed for easy interfacing with microprocessors/microcontrollers, for this purpose these ADCs have TRI-STATE output latches directly driving the data bus. These A/Ds appear like memory locations or I/O ports to the microprocessor and no special interfacing logic is required. Differential analog voltage inputs allow increasing the common-mode rejection and offsetting the analog zero input voltage value. In addition, the voltage reference input can be adjusted to allow encoding any smaller analog voltage span to the full 8 bits of resolution. This ADC has been designed to accommodate a 5V, 2.5 V or an adjusted DC voltage reference. This has been achieved in the design of the IC. This arrangement provides maximum applications flexibility. IC voltage regulators may be used for references if the ambient temperature changes are not excessive.

The ADC 0804 has a conversion time of 100 µSec, which is quite sufficient for many general purpose applications. The other features of this ADC are:

- Easy interface to all microprocessors, or operates "stand alone" (Free Running mode).
- Differential analog voltage inputs.
- Logic inputs and outputs meet both MOS and TTL voltage level specifications;
- On-chip clock generator.
- No zero adjustment is required.
- 0V to 5V analog input voltage range with single 5V supply

The ADC0801 series contains a circuit equivalent of the 256R network. Analog switches are sequenced by successive approximation logic to match the analog difference input voltage$[V_{(+)} - V_{(-)}]$ to a corresponding tap on the R network. The most significant bit is tested first and after 8 comparisons (64 clock cycles) a digital 8-bit binary code (1111 1111 = full-scale) is transferred to an output latch and then an interrupt is asserted (INTR makes a high-to-low transition). Issuing a second start command can interrupt a conversion in process. The device may be operated in the free-running mode by connecting INTR to the WR input with CS=0. To ensure start-up under all possible conditions, an external WR pulse is required during the first power-up cycle.

On the high-to-low transition of the WR input the internal SAR latches and the shift register stages are reset. As long as the CS input and WR input remain low, the A/D will remain in a reset state. Conversion will start from 1 to 8 clock periods after at least one of these inputs makes a low-to-high transition.

The digital control inputs (CS, RD, and WR) meet standard TTL logic voltage levels. These inputs are active low to allow an easy interface to microprocessor control busses.

The clock for the ADC can be derived from the CPU clock or an external RC can be added to provide self-clocking. The CLK IN pin 4 makes use of a Schmitt trigger. Heavy capacitive or DC loading of the clock R pin should be avoided as this will disturb normal converter operation. The clock frequency is given by the formula:

$$f_{CLK} \cong \frac{1}{1.1\ RC}$$

$$R \cong 10\ k\Omega$$

Noise spikes on the VCC supply line can cause conversion errors as the comparator will respond to this noise. A low inductance tantalum filter capacitor should be used close to the converter VCC pin and values of 1 µF or greater are recommended. If an unregulated voltage is available in the system, a separate 5V voltage regulator for the converter (and other analog circuitry) will greatly reduce digital noise on the VCC supply.

The AT89C2051 is a low-voltage, high-performance CMOS 8-bit microcomputer with 2K bytes of Flash programmable and erasable read only memory (PEROM). The device is manufactured using Atmel's high-density nonvolatile memory technology and is compatible with the industry-standard MCS-51 instruction set. By combining a versatile 8-bit CPU with Flash on a monolithic chip, the Atmel AT89C2051 is a powerful microcomputer which provides a highly-flexible and cost-effective solution to many embedded control applications.

The AT89C2051 provides the standard features like: 2K bytes of Flash, 128 bytes of SRAM, 15 I/O lines, two 16-bit timer/counters, a five vector two-level interrupt architecture, a full duplex serial port, a precision analog comparator, on-chip oscillator and clock circuitry. In addition, the AT89C2051 is designed with static logic for operation down to zero frequency and supports two software selectable power saving modes. The Idle Mode stops the CPU while allowing the RAM, timer/counters, serial port and interrupt system to continue functioning. The power-down mode saves the RAM contents but freezes the oscillator disabling all other chip functions until the next hardware reset.

The microcontroller has 15 I/O lines, divided into two ports viz P1 and P3. Port 1 is an 8-bit bidirectional I/O port. Port pins P1.2 to P1.7 provide internal pullups. P1.0 and P1.1 require external pullups. P1.0 and P1.1 also serve as the positive input (AIN0) and the negative input (AIN1), respectively, of the on-chip precision analog comparator. The Port 1 output buffers can sink 20 mA and can drive LED displays directly. Port 3 pins P3.0 to P3.5, P3.7 are seven bidirectional I/O pins with internal pullups. P3.6 is hard-wired as an input to the output of the on-chip comparator and is not accessible as a general purpose I/O pin. The Port 3 output buffers can sink 20 mA. Port 3 also serves the functions of various special features of industry-standard MCS-51 family.

With a limited hardware facility the microcontroller AT89C2051 applies some restrictions on certain instructions. All the instructions related to jumping or branching should be restricted such that the destination address falls within the physical program memory space of the device, which is 2K for the AT89C2051. This should be the responsibility of the software programmer. For example, LJMP 07E0H would be a valid instruction for the AT89C2051 (with 2K of memory), whereas LJMP 0900H would not. External DATA memory access is not supported in this device, nor is external PROGRAM memory execution. Therefore, no MOVX [...] instructions should be included in the program.

The dot matrix liquid crystal display (16*1 LCD) can display alphanumeric and symbols. The built-in controller & driver LSIs provide convenient connectivity between a dot matrix LCD and most 4 or 8 bit microprocessors or microcontrollers. All the functions required for a dot matrix liquid crystal display drive are internally provided. The CMOS technology makes the device ideal for Applications in hand held, portable and the other battery powered instruments with low power consumption.

The LCD display has following features:

- Easy interface with a 4-bit or 8-bit MPU.
- Built-in Dot Matrix LCD Controller with font 5×7 dots.
- Character generator ROM, which provides 160 characters with font 5×7 dots.
- Internal automatic reset circuit at power ON.
- Built-in Oscillator circuit. (No external clock required)
- Wide range of instruction functions: Clear Display, Cursor Home, Display ON/OFF, Cursor ON/OFF, Cursor Shift, Display Shift.

Fig. 1 shows the internal block diagram of the LCD module. The module is powered by the supply of $5V_{dc\ Regulated}$. The supply voltage is applied between the terminals VDD and VSS. The voltage applied at the terminal V0 controls the LCD contrast. The module has three control lines viz. RS (Register select), R/W (Read / Write), and E (Enable). The facility of backlighting is provided and the diodes can be illuminated by providing proper bias across A and K terminals.

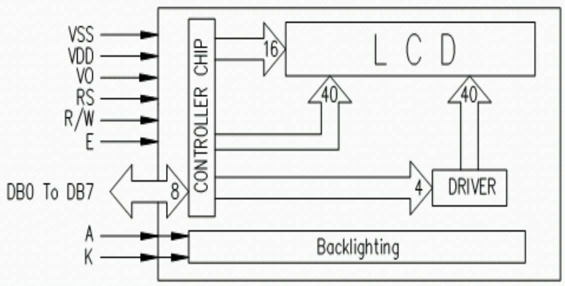


Fig. 1. LCD display internal block diagram

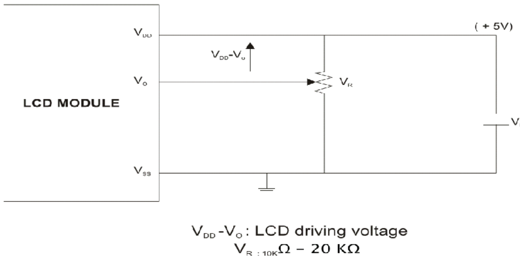


Fig. 2. LCD contrast adjustment

3. PRACTICAL:

The basic block diagram is as shown in the fig. 3. It consists mainly three blocks: ADC, Microcontroller and the LCD display. The analog signal voltage is applied to the input of the ADC block, which produces the digital approximation of the applied analog input voltage. The digital word is read by the microcontroller and calibrated to the useful range. The data is then converted to ASCII format and sent to LCD module for display.

The heart of the system is a microcontroller unit which employs ATMEL microcontroller AT89C2951. Fig. 4(a,b,c) shows the system layout organized around the MCU. The clock frequency is selected to be 11.0592 MHz in view of future planning, which may include serial data logger design using RS232 link with PC. The ADC data bus is connected to the port P1, while the control signals WR, RD and INTR are connected as below:

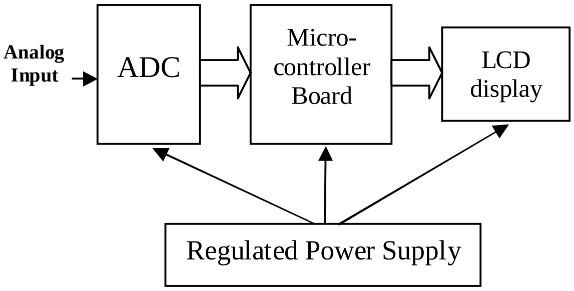


Fig. 3. System block diagram

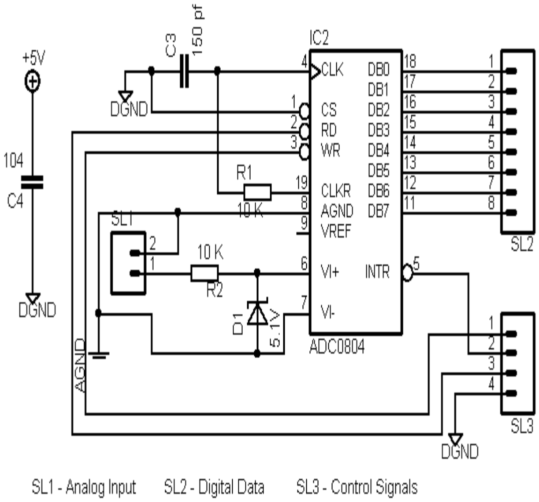


Fig. 4b. ADC unit circuit diagram

WR	:	P3.2
INTR	:	P3.3
RD	:	P3.4

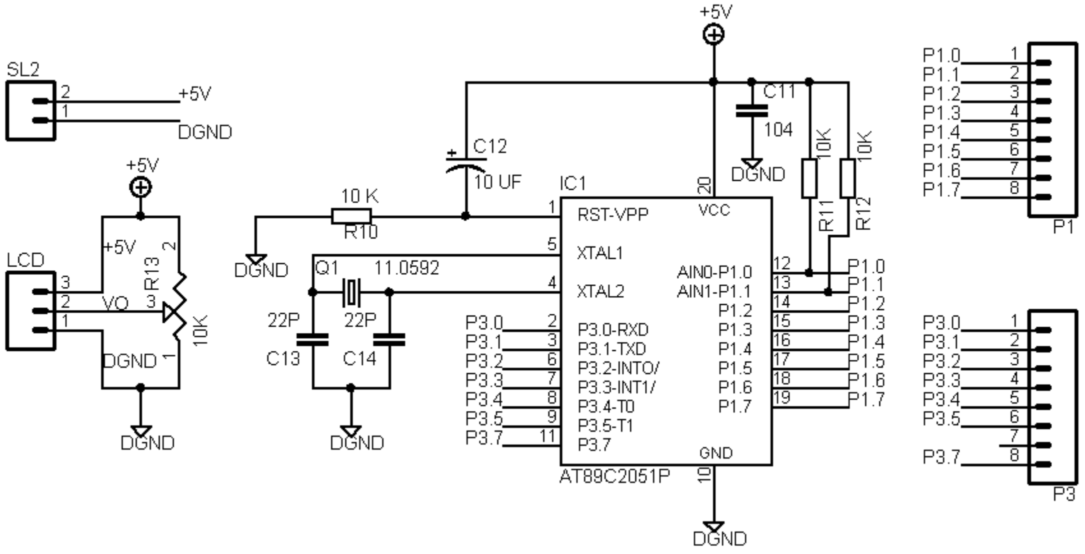


Fig. 4a. Microcontroller unit Circuit diagram

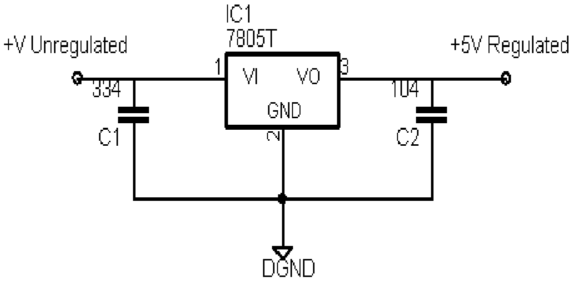


Fig. 4c. Power Supply unit circuit diagram

The ADC unit is wired as shown in the diagram. The analog ground point is separated from the digital ground points, to prevent digital noise from affecting the preciseness of ADC. A 5.1V zener diode is used to protect the analog input terminal of ADC in case excessive voltages are applied.

The system is powered by +5V regulated power supply obtained by utilizing National Semiconductor's Series Voltage Regulator IC LM7805. This regulator is easy to use and do not require any external components for regulatory action. Thus it is preferred in almost all non-critical, low power applications.

The software is written in MCS-51 Assembly language and transferred to the on-chip flash memory of the microcontroller. The flow chart of the program is shown in fig. 5. The major steps include Reading ADC, calculating result and display. On power up, the program executes this sequence infinitely until the power is disconnected. This gives continuous display of applied voltage to the ADC input.

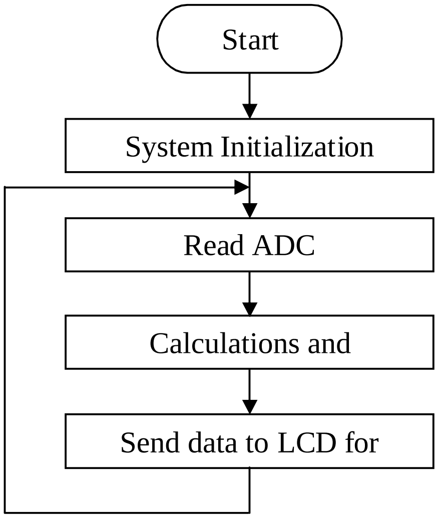


Fig. 5. Flow chart

The software written in Assembly Language is presented in the listing no. 1. The MAIN routine makes a call to the INIT, READ_ADC, CONVERT and DISPLAY subroutines. Function of each subroutine is listed below:

- The INIT routine initializes the operating environment for the system, by doing port configurations and initializing the LCD display with proper functionality.

- The READ_ADC routine reads the ADC data in handshake mode and returns the result in the accumulator. The sequence of the operations is 1) send start-of-conversion signal on WR pin (active low). 2) Poll INTR pin, which when low indicates End-of-conversion. 3) Read the ADC data by making its RD pin low. The data will be available on port P1.
- The Convert routine calibrates the ADC data to that of the input voltage. The system is designed to measure a range of 0-5 Volts. Thus a simple mathematical relation is used for the calibration. It includes three steps of calculations, which produce the three digit output ranging from 0.00V to 5.00V.

```
-   Step 1:
            R2 = (ADC_DATA/51),
            B= (ADC_DATA % 51)
-   Step 2:
            If (B>49)
            {
                    B=49
            }
            R3 = (B / 5),
            B = (B % 5)
-   Step 3:
            If (B>4)
            {
                    B=4
            }
            R4 = (B * 2),
```

Where R2:R3:R4 forms the three digits of the output.

- Finally the DISPLAY routine displays the 3-digit calibrated output on the 16*1 LCD display.

4. CONCLUSION:

Being simple it can be a good replacement of a voltmeter if the accuracy and precision requirements are not strict. The use of microcontroller makes the design cost effective and flexible. The basic functionality demonstrated in this article can be easily extended to more advanced forms such as multi-ranging, data logging, etc. The design can be a good guideline for calibrating various physical entities such as temperature, pressure, etc.

REFERENCES:

[1] Roy Choudhury, Linear Integrated Circuits, New Age International P. Ltd., 1998. P.P. 265-273.

[2] K.J.Ayala, The 8051 Microcontroller, Penram Int., 1996.

[3] National Semiconductors. www.national.com.

[4] Atmel Corporation. www.atmel.com

INTERMITTENT FAULTS FOR THE DEPENDABILITY VALIDATION OF COMMERCIAL MICROCONTROLLERS

Dr U.Eranna and Aladalli sharanabsappa,
Designation: Lecturer, Electrical & Electronics Engg.
Ballari Institute of Technology & Management.
sharanueee@gmail.com

ABSTRACT

Nowadays, advances in integration techniques have allowed rising microprocessors operating frequency as well as reducing their size and power voltage, achieving in this way a greater productivity. Nevertheless, these advances have a negative impact on reliability. As the new submicron technologies shrink device sizes, the rate of occurrence of faults increases. The consequences are that designers have to deal with an increasing number of different fault types due to manufacturing defects. Intermittent faults are expected to be a big challenge in modern VLSI circuits. Usually, intermittent faults have been assumed to be the prelude of permanent faults. Currently, intermittent faults due to process variations and residues have grown, being necessary to study their effects. The objective of this work has been to analyze the impact of intermittent faults, taking advantage of the power of the simulation-based fault injection methodology. Using as background faults observed in real computer systems, intermittent faults have been injected in the VHDL model of a microcontroller. The controllability and flexibility of VHDL- based fault injection technique has allowed to do a detailed analysis of the influence of some parameters of intermittent faults.

1. INTRODUCTION

It has been foreseen that intermittent faults will have a great impact in deep submicron technologies. The reduction of the feature size and the power voltage, together with the increase of the clock frequency, will raise the rate of transient faults. On the other hand, the complexity of the manufacturing process (that provokes residues and process variations) and special wear out mechanisms may increase the presence of intermittent faults [1]. Although errors induced by transient and intermittent faults manifest in a similar way, the last ones are activated repeatedly in the same place, and so they are usually grouped in bursts. On the other hand, whereas replacing the affected part eliminates an intermittent fault. Additionally, intermittent faults may be activated or deactivated by temperature, voltage or frequency changes [2]. The knowledge of intermittent faults is not So developed, as fewer observations of real faults and studies of their physical causes and mechanisms have been performed [2] [3]. Related to intermittent faults, there are some Questions difficult to answer: Where do intermittent faults happen? When do faults occur? How many times does a fault activate in a burst? How does the fault manifest at higher abstraction levels? Etc. To answer these questions, it is important to understand the physical mechanisms that take place in deep submicron technologies. But this research subject is complex, it is technology dependent, and it is still in an early phase of evolution. In order to study the impact of intermittent faults, I propose to use a methodology based on fault injection as a complement to the research on the physics of fault. Fault injection technique

allows a controlled introduction of faults in the system, not being necessary to wait for a long

time to log the apparition of real faults [1]. Particularly, VHDL based fault injection can be a very suitable option due to its flexibility as well as the high observability and controllability of all the model components [2]. It allows both to make an exhaustive variation of the fault parameters and to analyze the effects of faults on the system behavior. In these works, the influence of different fault and system parameters has been analyzed. The objective of this work is to deepen those studies, analyzing the method of fault injection using
simulation package for microcontroller circuit such as RAM, ROM and BUS.

A common method to study experimentally the dependability parameters of a microprocessor is Fault Injection [1]. This technique allows a controlled introduction time for logging the apparition of real faults.

1.1 <u>Existing system and its drawbacks</u>

1.1.1 Physical Fault Injection (or Hardware Implemented Fault Injection, HWIFI) Physical fault injection can be used for dependability validation later in the design process when the actual physical system, or a prototype of it, is available. Two main categories of physical fault injection are available. Those that require additional hardware for performing the fault injection, and those that require additional software. Fault injection in physical systems is important because it tests the actual implementation of fault handling mechanisms. However, physical fault injection often provides a more limited controllability and observability than the simulation based techniques; Software implemented Fault Injection (SWIFI)

1.1.2 software implemented fault injection

An emerging trend is to use additional software for injecting faults into physical systems, i.e. software implemented fault injection (SWIFI). The advantages of using SWIFI are cost-effectiveness and flexibility since no extra hardware is required. The technique also allows software defects to be emulated by changing the code. However, as for most physical fault injection techniques, the controllability and observability is somewhat limited. Also, the effects of physical faults may not always be properly emulated since the technique suffers from a lack of physical reachability, i.e. the ability to reach possible fault locations in a system can be inadequate.The various SWIFI
techniques can be divided into pre-runtime injection techniques and Runtime injection techniques depending on whether the faults are injected before the system starts executing the software or during the software execution.

1.2 <u>Proposed System</u>

1.2.1 Simulation Based Fault Injection

The simulation based fault injection technique injects faults into a model of the system, allowing the technique to be applied early in the design process when the actual system is not yet available. Consequently, one of the advantages is early detection of design faults, which thus reduces the cost for correcting such faults. Controllability and observability can also be very high using this technique, while the time overhead involved in simulations is often substantial, which puts practical limitations on the amount of hardware and software activity that can be simulated.

Intermittent faults occur due to unstable or marginal hardware. Manufacturing residues, process variations and special wearout processes can lead to such faults. In addition, wearout mechanisms may provoke that intermittent faults eventually end up in permanent faults.

We have selected a set of intermittent faults observed in real computer systems by means of fault logging [2], as well as fault mechanisms related to process variations and wearout [1][5]. Then, we have deduced a set of fault models at logic (gate) and RTL abstraction levels which can be simulated into VHDL models.

2. <u>Overview of presentation</u>

By understanding the importance of intermittent fault, it is necessary to study the effects of intermittent fault on the microcontroller circuit.

Intermittent fault may occur due to manufacturing defects.

If I consider a simple half adder circuit designed by universal gates (NAND) it is

necessary to understand its possible inputs with respective possible faultless outputs. This can be provided by truth table. if there is a intermittent fault in the circuit of half adder , user won't get desired output It may be short circuited or there may be a open circuit (disconnection of path) such effect of fault is termed technically as stuck at '1' and stuck at '0'. Once we design a circuit it's better to know the effects of such intermittent faults. In other words we can say that effect of faults should be determined before the circuit is in service so that if that fault occurs we can easily identify where and what actually happening in the circuit and can go for remedies. Thus it's necessary to study circuit behavior before we design a fault module for it.

So we need to make changes in the software module according to the possibilities of error occurrence of the circuit.

I have considered a RAM, ROM and BUS of computer at which fault module has been designed and its related wave forms is obtained by the help of simulation software that is modelsim SE 6.2 version. So that we can have a clear picture of stuck at '0' and stuck at '1' fault occurring.

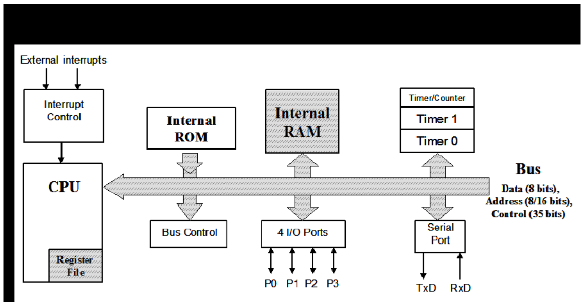
3. **Fault injection experiments**

The different fault injection experiments were carried out on the VHDL model of the 8051 microcontroller [17] running the Bubble sort sorting algorithm as workload.

Although the chosen system has not a deep submicron technology, the methodology used can be generalized to more complex microprocessors/microcontrollers. Fault injection experiments were performed using a fault injection tool called VFIT (VHDL)based Fault Injection Tool) [5], that runs on PC computers (or compatible) under Windows®. VFIT can apply different fault injection techniques on VHDL models (see Figure 1) at diverse abstraction levels. In this work we used the technique based on *simulator commands*, because it is not necessary to modify the VHDL code and it

is easy to apply. This technique consists on using the commands of the simulator to modify the value of the model signals and variables at simulation time. Nevertheless, it has limitations to inject some complex fault models [2].

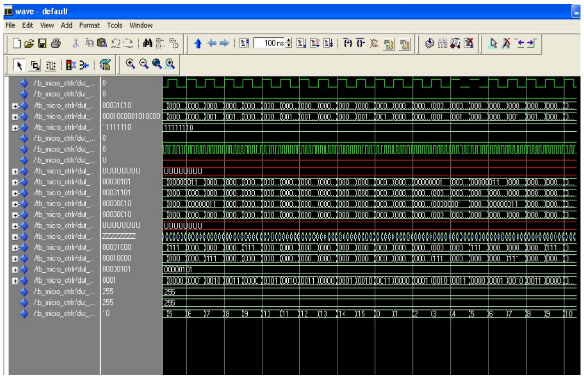
Fault injected in RAM,ROM and BUS as its block diagram is given below



4. **Result**

As fault has been predetermined by the simulation process we can have clear information of fault which may cause in the system which is under service.

Stuck at '0' and stuck at '1' fault will be assumed to be occurred in the system like RAM, ROM and BUS of computer and developed program is injected through modelsim SE simulator and the wave forms of fault and healthy system is observed as follows



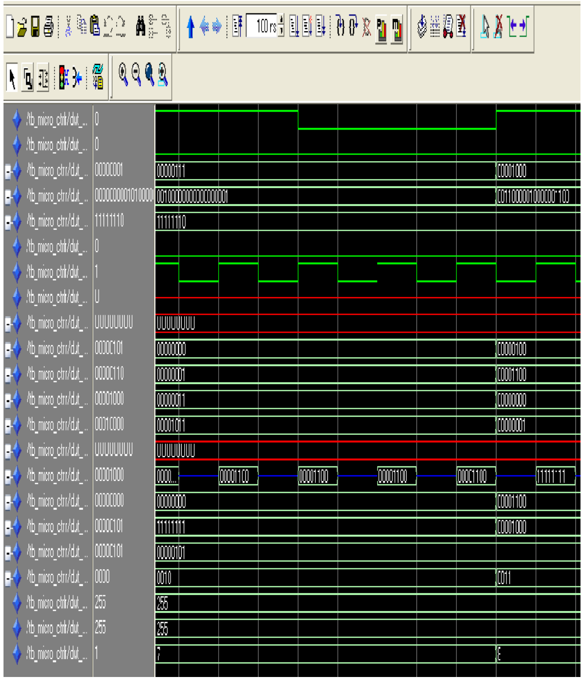

100 ns
/tb_micro_ctrl/dut_... 0
/tb_micro_ctrl/dut_... 0
/tb_micro_ctrl/dut_... 00000001 00000111 00001000
/tb_micro_ctrl/dut_... 0000000001010000 0010000000000000001 0011000001000000100
/tb_micro_ctrl/dut_... 11111110 11111110
/tb_micro_ctrl/dut_... 0
/tb_micro_ctrl/dut_... 1
/tb_micro_ctrl/dut_... U
/tb_micro_ctrl/dut_... UUUUUUUU UUUUUUUU
/tb_micro_ctrl/dut_... 00000101 00000000 00000100
/tb_micro_ctrl/dut_... 00000110 00000001 00001100
/tb_micro_ctrl/dut_... 00001000 00000011 00000000
/tb_micro_ctrl/dut_... 00010000 00001011 00000001
/tb_micro_ctrl/dut_... UUUUUUUU UUUUUUUU
/tb_micro_ctrl/dut_... 00001000 0000... 00001100 00001100 00001100 00001100 11111111
/tb_micro_ctrl/dut_... 00000000 00000000 00001100
/tb_micro_ctrl/dut_... 00000101 11111111 00001000
/tb_micro_ctrl/dut_... 00000101 00000101
/tb_micro_ctrl/dut_... 0000 0010 0011
/tb_micro_ctrl/dut_... 255 255
/tb_micro_ctrl/dut_... 255 255
/tb_micro_ctrl/dut_... 1 7 8

Fig. 1 & 2 Waveforms of stuck at '0' and stuck at '1' fault.

5. Conclusion

VHDL based fault injection provides information of fault occurs in microcontroller which in turn helps to determine the dependability of microcontroller. it is most reliable and controllability process.

6. References

[1] A.Benso and P. Prinetto, eds., "Fault Injection Techniques and Tools.

[2] D. Gil, J. Gracia, J.C. Baraza, and P.J. Gil, "Study, Comparison and Application of different VHDL-Based Fault Injection Techniques for the Experimental Validation of a Fault-TolerantSystem",Microelectronics Journal, vol. 34(1):41-51, 2003.

[3] J.C. Baraza, J. Gracia, D. Gil, and P.J. Gil, "A Prototype of a VHDL Based Fault Injection Tool:Description and Application", Journal of Systems Architecture, vol. 47(10):847-867, 2002.

[4] http://www.oregano.at

[5] Google search

BIOGRAPHY

Louis Amat was born in Toulouse, France:. In 1939. He obtained the Speciality Doctorate degree in 1965 from the University of Toulouse. During 1965-1968, he worked in the Electronical Laboratory of the Thomson-Houston Fi.ench company. He is now working in the Information Processing and Instrumentation Support Service at LAAS-CNRS. In this context. he is responsible for the design and the realization of the MESALINE fault-injector.
Authorized licensed

Cristian Constantinescu
Intel Corporation, RA1-329
2501 NW 229th Avenue
Hillsboro, OR 97124, USA.

THE EMBEDDED MICROCONTROLLER AS THE GREEN TECHNOLOGY IN DISTANT INEXHAUSTIBLE ENERGY SOURCE

Mahesh Choudari . Nithyesh C.L
Dept. of Telecommunication Engg. KIT,Tiptur
Email**:** choudari.86@gmail.com

***Abstract*—With an increased focus on the utilization of green technologies and greater demands on the electric power grid, renewable energy is an important form of current and future power generation. With remote generation deployments, such as those based on wind energy, a cost-effective communication system with global coverage using satellite technology would be advantageous. The monitoring of remote generators for performance and maintenance issues is certainly necessary for any distributed-generation system. To offer a cost-effective satellite solution, a cost-optimization algorithm for minimizing data transmission while maximizing relevant telemetry data is required. This paper proposes a lowcost smart communications architecture using an Iridium Satellite System 9601 short-burst data transceiver and simple microcontroller technology. The microcontroller allows for simple optimization routines to be performed on the locally stored data. This proposed system was implemented and tested and recommendations are drawn on the usability of the developed communication system for monitoring a remote generation site.**

***Index Terms*— Iridium, microcontroller, nonlinear programming, renewable energy, satellite communications, smart systems.**

I. INTRODUCTION

For distributed generation (DG) of electricity, it is important that the generation source, such as those based on renewable technologies, is integrated, monitored, and controlled for possible dispatch. Thus, a communications infrastructure that is cost-effective and reliable [1] is required to monitor and control the remote renewable energy source (RRES). Factors such as cost, telemetry requirements, control functions, and feasibility of various communication technologies determine the specifications of the required DG communication system. Generally, provided that the physical-layer infrastructure exists, communications via a high-speed network access point is desired. However, when not available, alternative communication strategies, such as point-to-point wireless radio modems [1], cellular modems, or satellite modems [2] (or a combination thereof) can be considered.

For low data volume and for applications having a high latency tolerance, this research considers an ISS as a viable alternative for RRES sites. Although data transmission utilizing satellite communications is usually costly in comparison to cellular and unlicensed technologies, smart data transmission, in combination with global coverage capability, supports a unique design solution that is cost-effective and flexible. The Iridium short-burst data (SBD) transceivers were launched, by Iridium Satellite LLC, a privately held company based in Bethesda, MD, mainly for application development in the areas of field-force automation and remote asset tracking [4]. The Iridium constellation consists of 66 low-earth-orbiting (LEO) satellites at about 780 km from the Earth’s surface in six polar orbits. The low-cost ISS 9601 transceiver is designed to support packet-mode SBD service and is designed for remote telemetry applications. It employs a time-domain duplex (TDD) approach, transmitting and receiving in an allotted time window with a 90-ms frame structure in the frequency range of 1616 to 1626.5 MHz. The L-band frequency is used for communication between the ISU and satellite and the Ka-band frequency for intersatellite communication. For telemetry data acquisition (DAQ) of an RRES, previous work by Dahal *et al.* [2] tested the SBD service of Iridium for basic data transmission and assessed the delay aspect of SBD service. It was envisaged that the ISS-based communication system could be used when no other form of communication system exists or as a standby backup system. In addition, this paper focuses on a smart approach to controlling the ISS transceiver for unattended continuous operation by using a programmable interface controller (PIC) microcontroller architecture.

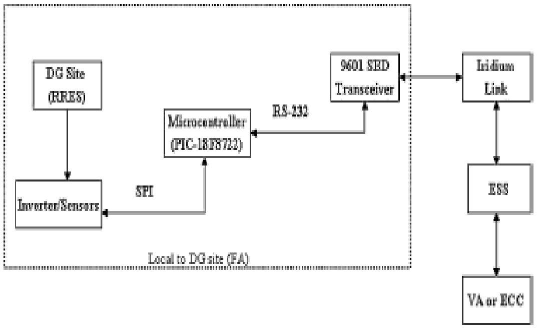
To make the ISS system price viable, a nonlinear constraint algorithm is developed to maximize relevant information content for data transmission. For example, consider a wind generator whose operational parameters include wind speed, temperature, power, and machine maintenance information. System constraints based on these parameters are used to formulate a mathematical model to determine the optimal configuration of the transmitted data, in this case, the ISS SBD data frame. This nonlinear programming algorithm can then be coded into the PIC microcontroller for control and operation of the SBD transceiver. This smart DAQ system could also be used for any other communication system to sense and deliver intelligent data pat-terns to an energy control center (ECC), especially where limiting operational costs are paramount. For example, patterns and content of the data sent could assist the ECC in making certain decisions, regarding preventive maintenance and unscheduled maintenance, which may be of long-term value for an organization. Cost analysis can provide insight to the monitoring requirements and maintenance scheduling, the importance of which will be determined by the specific installation requirements. The various elements of the SBD architecture consist of the remote field application (FA), the iridium subscriber unit (ISU), the Iridium satellite constellation, and the earth terminal controller SBD subsystem (ESS) located at the Iridium gateway, and the vendor application (VA). The ESS is responsible for storing and forwarding messages from the ISU to the VA and storing messages from the VA to forward to the ISU. The ISU communicates with the ESS via the Iridium satellite constellation. The interface between the VA and the ESS uses standard Internet mail protocols to send and receive messages.The ISU supports a maximum of 205 B for a mobile-originated (MO) message and a maximum of 135 B for a mobile-terminated (MT) message in one SBD session. MT messages are sent to the ESS from the VA by using a common e-mail address, identifying the specific ISU by encoding the unique ISU International Mobile Equipment Identity (IMEI) in the subject line of the e-mail. The data message is transmitted as a binary attachment with the e-mail. The MO messages from the FA are delivered to a specific e-mail address that is configured for the particular IMEI. It is only possible to configure one type of delivery system (i.e., either ISU-ISU or ISU-e-mail [5]). The communication system developed in this paper utilizes the latter delivery type. This paper is organized as follows. The basic ISS system and its hardware and software components are presented in Section

II. LOW-COST ARCHITECTURE DESIGN FOR IRIDIUM SBD COMMUNICATION

A. Introduction

As previously discussed, the proposed low-cost communication ISS architecture utilizes a basic 9601 SBD transceiver in combination with a PIC microcontroller (PIC 18F8722) to control the volume of data transfer on a “need to know” basis to the ECC. Some parameters to be monitored at the ECC from RRES sites include: voltage, current, real and reactive power, power quality (PQ), harmonics, transients, flicker and connection/ fault status [3], as well as the maintenance data (MD) related to the field equipment and environmental data (ED). The data collected at the site can thus be categorized as power generation data (PD), MD, and ED. The focus of this research is to formulate the optimal SBD frame, comprised of various numbers of PD, MD, and/or ED subframes to be transmitted to the ECC from the RRES. In terms of ISU communications, a single SBD frame of 205 B fills the MO message to be sent. The optimal configuration of the SBD frame is based on some predefined nonlinear cost function with constraints based on the physical

figure1:ISS/microcontroller architecture block diagram



model. In the end, the goal is to make the SBD service of ISS cost competitive, as the cost per session of SBD or per kilobyte of data via SBD service is relatively high when compared to more ubiquitous wireless technologies. The overall architecture block diagram of the data acquisition (DAQ) and communication system is shown in Fig. 1, where SPI represents the serial peripheral interface. The RRES contains the PD to be transmitted to the ECC and it is assumed that specialized sensors are present for the MD and ED collection. The FA reads PD, MD, and ED from the RRES inverter and the sensors, and subsequently logs the data in the electrically erasable programmable read-only memory (EEPROM) of the microcontroller. The data are clocked from the inverter/sensors to the PIC through the SPI connection at specified times and at specified locations in the EEPROM. The collected data at the end of 24 h will be optimized into a single SBD MO data frame of 205B and sent to the 9601 SBD transceiver through the RS-232 interface for transmission. Once the MO SBD frame is written to the ISS 9601 SBD transceiver, the satellite session is initiated which transfers the message to the ESS. At the ESS, the SBD message is converted into an e-mail message with an attachment and sent to the addresses configured for the IMEI SBD transceiver. For sending the MT message to the FA, an e-mail needs to be composed, from one of the e-mail addresses with a maximum message length of 135 B, with an attachment having an SBD extension and sent to the address sbd@iridium.com with the intended transceivers IMEI number in the subject line of the e-mail. Once queued in the ESS, a confirmation e-mail is sent to the VA and then the FA can at any point in time acquire the message with appropriate operations [6]. As expected, due to the latency of such a system, critical control of the RRES is not possible. However, current design strategies of modern inverter technology generally require local automation intelligence for shutdown and anti-islanding procedures.

B. Field Application Design

The main component of the FA hardware design is the PIC 18F8722 microcontroller chip on a prototyping board driven by a 20-MHz oscillator. The 9601 transceiver is connected via a DB9 RS-232 custom-made serial cable to the PIC 18F8722 for data transfer and to command the transceiver for satellite initiation. The hardware components required for the 9601 SBD transceiver connection to the PIC and antenna are shown in Fig. 2. The transceiver was powered by using a regulated 6-Vdc power supply from the power tray. The power tray used can regulate a 6 V supply output when given an input between 6 V to 30VDC. The power tray also has an RS-232 serial interface, enable/disable switch and status light-emitting diode (LED).The PIC8722 was powerd up with a 5-V dc adapter.The SBD transceiver was connected to a fixed-mast

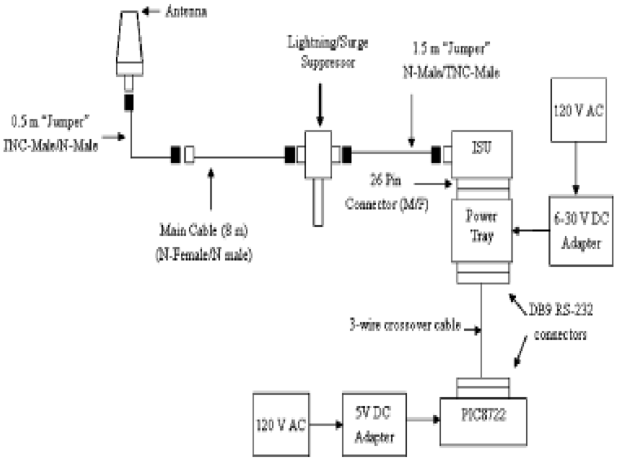


Figure2: Hardware assembly block diagram

helical antenna with an 8-m coaxial cable, flexible jumper cables, connectors, and a lightning/surge suppressor for protection. The insertion loss for the cables, connectors, and lightning arrestor was checked so that it does not exceed 3 dB as per the specifications. The software component of the FA was designed to control the data acquired from the inverter/sensors and more important, the volume of data transfer from the RRES to the ECC. The major design requirement is the cost-effective transmission of the most relevant PD, MD, and ED collected over a given time interval. The various components of the FA can be classified into five logical procedures: 1) acquire data and memory write; 2)optimize the SBD frame using nonlinear programming (NLP);3) control and communicate the MO message, comprised of a single SBD frame, to VA; 4) obtain the MT message at FA; and 5) check the status of the inverter. The code for the FA is written by using the CCS C compiler. The state diagram for the software architecture of the FA is given in Fig. 3, where RI indicates the ring indicator. It uses four of the timer interrupts and one pin interrupt available on the PIC8722 to perform specific tasks at specified times. The appropriate commands (the 9601 SBD transceiver can be controlled with the standard AT command) are written to the transceiver via the RS-232 connection. During the acquire data procedure, PD, MD, and ED subframes are stored over the day using the 1024-B EEPROM available on the PIC microcontroller. Although the DAQ rate and subframe size are design-dependent variables, for the implementation given in this paper, an interrupt (Interrupt 0) is invoked every hour for the writing of PD and ED subframes and on the 6th, 12th, 18th, and 24th h for the MD subframes. Correspondingly, there are 24 subframes of PD and ED collected and stored over a 24-h period stored. Also, there are four subframes of MD collected and stored over a 24-h period stored. Interrupt2 is invoked at the end of 24 h when the data are collected and written to the EEPROM. Once the 24-h data aggregation is complete, a smart algorithm based on NLP calculates the number of PD, MD, and/or ED subframes to optimally fill the SBD frame for the required MO message. Given the 205-B size of the SBD frame, 41 subframes of 5 B each are used to fill a single SBD frame. Also, Interrupt 3 is invoked every minute to check the status of the RRES.This interrupt is implemented so that an emergency condition can be reported to the ECC. Upon receiving a emergency-condition status from the inverter as “true,” the control in FA is passed to send an SBD message to the ECC, with a notification of the inverter status previously written to a predefined EEPROM location. Finally, Interrupt 4 is invoked once per day, at 00:15 h, for the initiation of the satellite session after writing the optimized SBD frame to the 9601 transceiver. The PIC microcontroller then returns to acquire the data for the next day, overwriting the required EEPROM addresses locations for the next 24 h. From a software design and implementation point of view, timing issues,state transitions, subframe sizes, interrupt service routines, and specific memory allocations can be customized for a particular RRES/ECC application. Using the PIC microcontroller allows for this flexibility. For ECC to RRES data transmission, a pin interrupt (Interrupt1) occurs when the RI pin goes high, indicating that there is an MT message waiting at the ESS. The code to retrieve the messages is invoked when RI goes high and the received MT message is written to EEPROM (135 B) for future retrieval and use by the RRES.

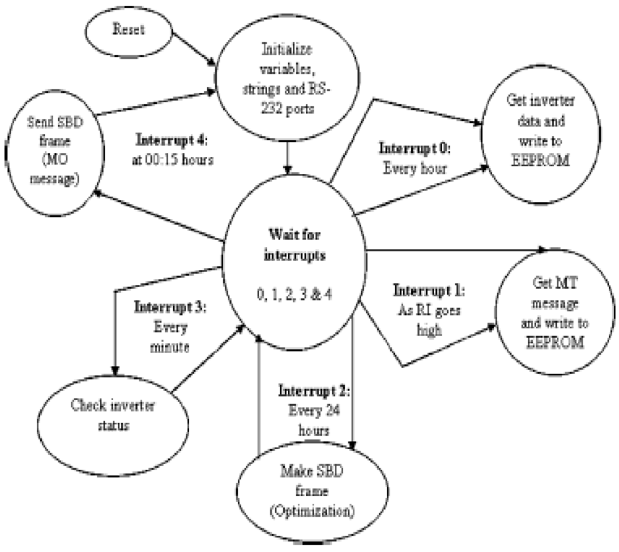

Reset
Initialize variables, strings and RS-232 ports
Send SBD frame (MO message)
Interrupt 4: at 00:15 hours
Get inverter data and write to EEPROM
Interrupt 0: Every hour
Wait for interrupts
0, 1, 2, 3 & 4
Interrupt 1: As RI goes high
Get MT message and write to EEPROM
Interrupt 3: Every minute
Check inverter status
Interrupt 2: Every 24 hours
Make SBD frame (Optimization)

figure3:Field application state diagram

C. Field Test Results

The hardware and software of the FA was tested for MO and MT messages of 205 B and 135 B, respectively. A 205-B MO message containing a single SBD frame was sent from the FA and was received in an e-mail attachment at the VA. The e-mail also has the MO message sequence number (MOMSN) and the MT message sequence number (MTMSN) giving the number of MO and MT messages sent and received by the particular IMEI. The time stamps of the sessions are in coordinated universal time (UTC) format where the time for a particular time zone can be expressed as positive or negative offsets from the UTC.The confirmation message also gives the message size, latitude, and longitude position of the ISU. The geolocation information is only an approximate location of the ISU and if not required in the application, could be disabled. For testing the MT message, an e-mail with the attachment content of 135 B of one's was sent to the ESS from the VA. The confirmation message of its storage in the ESS for polling by the ISU with 135 B of attachment size was received. As expected, the message was retrieved at the FA without errors. Additional information, such as the count of MT messages queued at the ESS to be polled by the IMEI, is also provided. The total delay depends on the transmission and propagation delays and the performance of the routing algorithm implemented in the satellites. The various components of the delays associated with the SBD services can be classified into five categories, that is: 1) modem processing time or transmission delay (depending on message size); 2) uplink delay; 3) intersatellite handoff and queuing delays; 4) downlink delay; and 5)E-mail delays. The modem processing delay for a 205-B message using an average throughput of 2.4 kb/s is calculated to be683 ms. Using the signal travel speed of 3 10 m/s and distance to the satellite and back as approximately 1600 km, the uplink and downlink delay works out to be 5.2 ms. The general queuing algorithms in the satellite nodes suggest that the maximum delays associated with queuing would be on the order of a few hundred microseconds [7]. Assuming a queuing delay per node of 300 s and four intersatellite handoffs, the total expected queuing delay is estimated to be 1.2 ms. Given that a typical distance between satellites is 400 km, the total intersatellite handoff delay is approximately 5.3 ms. Combining all delays (i.e. 683ms+5.2ms+5.3ms+1.2ms) yields 695 ms of theoretical satellite delay. Clearly, the modem-processing time is the dominant component of the total satellite delay. The timing stamps, when deciphered from the e-mail received from the field test, showed a time of 18 s for an MO message sent from FA to reach the VA. This corroborates well with the timing from the MO message study performed by Margaret [8].

Since the total satellite delay is estimated at 695 ms, about 99% of the total message delay time can be attributed to the e-mail delay (i.e., from the Iridium gateway to the VA). As the Iridium gateway processes messages once every 30 s [8], to collect the MT SBD message queued for a particular IMEI, the latency for the MT message to reach the ISU could be as high as 30 s. This high latency precludes the ISS solution from safety-critical requirements but should be satisfactory for other telemetry applications.

III. *SMART SYSTEM DESIGN ALGORITHM —A WIND POWER EXAMPLE*

A. Background

DG formulated by using wind-based RRESs has achieved a significant place in the green energy market [9]. Effective operation of these wind turbines, with a capital and operational cost of approximately U.S.$2000 per kilowatt, warrants the acquisition of telemetry data for power output, maintenance data, and specific environmental conditions. With a single SBD frame sent every 24 h, the main goal of the smart algorithm implemented on the PIC microcontroller is to optimally configure the SBD frame based on some predefined objectives. Correspondingly, an objective function based on specific decision variables and bounded by system constraints is required to formulate the respective mathematical model. For the decision variables, and are defined as the number of 5-B subframes of PD, MD, and ED, respectively. Using NLP techniques, an optimal solution yields the SBD frame configuration of, and subframes so as to maximize the objective function.

B. Physical Model

The growth of wind power energy systems for green power has also caused the appearance of many reliability issues mainly related to maintenance practices [9]. Therefore, acquiring system data on changing and unpredictable environmental conditions will validate foreseen or respond to unforeseen maintenance requirements. For example, increased lubricant temperature due to higher-than-expected wind speeds or external temperatures may accelerate the machine bearing wear. This can certainly improve the maintenance practices and reduce the reliability problems currently found with wind generators. Given a maintenance data focus, it is proposed that a power, maintenance, and environmental (PME) data constraint is utilized to find an optimal SBD frame with the

number of MD subframes being dependent on PD and ED data patterns recorded. In formulating the PME constraint, three regions of operation (PME1, PME2, and PME3) of the system are considered to decide on the weight of the MD data on the SBD frame configuration. The three regions are defined as:

1) PME1: power output or turbine speed over the day is lower than the rated values for a long time;

2) PME2: power output or turbine speed over a day generally runs on a rated level and the ambient temperature around normal conditions;

3) PME3: power output or turbine speed over the day is higher than the optimal operational values or if the ambient temperature or humidity data collected over the day is higher than nominal.

Fig. 4 shows the modeling of the physical behavior constraint of a wind turbine DG and explains the regions of operation of the system. Depending on the change of PD given by the power output (in kilowatts) or MD given by turbine speed (revolutions per minute) over the day1 and ambient temperature and humiditychanges2, it has been decided that subframes , and will fill the MO SBD message. The system should be sensitive to reporting more subframes in the MO message on a more regular basis if the system has operated in region PME3, as this could lead to a temperature increase of the lubricant leading to possible micro pitting of the gears. If the system has operated in region PME1 over the day, the subframes should again be reported relatively higher in the MO SBD message as this may not be a suitable environment for the high-performance self-lubricating bearings to function at their maximum efficiency. Also, if the system operated in region PME2 over the day, the MO message could have less N_{mi} subframes since the data required for the maintenance are presumed to be relatively low for such a situation.

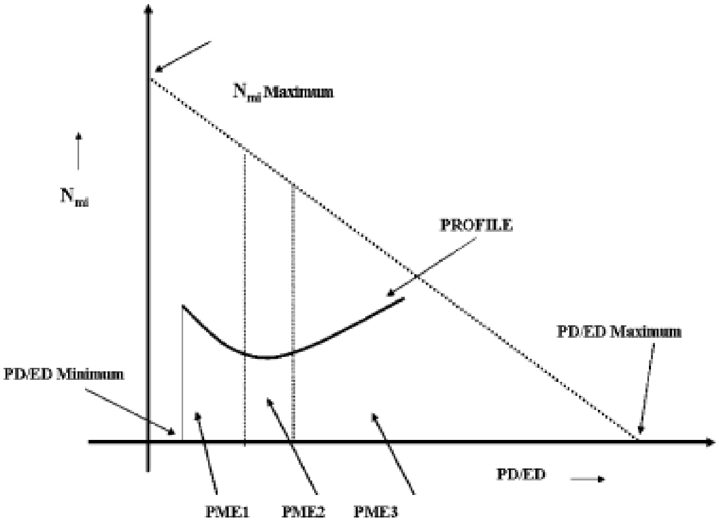


Figure 4: wind turbines region of operation

1Comment: signifying values of PD being higher/lower/rated that is indirectlyresponsible for increased/decreased temperature of the machine lubricant by friction.

2Comment: signifying values of ED which have an effect on the maintenance parameters of the lubricant.

C. Mathematical Constraint Modelling

For the system constraints, the profile for deciding the N_{pi}, N_{mi}, and N_{ei}subframes from the values of PD and ED collected over a day is conceptualized as a parabola shown in Fig. 4. A description of how the values of PD and ED collected over the day influence the number of subframes (N_{pi}, N_{mi}, and N_{ei}) contending to fill the SBD frame will be described mathematically. The subframe variables applied to the general equation of a parabola give the PME constraint inequality as

$$N_{mi} \geq K_1 \left(N_{pi} - \frac{K_2 - N_{ei}}{K_4}\right)^2 + K_3 \qquad (1)$$

where , K_1, K_2, K_3 and K_4 are the quadratic onstant, scaled offset, offset, and offset scalar, respectively, which are chosen so that the profile of the curve represents reasonable N_{pi}, N_{mi}, and N_{ei} for various conditions of PD and ED values collected over the day. Also, to make the system responsive and sensitive to the number of subframes of MD, the PME constraints could be automatically changed over a year, which is considered to be the regular maintenance period. For example, a larger N_{mi} could be used for the six months after the regular maintenance period. For the next six months, data could then be reduced, since the ECC has the confidence from the MD from the first six months, N_{mi} and that a regular maintenance appointment would be approaching soon. Equations shown in (2) and (3) are sample PME constraints for the first and the second half of the regular maintenance period.On substituting example K values into inequality (1), the two typical constraints are

$$N_{mi} \geq \frac{10}{20^2} \left(N_{pi} - \frac{80 - N_{ei}}{4}\right)^2 + 10 \qquad (2)$$

$$N_{mi} \geq \frac{10}{35^2} \left(N_{pi} - \frac{140 - N_{ei}}{4}\right)^2 + 2. \qquad (3)$$

Inequality (2) is designed for the first half of the maintenance period after the regular maintenance period where monitoring of the RRES's MD occurs more frequently. Inequality (3) is for the second half of the maintenance period where relatively less MD is required by the ECC. In each case, the offset reflects the minimum value.

The data byte size constraint on the system is that the number of subframes of the decision variables cannot exceed 41 since a subframe is designed to be 5 B long with an SBD size frame of 205 B. Therefore, inequality (4) expresses the data byte size constraint mathematically:

$$N_{pi} + N_{mi} + N_{ei} \leq 41. \qquad (4)$$

Also, the value of N_{pi}, N_{mi}, and N_{ei} should be integers greater than or equal to zero. Correspondingly, inequality (5) gives the other constraint mathematically as:

$$\begin{aligned} N_{pi} &\geq 0 \\ N_{mi} &\geq 0 \\ N_{ei} &\geq 0. \end{aligned} \qquad (5)$$

These constraints plotted together (inequalities (2) and (4) and inequality (5)), for $N_{ei} = 1$ and 41, give the solution points $N_{pi}, N_{mi}, N_{ei})$ under the area carved by the parabola and the triangular plane as illustrated in Fig. 4. The optimal solution can be found by using NLP with the given decision variables and constraints with a goal to communicate the most-needed data to the ECC. With regard to the optimal solution of the SBD frame configuration, a set of points solution $(N_{pi}, N_{mi}$ and $N_{ei})$ satisfying the constraints, is fed to the objective function to calculate which of the solutions yields the maximum value. A simple objective function is given by;

$$R = A\ N_{pi} + B\ N_{mi} + C\ N_{ei} \qquad (6)$$

where the coefficients A, B, and C are chosen to provide appropriate weights for the number of subframes (N_{pi}, N_{mi}, and N_{ei}), and indicate which data have more significance to the ECC3 .Also, to make the system adaptive over the summer and winter months of the year, the coefficients A, B, and C of the objective function can be varied to give appropriate weights to the subframes depending on the conditions on the field. During the summer months when the temperature and humidity may increase, the constant A could be reduced while increasing B and C so that more weight is on the N_{mi} and N_{ei} frames. On the contrary, during winter months, the objective function could be changed to reflect more PD as the environmental conditions are assumed to be more suitable for the rated operation of the RRES wind generators and turbines. As an example, consider the objective function weights $A = 0.5, B = 0.3$, and $C = 0.2$, which yields:

$$R = 0.5N_{pi} + 0.3N_{mi} + 0.2N_{ei}. \qquad (7)$$

The optimal solution of the NLP problem is the combination of N_{pi}, N_{mi} and N_{ei} that yields the largest value of R.

3The sum of the weights A, B, and C is limited to unity.

IV. CONCLUSION

The SBD service of Iridium was chosen as a viable platform for the DG communication system due to its global service availability as well as its low hardware and setup cost. A cost-effective interrupt-driven microcontroller architecture with the 9601 SBD Iridium transceiver was developed and tested for MO and MT messages and the timing, and delay analysis of the Iridium SBD service was analyzed. With the delay of MT/MO messages, it is seen that critical control functions, such as anti-islanding of the inverter, cannot be achieved by this system but could be used for noncritical telemetry data acquisition or for a backup system in case of a communications fault. In terms of system latency, the MO and MT message delays were mostly due to the processes at the gateway and the e-mail delays. This paper also presented the idea of optimizing the 205 B of the SBD frame using NLP for smart data transmission on a "need to know" basis at the ECC. This is required given the data-transmission cost of the Iridium system. The feasibility of a PIC microcontroller- based DAQ was investigated and was found to be capable of the design requirements of the proposed smart DAQ system. Finally, compared to the ISU-email approach, this research could be taken further by testing the system with the relatively fast ISU-ISU SBD service.

REFERENCES

[1] J. Meng and S. Barnes, "Distributed generation communications utilizing a real-time FPGA-based channel simulator in the ISM frequency band," *Int. J. Emerging Elect. Power Syst.*, vol. 8, no. 1, p. 3, 2007.

[2] U. D. Dahal, B. R. Petersen, and J. Meng, "Iridium communication system for data telemetry of renewable distributed generation system," in *Proc. 24th Biennial Symp. Communications*, Kingston, ON, Canada, Jun. 24–26, 2008, pp. 262–265.

[3] L. Chang, "Wind energy conversion systems," *IEEE Canadian Rev.*,no. 40, pp. 12–16, 2002.

[4] R. Hobby,An Introduction to the Iridium System. London, U.K., Oct.1998, [Electronic version]. Inst. Elect. Eng. Colloq. Commun. Offered by Advanced Satellite Systems.

[5] *"Iridium SBD Service Developers Guide,"* 1.2 ed. Feb. 2006.

[6] *"IRIDIUM 9601 SBD Transceiver Product Developers Guide (RevisionEd.),"* 1.24 ed.Dec. 2005.

[7] A. M. Jabbar, "Multi-Link Iridium satellite data communication system," M.Sc. dissertation, Univ. Kansas, Lawrence, 2001.

[8] M. Margaret, R. Rathbarn, Ed., "Measuring latency in iridium satellite constellation data services," Naval Academy, A291464. Annapolis, MD, Jun. 2005.

[9] J. Terradillos, M. Bilbao, J. I. Ciria, A. Malagan, and J. Ameya, "Oil analysis as an improvement tool for the behavior of windmill gears. Main problems detected through the lubricant condition" Prague, Czech Republic, 2006, Eur. Lubricating Gears Inst.

NAME: ***2)Mahesh Choudari, 9632921056***
choudari .86@gmail.com
Kalpataru Institute of Technology
Tiptur,Tumkur

An Expert Controller for a DC Motor Speed Control System

Nagabhushana Katte[¥], K. Nagabhushan Raju, P. Bhaskar*, and Parvathi C. S.*

[¥]*Department of ECE, Ballari Institute of Technology & Management, Bellary – 583 104, KA., INDIA.*

Department of Instrumentation, Sri Krishnadevaraya University, Anantapur – 515 003, A.P., INDIA.

**Department of Instrumentation Technology, Gulbarga University P. G. Centre, Raichur - 584 133, Karnataka, INDIA.*

nagkatte@gmail.com

Abstract–Paper emphasizes on design and development of fuzzy logic based an Expert Controller (EC) and its implementation for a real time DC motor speed control application. Fuzzification, inference mechanism, and defuzzification processes of the expert controller are characterized by triangular shaped membership functions, IF and THEN rules, and centre of gravity method respectively. Speed measurement of DC motor with 16-bits precision is achieved through the optical encoder, frequency to voltage converter and A/D converter and control actions given to the power actuator via D/A converter (16-bits) and PWM generator. Experimental results of the system are presented for step input, step and set point variations. Control algorithms are implemented using C language.

Index Terms–Expert controller, Fuzzy logic, DC motor, Speed control, and Computer based control.

I. INTRODUCTION

DC motors are used worldwide in many residential, commercial, industrial and utility applications. In many applications, DC motor speed control plays a key role in many processes; in addition, precision and quality in control of speed (with minimum overshoots and undershoots, fast rise and settling times) is desirable. Conventional control strategies usually require a mathematical model for designing the controller [1-3]. The inaccuracy of mathematical modeling of the plants usually degrades the performance of the controller, especially for nonlinear and complex control problems. Recently, the advent of the fuzzy logic implemented controllers (expert controllers) has inspired new resources for the possible realization of better and more efficient control [4-6]. They offer a key advantage over traditional control systems. That is, they do not require mathematical models of the plants and hence it is concluded that fuzzy logic is dominant for process control applications [7-9]. This gave the authors motivation for design and development of computer based expert controller for a DC motor speed control system and to study its performance under no-load and on-load conditions.

II. INSTRUMENTATION

The instrumentation system for measurement and control of speed of a DC motor is shown in Fig. 1. Specifications of the motor used in current investigation are presented in table 1. The personal computer plays a key role in acquiring the speed, computing the error and change in errors, evaluating the control actions through fuzzy implemented control algorithms and sending control actions to the motor system.

Description	Value
Rated voltage	± 12 V DC
Rated current	200 mA, at no load 290 mA, at full load
Maximum speed	3500 RPM, at no load 2400 RPM, at full load
Torque	50 gm-cm
Weight	150 gm

TABLE I. Specifications of the DC motor

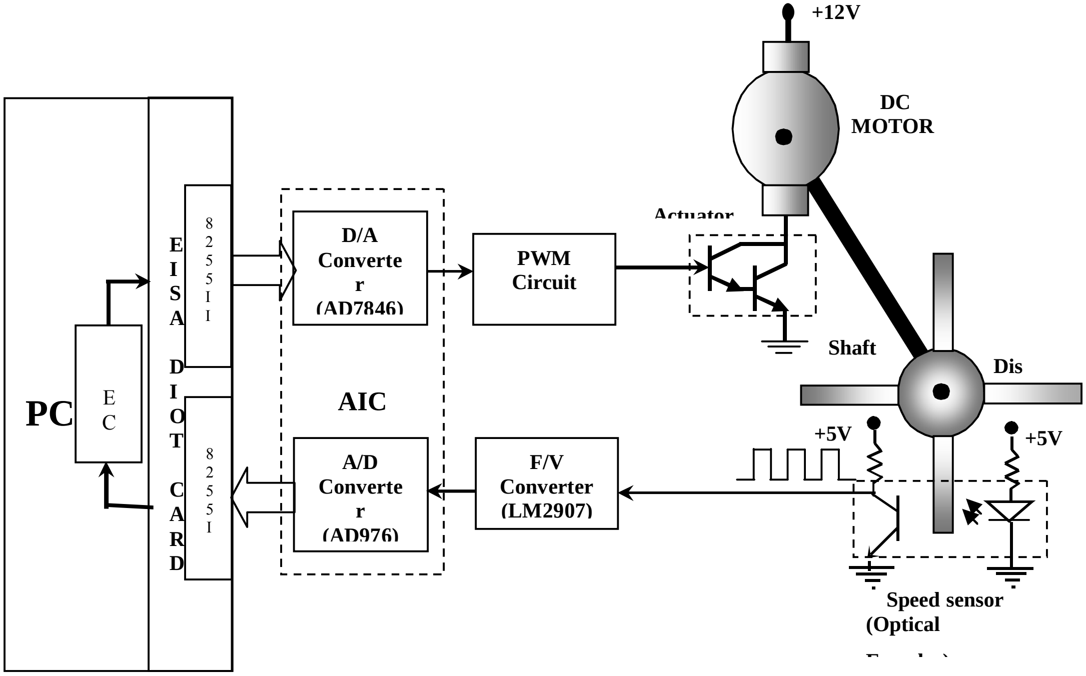


Fig. 1. Block diagram of the computer based DC motor speed control

1) Speed measurement

The optical encoder senses the speed of the motor and converts it into a train of TTL compatible pulses. Frequency of these pulses is directly proportional to the speed of the motor. This frequency is converted into proportional voltage by F/V converter. Computer acquires voltage through A/D converter available on analog interface card (AIC). This voltage in digital form is converted back to corresponding frequency by the equation $f = a_1*v + a_0$, where f is the frequency of the signal generated from optical encoder, v is the measured voltage of F/V converter, a_1 = slope of frequency v/s voltage graph, & a_0 = intercept on y-axis. The values of a_1 and a_0 are found from the plot of frequency v/s voltage as 80.156 and

0.04276 respectively. Hence the equation is given by:

$$f = 80.156 * v + 0.04276$$

Further this frequency is converted into speed in RPM by the equation:

$$\text{Speed} = (f * 60 \text{ seconds}) * (1/p) \text{ RPM}$$
$$= (f * 5) \text{ RPM}$$

where, p = number of pulses for one revolution. For the optical encoder used, 12 pulses are generated for one complete revolution.

2) Speed control

Speed control of the motor is achieved as follows; the error and change in errors are computed and applied to the PID and fuzzy expert control algorithms. The controller produces the control action according to the error. The computer then applies this control action, in the form of digital data, to the motor through D/A converter AD7846, pulse width modulator and actuator. The ON time of PWM wave varies with digital data. If digital data is more, ON time will be more and vice-versa. Hence, the power applied to motor through actuator will vary with PWM wave. This procedure is repeated till the motor reaches the desired speed. Thus the motor speed is controlled at the desired value.

III. EXPERT CONTROLLER

Methodology of design of EC using Fuzzy logic technique has been discussed. Two input linguistic variables of EC are given as, error e(k) = (set-point speed – measured speed)/set-point speed and change-in-error ce(k) = present error – previous error. These two inputs are defined on a universe of discourse with the seven membership functions (NL, NM, NS, ZE, PS, PM, and PL) as shown in Fig. 2.

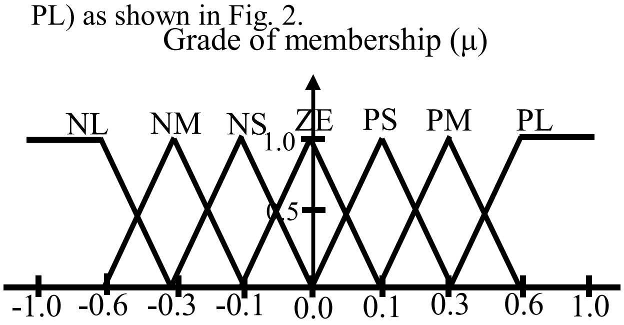
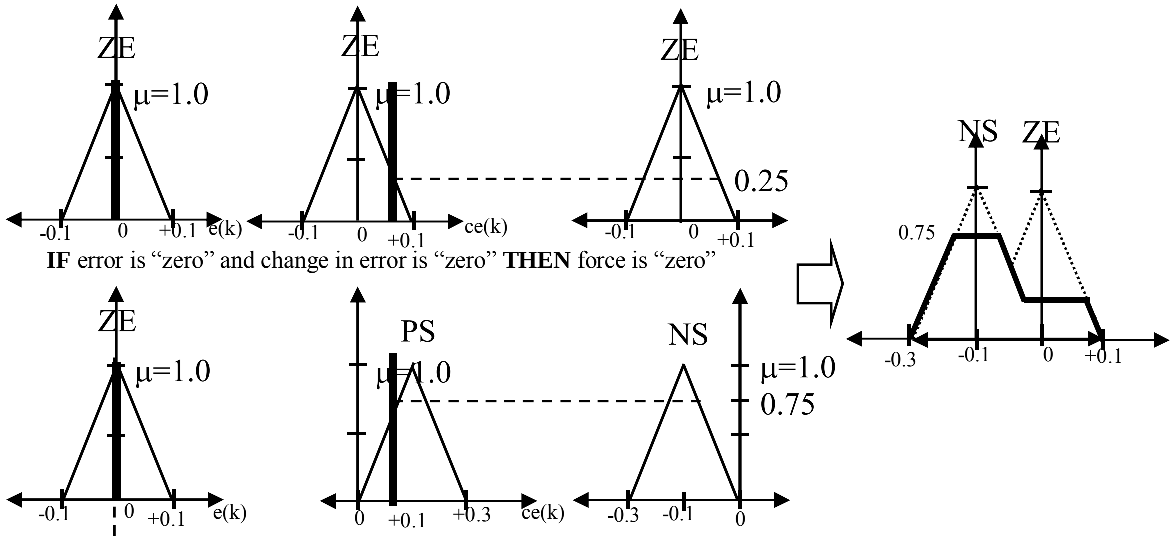


Fig. 2. Triangular membership functions

Fuzzy inference engine emulates the expert's decision making in interpreting and applying knowledge about how[7]to do good control [10,11]. It picks up (makes a decision) a control rule using IF and THEN statements from its rule base. Fig. 3 illustrates the possible decision-making process. Fuzzy evaluated rules represented by the implied fuzzy sets are combined using centre of gravity (COG) method of defuzzification and is given as [12],

$$cu(k) = \left(\sum_{i=1}^{i=n} \mu_{cu}(w_i).w_i\right) / \left(\sum \mu_{cu}(w_i)\right)$$

where, w_i is the support member value for the ith element, and $\mu_{cu}(w_i)$ is the value of grade of membership function for ith element.



IF error is "zero" and change in error is "zero" **THEN** force is "zero"

IF error is "zero" and change in error is "positive small" **THEN** force is "negative small"

Fig. 3. Fuzzy decision-making

As an example, the above illustrations are combined and given as,

cu(k) = ((.75)*(-.1) + (.25)*(0))/(.75+.25)
= -0.075

The fuzzy velocity control algorithm computes the control action applied to the process and is given as,

$$u = cu(k) + cu(k\text{-}1)$$

where, u is the final control value, cu(k) is present fuzzy computed control action and cu(k-1) is the previous control action.

IV. SOFTWARE DEVELOPMENT

Software of computer based EC for a DC motor speed control system is written in 'C' language. The software is based on the character user interface (CUI), it displays a text menu, which enables the user to monitor current value of the parameter (speed) being measured and controlled, and the controller parameters. It also features to enter the new set point; filename to store measured data; and tune the controller parameters etc. The details of the fuzzy software are presented in the flowchart shown in Fig. 4 (a) and (b).

V. EXPERIMENTAL VERIFICATION

Computer based expert speed controller has been designed and investigated for a DC motor control system. The efficacy of this controller for speed control of a DC motor is evaluated by applying several tests over a wide range of operating conditions. The performance indexes (in terms settling time, steady state error, overshoot, and under shoot) of the proposed controller are studied under no-load condition. Photograph of working model of the experimental

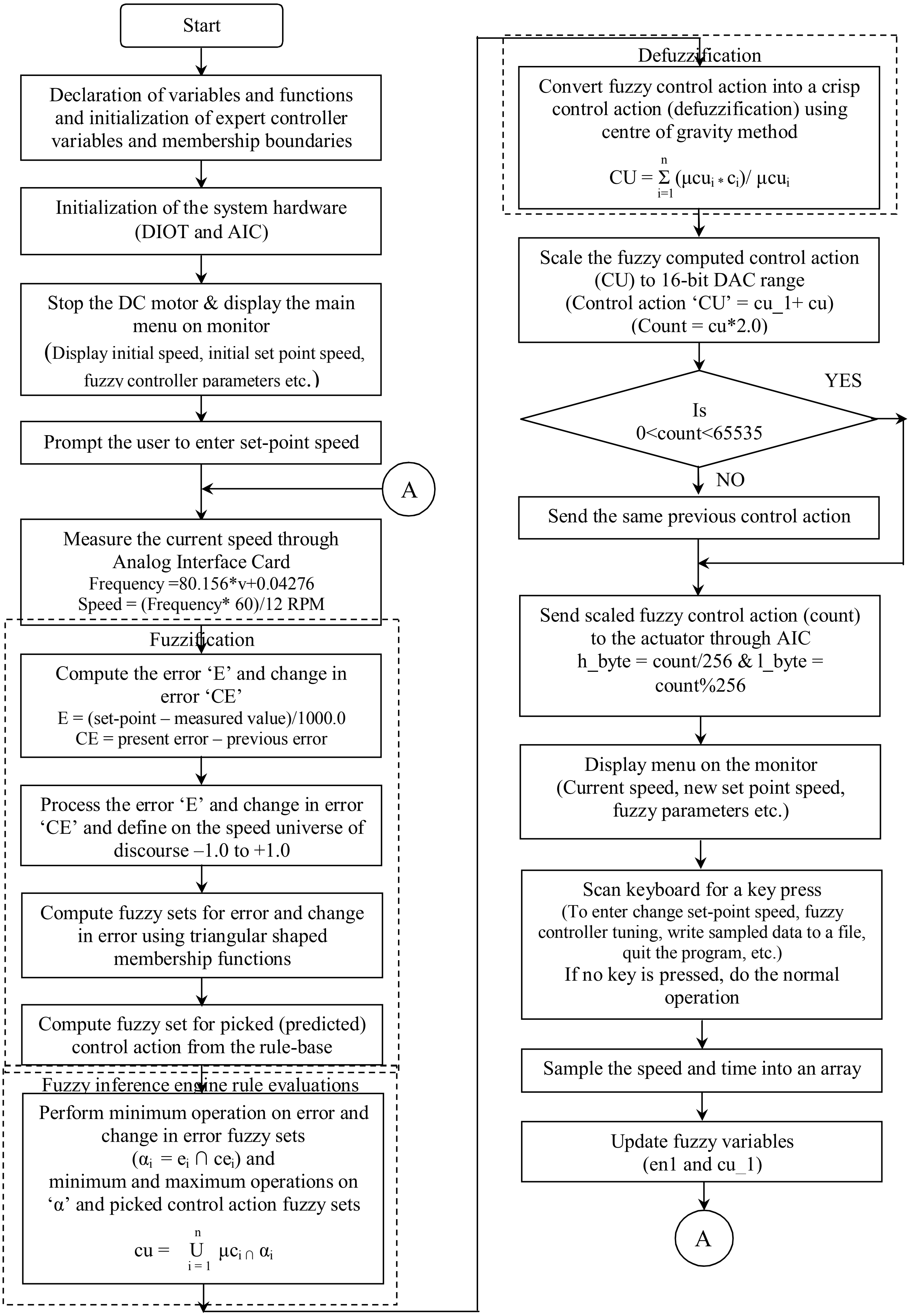

Start
Declaration of variables and functions and initialization of expert controller variables and membership boundaries
Initialization of the system hardware (DIOT and AIC)
Stop the DC motor & display the main menu on monitor
(Display initial speed, initial set point speed, fuzzy controller parameters etc.)
Prompt the user to enter set-point speed
A
Measure the current speed through Analog Interface Card
Frequency =80.156*v+0.04276
Speed = (Frequency* 60)/12 RPM
Fuzzification
Compute the error 'E' and change in error 'CE'
E = (set-point – measured value)/1000.0
CE = present error – previous error
Process the error 'E' and change in error 'CE' and define on the speed universe of discourse –1.0 to +1.0
Compute fuzzy sets for error and change in error using triangular shaped membership functions
Compute fuzzy set for picked (predicted) control action from the rule-base
Fuzzy inference engine rule evaluations
Perform minimum operation on error and change in error fuzzy sets
(α_i = $e_i \cap ce_i$) and
minimum and maximum operations on 'α' and picked control action fuzzy sets
cu = $\bigcup_{i=1}^{n} \mu c_i \cap \alpha_i$
Defuzzification
Convert fuzzy control action into a crisp control action (defuzzification) using centre of gravity method
CU = $\sum_{i=1}^{n} (\mu cu_i * c_i) / \mu cu_i$
Scale the fuzzy computed control action (CU) to 16-bit DAC range
(Control action 'CU' = cu_1+ cu)
(Count = cu*2.0)
Is 0<count<65535
YES
NO
Send the same previous control action
Send scaled fuzzy control action (count) to the actuator through AIC
h_byte = count/256 & l_byte = count%256
Display menu on the monitor
(Current speed, new set point speed, fuzzy parameters etc.)
Scan keyboard for a key press
(To enter change set-point speed, fuzzy controller tuning, write sampled data to a file, quit the program, etc.)
If no key is pressed, do the normal operation
Sample the speed and time into an array
Update fuzzy variables (en1 and cu_1)
A

Fig. 4 Flowchart for expert controller implementation

system is presented in Fig. 5. Fig. 6 shows the comparison of step input responses of PID and EC and Table 2 the corresponding numerical results of the computer based DC motor speed control system. The transient response of EC is better in terms of reduced overshoot and better steady state response with reduced steady state error and settling time. So it is concluded that the proposed controller has better time response when compared to PID. The performance of the proposed controller is also studied for set point and step variations. Fig. 7 shows the response of EC for set-point variations and step variations. The EC responds quicker when step is changed and adopts well for different set point variation. The numerical results in this paper confirm the validity of the proposed speed control of DC motor. The results have demonstrated that acceptable control performance can be achieved using triangular shaped seven membership EC. The results of the experiments on the dc motor speed control demonstrate that the performance of the EC is better than that of a commercial PID controller.

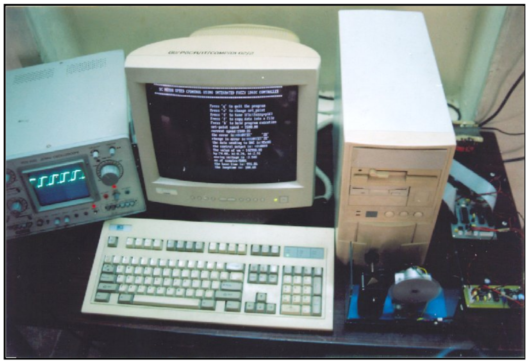


Fig. 5. Photograph of the experimental system

TABLE II. Numerical results of the computer based DC motor speed control system

DC MOTOR SPEED CONTROL SYSTEM RESULTS For a step size of 1000 RPM (0-1000RPM)						
Controller	Sampling Interval (Seconds)	Maximum		Settling time (Seconds)	Steady-state Error (RPM)	Remarks
		Overshoot (RPM)	Undershoot (RPM)			
PID	0.01879	0.95	0.63	5.5616	0.52	
EC	0.01971	4.58	0.69	3.6717	0.38	Better

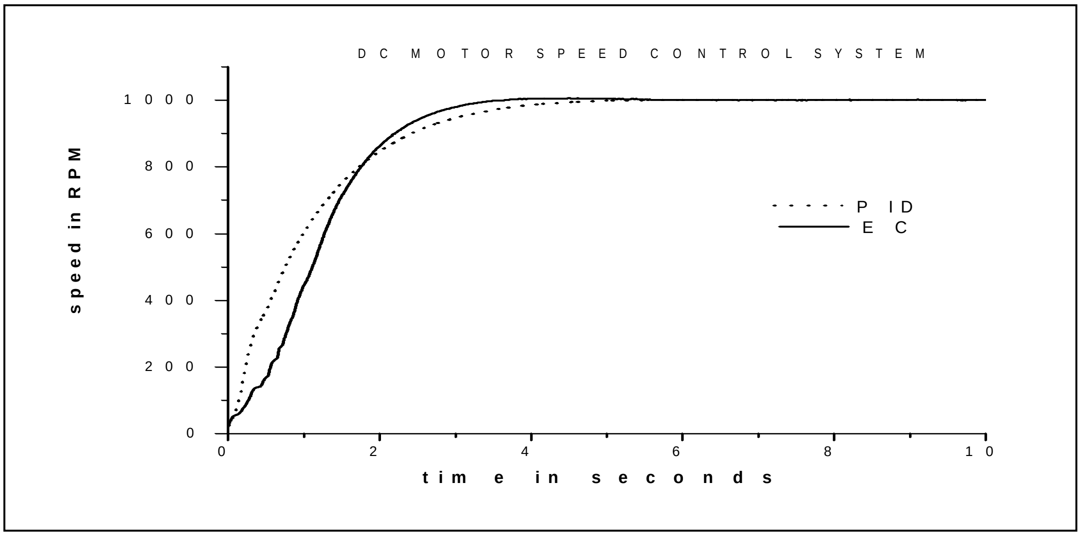
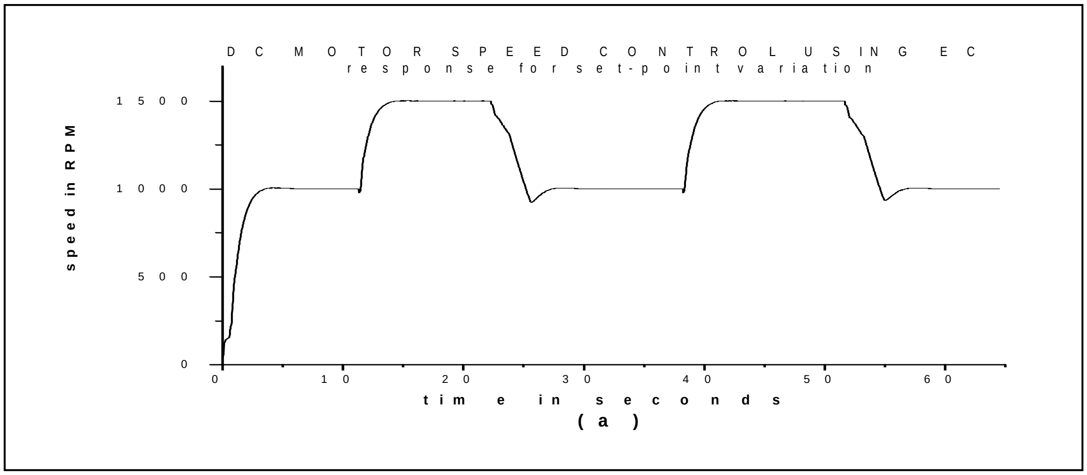


Fig. 6. Comparison of step input responses of PID & EC



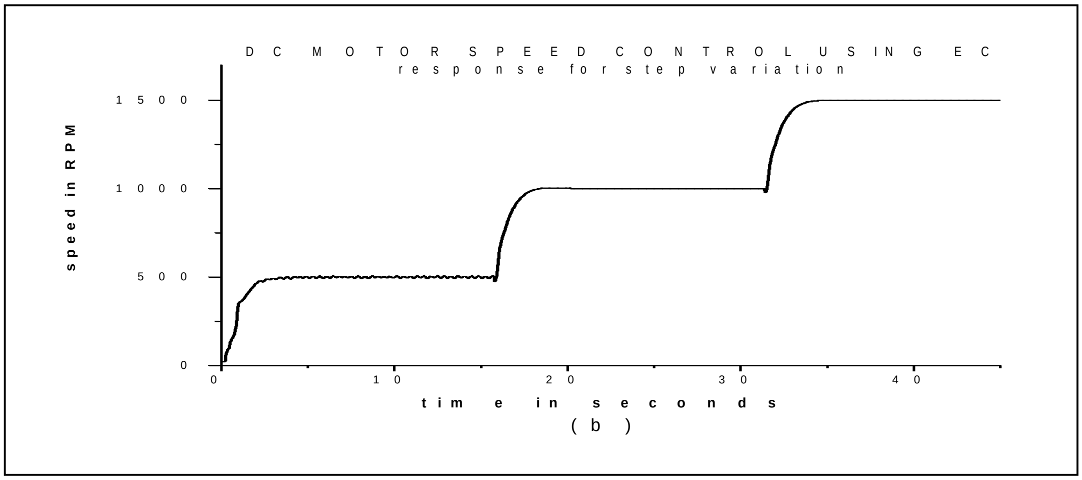


Fig. 7. Response of EC for (a) set-point variations (b) step variations

REFERENCES

[1] B. G. Liptak, *Instrument Engineers' Handbook – Process Control*, Butterworth Heinemann Ltd., Oxford, 1995.

[2] Paul-Hai Lin, SantaiHwang and John Chou, "Comparison of fuzzy logic and PID controls for a DC motor position controller," *IEEE Industrial Applications Society Annual Meeting,,* vol. 3,pp. 1930-1935, 1994.

[3] J. Nagrath and M. Gopal, *Control Systems Engineering,* 3rd ed., New Age International (P) Ltd., New Delhi, 2002.

[4] P. Guillemin, "Fuzzy logic applied to motor control," *IEEE Trans. Ind. Applications,* vol. 32, pp. 51-56, 1996.

[5] W. W. Tan, A. L. Dexter, " A self learning fuzzy controller for embedded applications," *Automatica,* vol. 36, pp. 1189-1198, 2000.

[6] Bao-Gang Hu, George K. I. Mann, and Raymond G. Gosine, "A systematic study of fuzzy PID controllers-function based evaluation approach," *IEEE Transactions on Fuzzy Systems,* vol. 9, 2001.

[7] Kickert W. J. M. and Lemke Van Nauta H. R., "Application of fuzzy logic controller in a warm water plant," *Automatica,* vol.12, pp. 301-308, 1976.

[8] Mamdani E. H. and Assilian S., "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man Machine Studies*, vol. 7, pp. 1-13, 1975.

[9] Y. Tipsuwan and M. Y. Chow, "Fuzzy logic microcontroller implementation for dc motor speed control," *IEEE IECON'99,* vol. 3, pp. 1271-1276, 1999.

[10] Chun Chen Lee, "Fuzzy logic in control systems: Fuzzy logic controller – Part I, II," *IEEE Transactions on systems, man and cybernetics,* vol. 20, no.2, March/April 1990.

[11] Li Zheng, "A practical guide to tune of PI like fuzzy controllers," *IEEE Int. Conf. On Fuzzy System,* pp.633-640, 1992.

[12] Hung T, Nguyen and Derek Sands, "Real-time self-organizing fuzzy logic controller for DC servo," Proc. of the European Power Electronics Association, pp.174-179, 1993.

Successful weaning from respiratory support system with Neural technology and Evolutionary algorithm

Name of the guide: Prof V. C. Patil
Name: S. Firdosh Parveen
Designation: Student
College: BITM, Bellary
Phone No: 9945797763, 9243206220
Email id: firdoseks@gmail.com

Abstract

Patients require mechanical ventilator support when the ventilator or gas exchange capabilities of their own respiratory system fail. The process of discontinuation of Mechanical ventilation from the patients is called Weaning. Unnecessary delays in discontinuation can lead to a host of complication and even death. The proposed method is to achieve successful weaning using genetic algorithm. Performance comparison of genetic algorithm with gradient descent is carried, where genetic algorithm gives much accurate results compared with gradient descent. The objective of the research is to provide very high performance in real time operation, which can improve current diagnosis very much.

KEYWORDS: Artificial Neural Networks, Genetic Algorithm, Gradient Descent, Mechanical Ventilation, Successful Weaning.

I INTRODUCTION

Mechanical ventilation is a method to mechanically assist or replace spontaneous breathing. Mechanical ventilation is typically used after an invasive intubation, a procedure wherein an endotracheal or tracheostomy tube is inserted into the airway. It is used in acute settings such as in the ICU for a short period of time during a serious illness. It may be used at home or in a nursing or rehabilitation institution if patients have chronic illnesses that require long-term ventilation assistance [1].

Mechanical ventilation is often a life-saving intervention, but carries many potential complications including pneumothorax, airway injury, alveolar damage, and ventilator-associated pneumonia. Mechanical ventilation, although the mainstay of treatment for respiratory distress syndrome, can result in physical trauma to lung tissue. Respiratory failure requiring mechanical ventilation has a large impact on hospital economics and resources.

The timing and method of discontinuation from mechanical ventilation remains an important clinical problem. Mechanical ventilation can result in life-threatening complications and therefore should be discontinued as soon as possible. However, premature attempts at weaning from respiratory support can result in failure and reinstitution of mechanical ventilation, which carries an enhanced risk of morbidity and mortality. Therefore, it is no surprise that many different strategies for successful weaning have been described in the medical literature.

All patients should be evaluated daily to determine if they are candidates for discontinuation of mechanical ventilation. To be considered a candidate, a patient should meet four criteria: (1) evidence of reversal or stability of the cause of acute respiratory failure; (2) adequate oxygenation as indicated by $Pao_2/Fio_2 > 150$ to 200, PEEP(Positive End Expiratory Pressure) in the range of ≤ 5 to 8 cm H_2O, $Fio_2 \leq 0.4$ to 0.5, and $pH > 7.25$; (3) hemodynamic stability, as defined by the absence of active myocardial ischemia and the absence of clinically significant hypotension and (4) ability to make an inspiratory effort[2].

A. Artificial Neural Networks

1. Architecture: The design and implementation of intelligent system with human capabilities is the starting point to design Artificial Neural Networks

(ANN). Artificial neural networks are computational systems whose architecture and operation are inspired from the knowledge about biological neural cells (neurons) in the brain [3].

ANNs is a network of many simple processors called units, linked to certain neighbors with varying coefficients of connectivity called weights that represent the strength of these connections. The basic unit of ANNs called an artificial neuron, simulates the basic functions of natural neurons. It receives inputs process them by simple connections and threshold operations and outputs a result.

ANN's can be divided into feedforward and recurrent classes according to their connectivity. An ANN is feed forward if their exists a method which numbers all the nodes in the network such that there is no connection from a node with a large number to a node with a smaller number. All the connections are from the node with a small number to the node with a larger number. An ANN is recurrent if such a numbering method does not exists.

Figure.1 shows the schematic representation of a multilayer perceptron with eight input neurons, two hidden layers with eight hidden neurons and one output layer with single neuron. Each of the input neuron connects to each of the hidden neurons, and each of the hidden neurons connects to the output neurons

The architecture of an ANN is determined by its topological structure i.e., the overall connectivity and transfer function of each node in the network.

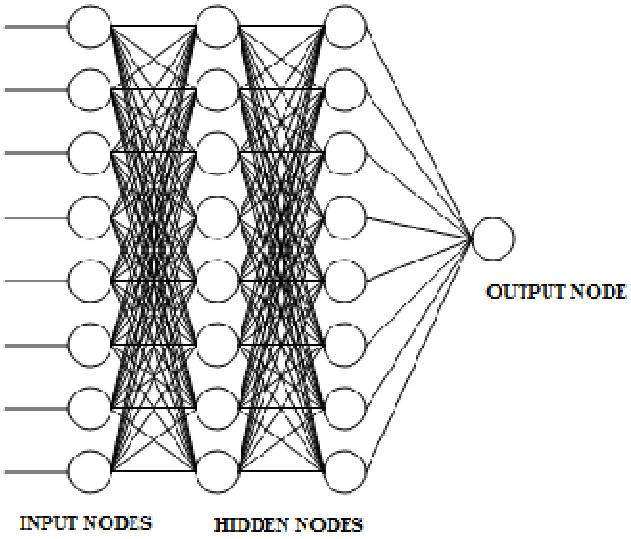


Fig.1 Schematic Representation of a Multi Layer Perceptron

2. Learning in ANN: Learning in ANN's is accomplished using examples. This is also called "training" in ANN's because the learning is achieved by adjusting the connection weight in ANN's iteratively so that trained (or learned) ANN's can perform certain tasks. Learning in ANN's can roughly be divided into supervised, unsupervised and reinforcement learning. Supervised learning is based on direct comparison between the actual output of an ANN and desired correct output also known as the target output. It is often formulated as the minimization of an error function such as the total mean square error between the actual output and the desired output summed over all available data. A gradient descent- based optimization algorithm such as back propagation (BP)[4] can then be used to adjust connection weights in the ANN iteratively in order to minimize the error. Reinforcement learning is a special case of a supervised learning where the exact desired the output is unknown. It is based only on the information of whether or not the actual output is correct. Unsupervised learning is solely base don the correlations among input data. No information on the "correct output" is available for learning.

The essence of a learning algorithm is the learning rule, i.e., a weight updating rule which determines how connection weights are changed[5].

B. Evolutionary algorithm

Evolutionary algorithms are stochastic optimization algorithms based on the mechanism of natural selection and natural genetics [6].They perform parallel search in complex search spaces. Evolutionary algorithms include genetic algorithms, evolution strategies and evolutionary programming. We deal with genetic algorithms in this paper. Genetic algorithms (GA's) were originally proposed by Holland [7].GA's have been applied to many function optimization problems and are shown to be good in finding optimal and near optimal solutions. Their robustness of search in large search spaces and their domain independent nature motivated their applications in various fields like pattern recognition, machine learning, VLSI design, etc.

C. Problem statement

Improper weaning may lead to very serious consequences, such as prolonged ventilation time, pneumonia and even death. Therefore, every effort must be extended to alleviate the problems mentioned. This paper will demonstrate which patients can be successfully weaned. The objective of the project is to

1. Develop a Dynamic Expert system for weaning the mechanical ventilation with very high accuracy and precision.
2. Evolutionary computation (Genetic algorithm) will be involved with neural technology for further performance enhancement.
3. To develop a very robust design using distributed architecture.
4. Performance comparison of genetic algorithm with Gradient descent algorithm.
5. Graphical and numerical methods for performance and comparative analysis.
6. To create low cost, high efficient health care environment in ICU.

This paper is organized as follows: section 2 discuses the related work, section 3 describes the methodology used, section 4 describes the experimental results, and section 5 concludes the paper.

II Related work

Two large (and related) evidence-based projects were conducted to review the results of published studies and evaluate the different strategies for successful weaning from mechanical ventilation. The first of these was commissioned by the Agency for Healthcare Policy and Research, who asked the McMaster University Evidence-Based Practice center to evaluate the issues surrounding ventilation, weaning and discontinuation. This group was directed to address five specific questions: (1) When should weaning be initiated? (2) What criteria should be used to determine when to begin the weaning process? (3) What are the most effective methods of weaning? (4) What are the optimal roles for nonphysician health-care providers in the weaning process? and (5) What is the value of using clinical practice algorithms or protocols in the weaning process? The McMaster project reviewed > 5,000 reports and identified 154 publications, which they used to create their comprehensive review. The second group was a task force put together by the American College of Chest Physicians, the Society for Critical Care Medicine, and the American Association for Respiratory Care. This group was charged to create clinical practice guidelines based on the McMaster report and their own literature review and expert consensus opinion [2][8].

III METHODOLOGY

The performance of the proposed method is demonstrated by employing the genetic algorithm. The performance criteria used in this research is to achieve the successful weaning of the patients. This paper compares the results of genetic algorithm and gradient descent.

The data is collected from the Royal infirmary of Edinburgh intensive care unit, Scotland. The data include the patient's respiratory parameters like Tidal Volume (VT), Respiratory Pressure (RR), Minimum Ventilation (VE) and Negative Inspiratory Pressure (NIF) are available at every instant of time and based on these values the patient shall be examined for weaning decision.

A. Gradient descent

Gradient descent is a function optimization method which uses the derivative of the function and the idea of steepest descent. The derivative of a function is simply the slope. So if we know the slope of a function, then it stands to reason that all we have to do is somehow move the function in the negative direction of the slope and that will reduce the value of the function. Gradient descent is an iterative method, so the idea is as follows:

- Compute the derivative of the function with respect to its independent variables. We can denote this derivative as dF(x), where F(x)) is the function to be minimized, and x is the vector of independent variables.
- Change the value of x as follows:
  xn+1 = xn – h dF(xn),
  where the subscript n refers to the iteration number, and h is a step size which must be chosen so that we don't take too big or too small of a step. Too big of a step will overshoot the function minimum, and too small of a step will result in a long convergence time.
- Repeat the above two steps until we converge to a minimum of the function F(x).

Gradient descent is an attractive optimization method in that it is conceptually straightforward and often converges quickly. Its drawbacks include the fact that the derivative of the function must be available and it converges to a local minimum rather than global minimum. So in this proposed work training is given using this method, but when tested the expected output is different from the output given by the machine

B .Genetic Algorithm

GA has blossomed rapidly due to the easy availability of low cost but fast speed small computers. The complex and conflicting problems that required simultaneous solutions, which in past were considered deadlocked problems, can now be obtained with GA. However, the GA is not considered a mathematically guided algorithm. The optima obtained are evolved from generation to generation without stringent mathematical formulation such as the traditional gradient–type of optimizing procedure. Infact; GA is much different in that context. It is merely a stochastic, discrete event and a non linear process. The obtained optima are an end product containing the best elements of previous generations where the attributes of a stronger individual tend to be carried forward into the following generation. The rule of the game is "survival of the fittest will win" [9].

A simple genetic algorithm can be summed up in seven steps as follows [10]:

1. Start with a randomly generated population of n chromosomes.
2. Calculate fitness of each chromosome.
3. Select a pair of parent chromosomes from the initial population.
4. With a probability *Pcross* (the 'crossover probability' of the 'crossover rate'), perform crossover to produce two offspring.
5. Mutate the two offspring with a probability *Pmut* (the mutation probability).
6. Replace the offspring in the population.
7. Check for termination or go to step 2.

Each iteration of the above steps is called a generation. The termination condition is usually a fixed number of generations typically anywhere from 50 to 500 or more. Under certain other circumstances, a check for global minimum is done after each generation and the algorithm is terminated as and when it is reached [11]. The chief aspect of this method is the representation of the parameter as strings of binary digits of 0 and 1. This composition allows simple crossover and mutation functions that can operate on the chromosomes.

C. Comparison between Genetic algorithm based training and Gradient based training

The genetic algorithm based training approach is attractive because it can handle the global search problem in a vast, complex, multimodal and non differentiable surface. It does not depend on the gradient information of the error function and thus is particularly appealing when this information is unavailable or very costly to obtain or estimate.

Genetic algorithm is less sensitive to initial condition of training. They always search for a globally optimal solution, while a gradient descent algorithm can only find a local optimum in a neighbourhood of the initial solution.

Genetic algorithm based training algorithm is significantly faster than methods that use the GDR. For the three tests in his paper [12], the Genetic algorithm based training algorithm "took a total of about 3 hours and 20 minutes, and the GDR took a total of about 23 hours and 40 minutes".

IV EXPERIMENTAL RESULTS

A computer simulation has been developed to achieve successful weaning using genetic algorithm. The simulations have been carried out using MATLAB. Various networks were developed and tested with random initial weights. The data is collected from the intensive care unit of royal infirmary of Edinburgh, Scotland. When NN are trained using genetic algorithm and later tested gave the correct results for which patient to be weaned or not where as the gradient descent method does not give the correct result. The results of genetic algorithm and gradient descent are shown in the graph below.

Graph for Gradient descent

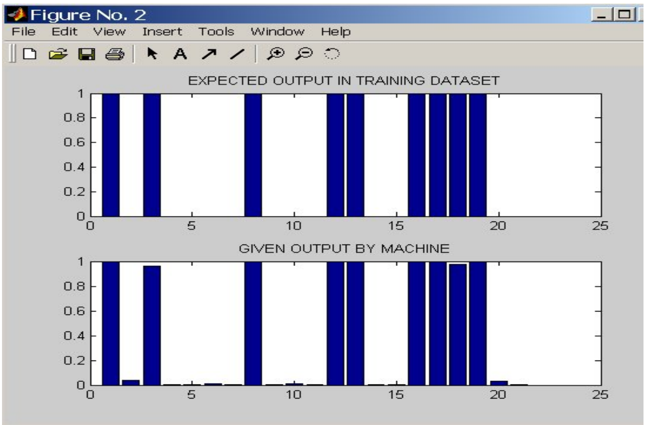


Fig 2 Snapshot showing the expected output in training dataset and output given by the machine

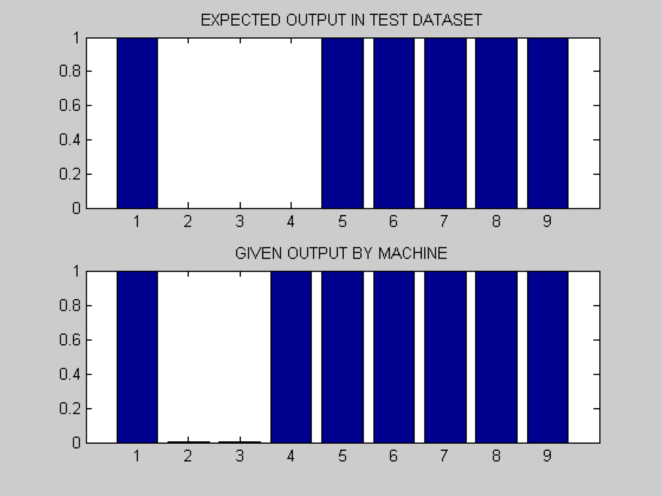


Fig 3 Snapshot showing the expected output in test dataset and output given by the machine

Graphs for Genetic algorithm

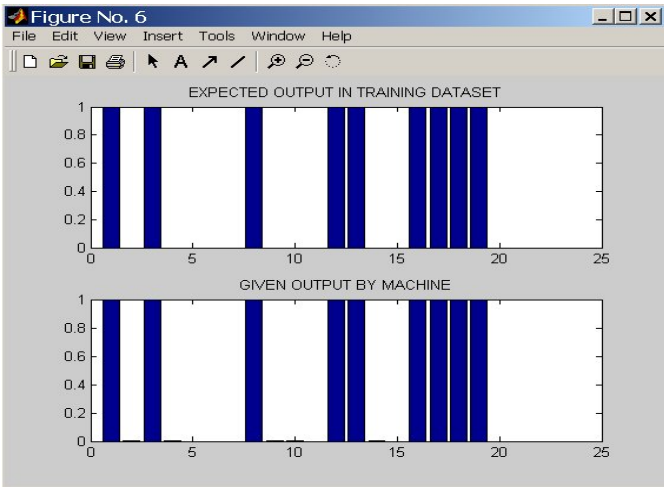


Fig 4 Snapshot showing the expected output in training dataset and output given by the machine

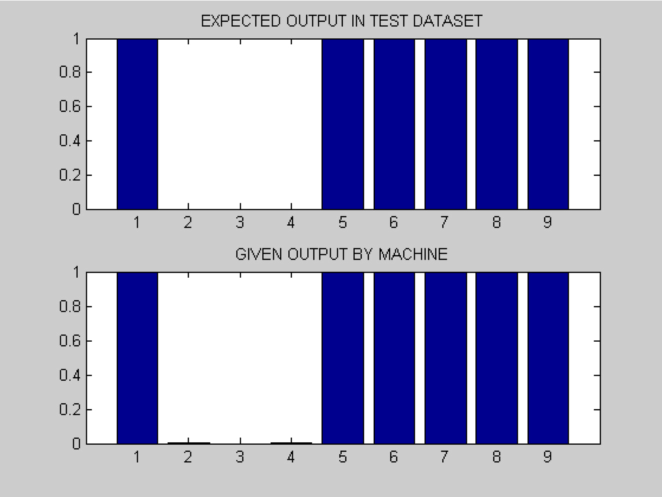


Fig 5 Snapshot showing the expected output in test dataset and output given by the machine

V CONCLUSION

Mechanical ventilation can have life treating complications it should be discontinued at the earliest possible time. the process of discontinuing mechanical ventilation termed" weaning" is one of the most challenging problems in intensive care and it account for a considerable proportion of the work load of staff in an ICU. The discontinuation of mechanical ventilation needs to be carefully timed premature discontinuation place severe stress on the respiratory and cardiovascular system which can impede the patients recovery. Unnecessary delay in discontinuation can lead to a host of complication.

Clinicians are generally inaccurate in weaning decision because there is no accurate predefined rule. Accuracy is only on average around 60% with trail and error method.

This paper demonstrates the successful weaning of the patients from mechanical ventilation with Neural Networks and evolutionary computation i.e., genetic algorithm.

VI REFERENCES

1. Collice, Gene L "Historical prespective on the development of mechanical ventilation". in Martin J Tobin. Principles and practice of Mechanical ventilation 2nd edition. Newyork: Mc graw-hill.
2. Neil R. Mac intyre MD Current Issues in Mechanical Ventilation For Respiratory Failure, 01 Nov 2005.
3. Madiha J. Jafri, Vince D. Calhoun (2006), Functional Classification of Schizophrenia Using Feed Forward Neural Networks, Proceedings of the 28th IEEE EMBS Annual International Conference, pp.6631-6634.
4. G.E.Hinton, "Connectionist learning procedures," Artificial Intel., vol. 40, no. 1–3, pp. 185–234, Sept. 1989.
5. J. Hertz, A. Krogh, and R. Palmer, Introduction to the Theory of Neural Computation. Reading, MA: Addison-Wesley, 1991.
6. D. B. Fogel, "An introduction to simulated evolutionary optimization", *IEEE Trans. Neural Networks,* vol. 5, no. 1, pp. 3–14, 1994.
7. J. H. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: Univ. of Michigan Press, 1975.
8. Chairman – Neil R. Mac Intyre M.D, F.C.C.P A collective task force facilitated by the American Collage of Chest, The American Association for respiratory care; and The American collage of critical care medicine, December, 2001.
9. D. E. Goldberg, *"Genetic Algorithms in Search, Optimization and Machine Learning"*, Reading, MA: Addison-Wesley, 1989.
10. Marco Russo, "Fu Ge Ne Sys – A Fuzzy Genetic Neural System for Fuzzy Modeling", *IEEE Transactions on Fuzzy Systems*, vol6, no,3,August 1998, pp 373 – 387.
11. Melanie Mitchell, *"An Introduction to Genetic Algorithms"*, A Bradford Book MIT Press, 1997.
12. D. L. Prados, "Training multilayered neural networks by replac-ing the least fit hidden neurons," in Proc. IEEE SOUTHEAST-CON'92, vol. 2, pp. 634–637.
13. Gilat, Amos, Matlab: An introduction with applications, 2nd edition,2004. John Wiley and sons.

Hardware Trojan

[1]Divya Rani P, [2]Veena, [3]Shama R B, [4]Shrusti H M, [5]R.Manjula
E-mail: [1]pdivyarani@gmail.com, [2]veena.minajagi@gmail.com, [3]shm542@gmail.com, [5]manjularaja5@rediffmail.com
5th semester, ECE,
Ballari Institute of Technology & Management
Bellary, Karnataka, INDIA

***Abstract*: Electronic devices such as mobile phones, laptops, music players have become an integral part of our daily life. Do we ever consider the possibility that these could be used to compromise our privacy and leak our deepest digital secrets? The migration of IC fabrication to low-cost foundries has made ICs vulnerable to malicious alterations that could, under specific conditions, result in functional changes and/or catastrophic failure of the system in which they are embedded. We refer to such malicious alternations and inclusions as Hardware Trojans. The modification(s) introduced by the Trojan depends on the application, with some designed to disable the system or degrade signal integrity, while others are designed to defeat hardware security and encryption to leak plain text information. This paper explores the wide range of malicious alternations of ICs that are possible and proposes a general framework for their classification. The taxonomy is essential for properly evaluating the effectiveness of methods designed to detect Trojans, several Trojan detection strategies and the classes of Trojans each is most likely to detect.**

Keywords- Hardware Trojan, *loose distribution, tight distribution Always-on, Condition-based, Failure Analysis-based Techniques, Side Channel Signals Analysis, ATPG-based Trojan Detection Techniques, Trojan Detection Challenges.*

1. Introduction

Electronics plays an important role in:

✴ Storage and communication of confidential information

✴ Management and control of important equipment

Today's business is global and high-security applications such as banking or government systems, military, finance, power or the political sector are facilitated by the integrated off-the-shelf silicon devices. Their security often relies on hardware based security modules. The majority of current PCs and laptops are sold with built-in trusted platform module (TPM) chips. Since these personal devices can be physically accessed by their owners, their security often relies on hardware based security modules. Security modules implemented in silicon must be more trustworthy. The fabrication of integrated circuits manufactured in untrustworthy factories cannot be ignored. The hardware integrity, i.e. a chip has no modifications in comparison with the original chip design, is not ensured. An adversary can introduce a Trojan designed to disable and/or destroy a system at some future time (we call it Time Bomb) or the Trojan may serve to leak confidential information covertly to the adversary. Trojans can be implemented as hardware modifications to application specific ICs (ASICs), commercial off the shelf (COTS) components, microprocessors, or digital signal processors (DSPs), or as firmware modifications, e.g., to field programmable gate arrays (FPGA) bit streams. Unfortunately, the detection of such inclusions is difficult for several reasons: 1) Nanometer IC feature sizes and system complexity make detection through physical inspection and destructive reverse engineering difficult and costly. Moreover, destructive reverse engineering does not guarantee that ICs not destructively inspected are Trojan-free. 2) Trojan circuits are by design activated under very specific conditions, which makes it difficult to activate and detect those using random stimuli. Moreover, existing automatic test pattern generation (ATPG) methods used in manufacturing test for detecting defects do so by operating on the netlist of the Trojan-free circuit specification. Therefore, existing ATPG algorithms cannot target Trojan activation/detection directly. In order to develop methods designed to improve IC TRUST, it is essential to first define taxonomy for Trojans. The Trojan classification is base on the several fundamental characteristics of Trojans, including their physical, activation and action characteristics. Once a framework is established, we consider detection strategies and the metrics on which they can be evaluated, such as the complexity of the method and the amount of effort needed to establish trust.

Definition: Hardware Trojan is malicious alteration of hardware that could, under specific conditions, result in functional changes of the system. It is completely characterized by its physical representation and its behavior.

2. Taxonomy Of Trojan

Malicious alternations to the structure and function of a chip can take many forms. We decompose the Trojan taxonomy into three principle categories as shown in Figure 1, i.e., according to their physical, activation and action characteristics. The physical characteristics of a Trojan are further partitioned into four categories; type, size, distribution, and structure. Our proposed taxonomy, therefore, describes Trojans using six attributes, including four physical, one activation and one action attribute. Although it is possible for Trojans to be hybrids of this classification, e.g., have more than one activation characteristic, we believe this taxonomy captures the elemental characteristics of Trojans and will be useful for defining the capabilities of various detection strategies.

2.1. Trojan Physical Characteristics

The physical characteristics category describes the various hardware manifestations of Trojans (see Figure 2).

A. Type: The *type* category partitions Trojans into functional and parametric classes. The functional class includes Trojans that are physically realized through the addition or deletion of transistors or gates, while parametric refers to Trojans that are realized through modifications of existing wires and logic. The parametric Trojan modifies the original circuitry, e.g. thinning of wires, weakening of flip-flops or transistors, subjecting the chip to radiation, or using Focused Ion-Beams (FIB) to reduce the reliability of a chip.

B. Size: The *size* category accounts for the number of components in the chip that have been added, deleted or compromised. Size of a Trojan can be an important factor during activation. A smaller Trojan has a higher probability for activation than a Trojan with larger number of inputs.

C. Distribution: The *distribution* category describes the location of the Trojan in the physical layout of the chip. For example, a *tight distribution* describes a Trojan whose components are topologically close in the layout while a *loose distribution* describes Trojans that are dispersed across the layout of the chip. Figure 2 shows some examples. Note that the distribution of Trojans depends on the availability of dead spaces on the layout. If very small dead spaces are available on the layout, then the adversary may be forced to place and route smaller portions of the Trojan in different dead spaces. Note that here we assume that the adversary may not change the physical layout dimension of the design.

D. Structure: If the adversary is forced to regenerate the layout to be able to insert the Trojan, then the chip dimensions change. This change could result in different placement for some or all the design components. Any changes in physical layout can change the delay and power characteristics of chip which will make it easier to detect the Trojan. In order to minimize the probability of detection, an adversary is likely to adopt a strategy whereby the physical ‘footprint’ of the Trojan is as small as possible. We use the term *stealthy physical footprint* to describe the adversary’s objective in this regard. For small, tightly coupled parametric Trojans, the goal is easily achieved because parametric Trojans can be introduced by changing the geometry of a single wire or transistor. For functional Trojans, size and distribution have significant impact on the physical footprint of the Trojan. For larger sizes, distributing the Trojan across the layout can improve the stealthy physical footprint criteria because detecting the Trojan based on, for example, an anomaly in a localized power or leakage signature, is more difficult. However, distributing the Trojan across the layout can actually worsen its physical footprint in other respects. For example, the length of the wires connecting the Trojan increases significantly, which changes the capacitance distribution of the Trojan-free chip and increases the chances that the Trojan will change the delay characteristics of the chip. For this reason, tightly coupled Trojans may be more attractive, particularly if power/leakage hiding techniques, such as power gating through transistor stacks, are used to reduce its footprint. Figure 3 shows six examples for Trojans with different physical characteristics. We will investigate this further in the context of detection strategies in Section 4.

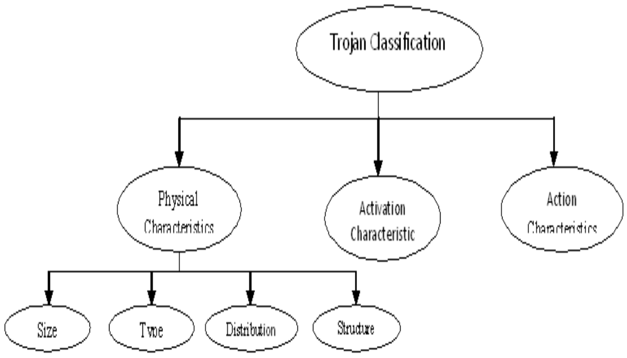


Figure1. Taxonomy of Trojan

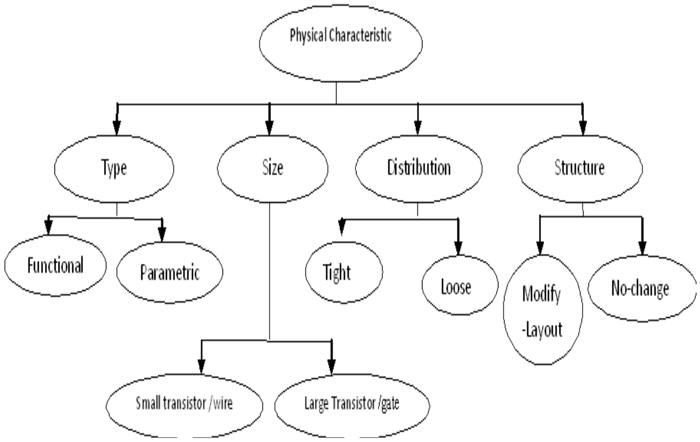
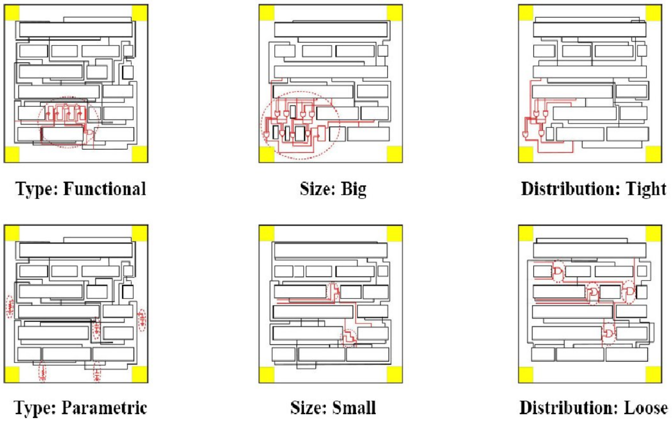


Figure2. Physical Characteristics of Trojan



Type: Functional **Size: Big** **Distribution: Tight**

Type: Parametric **Size: Small** **Distribution: Loose**

Figure 3. Examples for various Trojan physical characteristics

2.2. Trojan Activation Characteristics

Activation characteristics refer to the criteria that cause the Trojan to become active and carry out its disruptive function. The adversary who inserted the Trojan will make it difficult for the user of the chip to activate it, in an effort to prevent 'accidental' activation and detection during the testing phase(s) of the chip and system. Therefore, activation of a Trojan can be considered a 'rare event' from a statistical perspective. We use the term *stealthy activation* to describe the adversary's objective in this regard. We partition Trojan activation characteristics into two sub-categories, labeled *Externally-activated* and *Internally-activated.*

Externally-activated Trojans: In Externally-activated category, the Trojan can be activated externally by adversary in his/her time of choosing. This can be done by embedding a receiver or antenna on chip and controlling it through external signals. This can also be done by accessing the internal registers and forcing them to specific date to extract secret keys. The Internally-activated category is divided into two subclasses, labeled *Always-on* and *Condition-based* as shown in Figure 4. ***Always-on***, as the name implies, indicates that the Trojan is always active and can disrupt the function of the chip at any time. This class covers Trojans that are implemented by modifying the geometries of the chip such that certain nodes or paths in the chip have a higher susceptibility to failure. We referred to these types of Trojans as 'parametric' in the *type* subclass of the physical characteristics class. In order for Always-on Trojans to meet the stealthy activation criteria, the adversary necessarily needs to insert them on nodes or paths that are rarely exercised. In the test community, such nodes and paths are referred to as *hard-to-detect faults* because the conditions needed to detect faults on them are difficult to determine and are statistically unlikely to occur using random and structural stimuli. The ***Condition-based*** subclass includes Trojans that are 'inactive' until a specific condition is met. The activation condition can be based on the output of a sensor that monitors temperature, voltage or any type of external environmental condition, e.g., electro-magnetic interference (EMI), humidity, altitude, atmospheric pressure, etc. Or it can be based on an internal logic state, a particular input pattern or an internal counter value. The Trojan in these cases is implemented by adding logic gates and/or flip-flops to the chip, and therefore is represented as a combinational or sequential circuit.

We believe that fully activation for logic-based Trojans is an NP-complete problem. Fully activation of Trojans depends on the number of Trojan inputs and the states of the circuits the Trojan is observing. The partial activation of Trojans depends on the method(s) implemented to detecting and isolating Trojans. If part of a Trojan is activated, the power consumed by the gates included in that part will contribute to the total power consumption. An important distinguishing characteristic between Always-on and Condition-based Trojans is the former is nearly 'invisible' when it is inactive while the latter is always visible to some degree when inactive. We define 'invisible' as undetectable when methods that measure the chip's digital and/or analog properties, including power consumption, are applied while the chip is undergoing some form of testing. The fact that Always-on Trojans are defined as subtle modifications to existing wire and transistor geometries indicates that the chip that embeds them will behave identically to a Trojan-free chip, under the condition that the nodes or paths altered by the Always on Trojan are not exercised by such tests, i.e., the Trojan remains inactive. In contrast, a Condition-based Trojan, which needs sensors or logic components to monitor for the activation condition, consumes power at some level and/or adds load to wires of the original circuit, which in turn changes the delay characteristics of the chip. These subtle changes to the analog characteristics of the chip occur even while the Trojan remains inactive. This implies that detecting an Always-on Trojan will necessarily require its activation while detecting a Condition-based Trojan can be accomplished *without fully* activating it, in situations where the detection method incorporates, e.g., an analysis of the chip's power consumption characteristics.

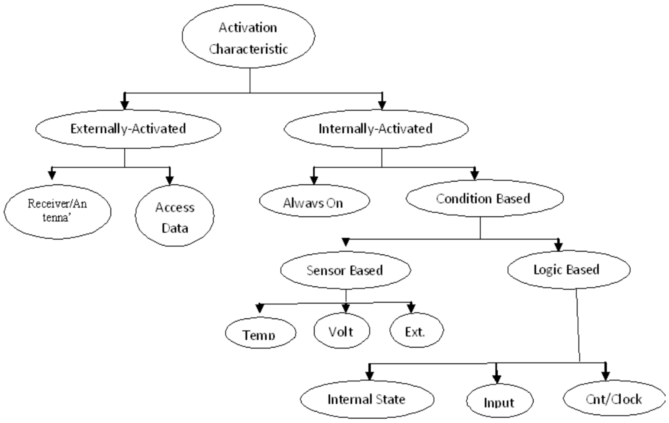


Figure 4. Trojan activation characteristics

2.3. Trojan Action Characteristics

Action characteristics identify the types of disruptive behavior introduced by the Trojan. The classification scheme shown in Figure 5 partitions Trojan actions into three categories; *Modify-function, Modify-specification,* and *Transmit-info* (see Figure 5). As the name implies, the Modify-function class refers to Trojans that change the chip's function through additional logic or by removing or bypassing existing logic. The Modify-specification class refers to Trojans that focus their attack on changing the chip's parametric properties, such as delay. The latter class represents parametric Trojans that modify wire and transistor geometries. Lastly, the Transmit-info class refers to Trojans that transmit key information from design mission mode to an adversary. An important distinguishing characteristic between modify-function and modify-specification Trojans concerns their capabilities. The nature of modify specification Trojans restricts their disruptive capabilities to actions that result in system failure. This is true because modify-specification Trojans are implemented as modifications to existing wires and transistors. Therefore, new capabilities are not possible. In contrast, the capabilities of modify-function Trojans are essentially limitless. As the examples illustrate, modify-function Trojans, once activated, can change virtually any characteristic of the chip or can introduce completely new functionality such as broadcasting confidential information over the power buss. The type of malicious behavior introduced by the Trojan once activated depends greatly on the application. The following list gives some examples but is by no means an exhaustive list. The examples each list some of the attributes associated with the Trojan as they relate to the taxonomy discussed above. The labels following the slash give the subclass. The keyword *any* is used to indicate any class or subclass is assignable under that category.

• Trojan introduces a hard short or a delay fault in a rarely exercised section of the circuit.

Physical: parametric, small, tight, **Activation**: always on, **Action**: modify-spec. /*any*

• Trojan disrupts clock to shut down the chip or affect its synchronization.

Physical: functional, large, *any*, **Activation**: condition based/ *any*, **Action**: modify-function/add

• Trojan disables encryption.

Physical: functional, large, *any*, **Activation**: condition based/ *any*, **Action**: modify-function/bypass.

• Trojan changes functional output of the chip resulting in Catastrophic failure.

Physical: functional, large, *any*, **Activation**: condition based/ *any*, **Action**: modify-function/*any*

• Trojan leaks information through side channels.

Physical: functional, large, *any*, **Activation**: condition based/ *any*, **Action**: modify-function/add

• Trojan destroys the operating environment of original circuit, such as shutting down power, generating noise to disrupt critical signals, or increasing the thermal gradients on the chip, possibly causing it to burn out.

Physical: functional, large, *any*, **Activation**: condition based/ *any*, **Action**: modify-function/add

Although the modify-function classes appear to be more attractive based on their action capabilities, they also are more ‘risky’ from the adversary’s perspective. When describing the Trojan activation characteristics, we indicated that Trojans in the condition-based class are never completely invisible, because the additional logic they incorporate needs to continuously monitor for the activation condition. In order to accomplish this, the Trojans need to ‘connect’ to the existing logic. The connection(s) introduce capacitive load to the wires and power grid to which they connect, that in turn, changes the power consumption and thermal characteristics of the chip. Moreover, the logic added to determine activation needs to perform ‘compare’ operations that also consume power and introduce temperature variations. The power and temperature anomalies introduced by the Trojans in the modify-function class can be considered a *secondary* action characteristic of the Trojan. The secondary action characteristics of the Trojan are undesirable, from the adversary’s perspective, but unavoidable, and their presence can be leveraged by certain Trojan detection methods are described in the upcoming section.

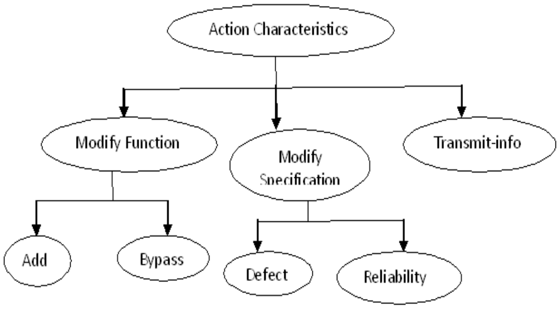


Figure 5. Trojan action characteristics

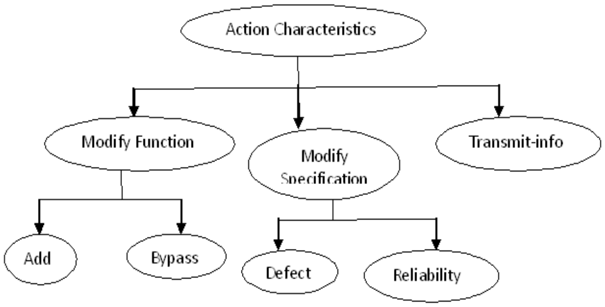


Figure 5. Trojan action characteristics

3. TROJAN SECURITY IMPLICATION

A consequence of the proliferation of microelectronics is the increasingly important role it plays in the manipulation and communication of confidential information and in the management and control of equipment. This type of microelectronics-enabled automation also makes such systems vulnerable to attack. The software threat to security is well known and many techniques have been proposed and implemented to protect systems [8]. However, threats originating in the actual hardware are new and are disruptive to software security layers that run on the hardware. This is true because software security mechanisms can be easily bypassed by malicious hardware, and such hardware is extremely difficult to detect given the trends in system complexity and IC technology. Below, we provide a list of applications that could potentially be targets for adversary to insert hardware Trojans. We acknowledge that this is not meant to be an exhaustive list of applications that are vulnerable to hardware-based security issues, but rather serves to highlight specific present and future systems that are likely to be targets for adversaries. The systems that are vulnerable to attack will continue to grow over time, which drives the need to develop more sophisticated methods for hardware authentication.

- **Military applications:** weapon control systems, battlefield communication systems, battlefield information collection and decision making systems, etc.
- **Aeronautic and astronautic applications:** space shuttle electronics, satellite electronics, aeroplane electronics, etc.
- **Civilian security-critical applications:** confidential business information management systems (banking systems, stock market information management systems, firewalls protecting company propriety information, etc.), confidential personal information management systems (medical record systems, personal financial record systems, etc.).
- **General purpose applications:** any electronic systems used to manage confidential information.
- **Secure IP:** fableless semiconductor industry, core developers.
- **Transportation security:** Secure search and information management systems, emergency awareness systems, etc.

4. TROJAN DEMONSTRATIONS

The presence of Trojans can be demonstrated by the way the leak information in the form of Thermal Energy, Optical Energy and Electro-magnetic Radiation Energy.

Thermal

✴An external resistor is electrically modulated creating thermal emission.

✴The micro-controller or other parts of the circuit are quickly saturated with operations, creating thermal emission.

✴The thermal signal is sensed using an IR camera.

Optical

✴An external LED is electrically modulated at a rate undetectable by human eye.

✴The optical signal is sensed using an optical-to-audio amplifier.

Radio

✴An external I/O pin is modulated causing radio emission.

✴The radio signal is sensed using radio receiver and post processing received signal on PC.

5. TROJAN DETECTION

In this section, we outline the general approaches for detecting Trojans. Trojan detection methods can be applied immediately after the chip is returned to the customer, either as a die on a wafer or as a packaged chip, and/or they can be applied continuously during the lifetime of the system. For the latter case, board level support systems, such as trusted companions, are needed to carry out the monitoring. Although these types of approaches are of interest, the focus of this work is on 'time-zero' detection methods, i.e., methods applied before the chip is installed in the target system. We refer to this phase as Silicon Design Authentication that is done after manufacturing testing phase.

In general, there are three basic approaches for detecting Trojans that we explain them in the following.

5.1. Failure Analysis-based Techniques

The first involves applying sophisticated failure analysis techniques such as scanning optical microscopy (SOM), scanning electron microscopy (SEM), pico-second imaging circuit analysis (PICA), voltage contrast imaging(VCI), light-induced voltage alternation (LIVA), charge induced voltage alternation CIVA, etc. [3]. Although these techniques can be effective for authentication purposes, they are also extremely time consuming and expensive. Moreover, many require the sample (chip) to be prepared by backside thinning and de-processing operations. Obviously, this approach is not suited for applications in which every chip needs to be authenticated. Another drawback is that many of these techniques are becoming increasingly ineffective for technologies in the nanometer domain. An important issue is that the adversary will most likely insert Trojans randomly in chips. Therefore, spending a large amount of time on each chip for authentication will be prohibitively expensive. As a result, new and efficient methods are required to detect Trojans with higher confidence level and minimum authentication time.

5.2. ATPG-based Trojan Detection Techniques

The second approach involves the use of 'standard' VLSI fault detection tools, such as automatic test pattern generation (ATPG). Detection of a Trojan is accomplished by applying a digital stimulus and inspecting the digital output of the chip. The digital stimulus is derived using the netlist of the chip. For Trojans of the parametric type as described in Section 2.1, the netlist of a chip is the same with and without the Trojan. This is true because parametric Trojans are introduced into the existing logic of the chip by violating design rules, i.e., thinning a wire, etc. Therefore, ATPG can be modified to target parametric Trojans. Given their stealthy activation criteria, ATPG directed to generate tests for nodes and paths that are hard to-detect, i.e., difficult to control and/or observe, is likely to yield the best results for activation and detection of Trojans. Unfortunately, ATPG is not effective for the functional Trojans, which are represented as inserted, additional logic. Without knowledge of this logic and how it is connected to the original logic in the chip, it is impossible for ATPG to perform a directed search for a vector or state that causes activation. Bear in mind, that if the activation criteria can be determined, then detection would be trivial in many cases, assuming the Trojan modifies the internal state or an output of the chip in some fashion. However, for Trojans that activate and leak information over side channels, e.g., the power supply, digital testing methods are not effective. Therefore, for functional Trojans, an ATPG approach is hindered by two problems, one that deals with activation and another that deals with detection. A third approach that can potentially solve these problems involves the measurement and analysis of the chip's side-channel signals [4] [5] [6]. For example, it is possible to stimulate the chip using digital stimuli and then measure the analog response signals of the chip, such as the transient or quiescent power supply current1.

5.3. Side Channel Signals Analysis

Another possibility is to stimulate the power grid directly by driving it with a sine wave at one position and measuring its response at another. The analog nature of the side-channel response signals enables the use of highly sensitive detection techniques. Such techniques may be able to detect functional Trojans without activating them, i.e., through the measurement of their *secondary* action characteristics as described in Section 2.3. For example, we indicated that functional Trojans are never completely inactive because of the need to continuously monitor for the activation conditions. Consider a Trojan that activates based on a specific state of a data bus in the chip. The implementation of the Trojan, in this case, requires some type of comparator to be installed that monitors the wires of the data bus. The logic of the comparators, e.g., AND gates, switches as the data bus changes and therefore consumes power. Side-channel signal analysis can potentially detect the power anomaly introduced by the operation of the comparator. Other side-channels signals include electro-magnetic field variations, temperature variations, voltage variations, etc., that occur at various locations across the chip. New methods can be developed that use such signals to detect and isolate hardware Trojans. Moreover, the highly sensitive nature of side-channel analysis techniques may allow detection of tightly coupled functional Trojans even without the application of a digital stimulus. The presence of the Trojan logic gates adds capacitance to the power grid. The presence of the additional capacitance changes in impulse response of the power grid. The impulse response of the power grid can be tested by injecting an analog stimulus onto the grid at one place and measuring the response at another. The effectiveness of side-channel-based measurement and analysis techniques can be improved by adopting design-for-hardware-trust (DFHT) techniques, which, for example, add circuitry to support the measurement and analysis processes. On-chip voltage and temperature sensors can be installed to increase the level of sensitivity of side-channel measurement and analysis techniques by providing local observability at various positions across the 2-D layout of the chip. The DFHT strategy must also incorporate a validation strategy for the on-chip support circuits because of the potential of the adversary to sabotage the sensors.

5.4. Trojan Detection Challenges

Depending upon the method used for Trojan detection, there seem to be extremely difficult challenges associated with the method. The taxonomy and discussion presented above suggest that detection strategies that depend on activating condition-based Trojans through the application of test patterns and detecting them through an analysis of the circuit's logic response may not be effective. Considering an intelligent and determined adversary, the Trojans can be inserted such that the probability of accidental detection using test patterns (functional, structural, and random) will be extremely low. Assume a Trojan with *n* number of inputs. Also, assume that *pi* is the probability of justifying a 0 or 1 on *i*th input of the Trojan circuit. If the Trojan is inserted deep into the circuit, *pi* will be extremely low. The probability (*P*) of activating this *n*-input Trojan and propagating its effect using these patterns would be:

P = *P* (activation). *P* (propagation)

P (activation) = _ni=1 *pi*.

Assume *pi*=10-3 and *n*=10, then *P*=10-30.

Considering *P* (propagation), the probability of successful propagation of the Trojan's effect if the Trojans output is connected to the circuit (e.g. Modify-function Trojan), can also worsen *P*. Note that *P* (propagation) is circuit topology dependent. This clearly demonstrates that relying on input patterns for Trojan detection may not seem to be an effective solution. When using side-channel signal analysis methods for detecting hardware Trojans, the circuit process variations will be a major bottleneck. For instance, process variations significantly impact circuit leakage currents therefore; using IDDQ like methods to detect a Trojan will suffer from inaccuracy. Trojan type and size also play an important role in detection sensitivity. Larger Trojans will consume more leakage power and are easier to be partially activated which will consume more switching current. Smaller Trojans are harder to be detected using leakage and switching current analysis since they contribute negligibly to the total power in the circuit but easier to be activated using functional or structural patterns. Detection of Trojans based on switching current analysis can be effective only if efficient patterns are generated and applied. The challenge here is to generate patterns that cause maximum switching in a small region in the circuit and minimum switching in other regions. Considering a small number of primary inputs in large and complex designs, this would seem to be a challenging task. The scan flip-flops can be used to facilitate the problem, however, the patterns must be shifted into the scan chain which makes the process significantly slow. The main advantage of using scan is in its significantly increased controllability and observability. This would cause increased switching in the circuit. New techniques must be developed to generate localized switching in the circuit to increase detection and isolation accuracy. An inserted Trojan in the circuit can in fact impact the circuit delay characteristics. Delay test methods can be used to detect such Trojans however the deficiency of current transition delay ATPG methods will challenge its efficiency. When a Trojan is inserted into a circuit, equivalent to the gate capacitance will be added to the total capacitance of the path the Trojan is taping the signal from. The amount of delay is small therefore novel methods must be developed to detect such small delay in the circuit induced by Trojans. We acknowledge that process variations can also cause small delay to the circuit and in turn cause uncertainty during detection. The Trojan can however cause a multi-path small delay injection which could potentially be detected using efficient delay testing. Also, note that depending on the type of the Trojan, we need to devise appropriate detection strategy. Some Trojans are easier to be detected using power analysis methods and some others are easier to be detected using delay analysis. Since the type and size of Trojans are not known to us, it is recommended to use both methods to target Trojans during silicon design authentication phase to increase the probability of detection. Another issue that must be addressed during Trojan detection is the time taken to verify the authenticity of each chip. A reasonable assumption is that the adversary will most likely insert Trojans randomly in a large batch of chips. Therefore, the authentication time will be significantly important for large volume of chips fabricated in an untrusted foundry.

6. Conclusion

A Trojan classification scheme is presented in this paper that partitions Trojans according to their physical, activation and action characteristics. The taxonomy can be used in conjunction with the Trojan detection methods outlined to help define their effectiveness and capabilities.

✴Hardware Trojans are a new and emerging threat.

✴Systems at risk include military systems, financial systems and even household appliances.

✴The purpose of our work is to demonstrate the dangers of Hardware Trojans.

✴We have also given Trojan detection methods.

REFERENCES:

[1] Dakshi Agrawal, Selcuk Baktir, Deniz Karakoyunlu, Pankaj Rohatgi and Berk Sunar: Trojan Detection using IC Fingerprinting, IBM T.J. Watson Research Center, Yorktown Heights, Electrical \& Computer Engineering Worcester Polytechnic Institute, Worcester, Massachusetts, Nov 10, 2006

[2] Xiaoxiao Wang, Mohammad Tehranipoor and Jim Plusquellic: Detecting Malicious Inclusions in Secure Hardware, Challenges and Solutions, 1st IEEE International Workshop on Hardware-Oriented Security and Trust (HOST'08), 2008

[3] J. Soden, R. Anderson and C. Henderson, “IC Failure Analysis Tools and Techniques -- Magic, Mystery, and Science”, International Test Conference, Lecture Series II “Practical Aspects of IC Diagnosis and Failure Analysis: A Walk through the Process”, 1996, pp. 1-11.

[4] A. Germida, Z. Yan, J. Plusquellic and F. Muradali, “Defect Detection using Power Supply Transient Signal Analysis”, International Test Conference, Sept. 1999, pp. 67-76.

[5] J. Plusquellic, D. Acharyya, A. Singh, M. Tehranipoor and C. Patel, “Quiescent-Signal Analysis: A Multiple Supply Pad IDDQ Method”, Design and Test of Computers, Volume 23, Issue 4, April 2006, pp. 278-293.

[6] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, B. Sunar, “Trojan Detection using IC Fingerprinting”, Symposium on Security and Privacy, 2007, pp. 296-310.

[7] J. Viega and G. McGraw, Building Secure Software, Addison-Wiley, 2

[8] To view a video of our hardware Trojan demonstrations please visit this link: http://www.cvorg.ece.udel.edu/defcon-16

[9]Mainak Banga and Michael S. Hsiao: A Region Based Approach for the Identification of Hardware Trojans, Bradley Department of Electrical and Computer Engineering, Virginia Tech., Host'08, 2008

ABS 01:

4G-Technology Future Warriors

K.P.Sreedhar (07931A0408)
IV B.Tech ECE
Id:sridhar_kp@ymail.com

V.Hashmi (07931A0414)
IV B.Tech ECE
Id:vallipallihashmi@yahoo.com

B.Swetha (07931A0433)
IV B.Tech ECE
Id:bswetha.reddy@yahoo.com

Lehana (07931A0456)
IV B.Tech ECE
Id:lahari.sony@gmail.com

Abstract

Information is power, nowhere is this truer than on the battlefield, where the ability to communicate clearly and rapidly pass on information spells the difference between survival and death? 4G (4th Generation) is the technology that is going to drive a soldier in the field in future. The key to empowering the military with tactical broadband voice, video and data is 4G communications technology. This technology adopts Wireless technology on the platform of fixed networks, advanced antennae technologies and more advanced wireless security technologies. Next thing is about the gear for the future warrior. Our system provides a enhanced power of vision, which provides Ground Guidance, Unit Detection, Soldier Status, Target Hand-Off and provides the Soldier Rescue during the battle. The uniform along with the armor, onboard computer which will monitor soldiers' overall physiological and psychological picture of how they are performing in the battle zone and enhanced human performance which weighs 50 pounds from head to toe against 120 pounds of the current day system present.

ABS 02:

Automated Aero Preceptor

S.VINAY KUMAR
4th B.Tech E.C.E
Email id:vinaykuar.somi@gmail.com

B. Rohith
4th B.Tech E.C.E
Email id:brungi.rohith@gmail.com

S.Shahid Basha
4th B.Tech E.C.E
Email id:shahzayed418@gmail.com

P.Kiran Kumar
4th B.Tech E.C.E
Email id: kiranben431@gmail.com

Abstract

Research projects involving unmanned aerial systems are popping up like drones in the atmosphere .The activity has come about as Unmanned Airborne Vehicles (UAVs) are increasingly being considered and used for new purposes, ranging from border patrol and crowd control to post-disaster rescue efforts.

This paper emphasizes on the development of architecture for Unmanned Airborne Vehicles (UAVs). In this paper we present a system which integrates computer vision and decision-making in an autonomous airborne vehicle that performs traffic surveillance tasks. The main factors that make the integration of vision and decision-making a challenging problem are:

1) the qualitatively different kind of information at the decision-making and vision levels,
2) The need for integration of dynamically acquired information with a prior knowledge, e.g. Geographic Information System (GIS) information,
3) The need of close feedback and guidance of the vision module by the decision-making module.

Given the complex interaction between the vision module and the decision making module we propose the adoption of an intermediate structure, called Scene Information Manager, and describe its structure and functionalities.

ABS 03:

An Insight To The Fourth Generation Communication Systems

K.Siva Krishna, T.Shankarnath Goud, S.Suraj Siddhartha, D.Chandra Babu Naidu
Dept. of E.C.E., S.K.T.R.M.College of Engg.NH-7, P.O.Kondair-509125, A.P.
Email:Krishna.siva481@gmail.com,chandrababunaidu.dondeti@gmail.com,tshankarnadh@gmail.com,buddha_siddhu@yahoo.com

Abstract

Our paper gives a brief idea of how the mobile communication has evolved and its generations. There are different generations have been developed since then each generation coming out overcoming previous generation drawbacks. Based on the study, 4G mobile technology is in a determining and standardization stage. Although 4G wireless technology offers higher data rates and the ability to roam across multiple heterogeneous wireless networks, several issues require further research and development. Since 4G is still in the cloud of the sensible standards creation, ITU and IEEE form several task forces to work on the possible completion for the 4G mobile standards as well. Under these circumstances, we will try to present about the current trends and its underlying technologies to implement the 4G mobile technology. This paper also shows some of the possible scenarios that will benefit the 4th generation technology. One such technology in fourth generation is OFDM. OFDM based technique looks more promising as a 4G standard surpassing the 3G standards .So, a complete review of OFDM is provided explaining its spectrum utilization and also how signals are multiplexed using OFDM and transmitted using MIMO technology.

ABS 04:

CMOS Active Pixel Sensors

S.SHANWAZ (07931A0403) K.SIVA KRISHNA (07931A0409)
Email id:shanwaz403@gmail.com Email id:ksiva4@gmail.com

Abstract

Output images from the sensors more likely are not optimal results for display or further processing mainly because of noise, blurriness and poor contrast. In order to prevent these problems, image processors typically accompany the image sensors as a part of the whole camera system. The integration of image sensors and processing circuits on a single monolithic chip, called smart sensing, is done to obtain better performance from sensors and make the sensing and processing system more compact. It has become a popular idea. The integration of image acquisition and processing on the same focal plane has potential advantages through low fabrication cost, low power, compact
size, and fast processing frequency. Noise and cross-talk can also be reduced through
monolithic connections instead of off-chip wires, which are the only transfer medium between two separated chips.

hus, we propose for integrating image processing with CMOS active pixel sensors on a single chip.

The CMOS active pixel sensor (APS) is a second generation solid state sensor technology that was invented and developed at JPL. The goal of the advanced imager technology effort at JPL has been the development of a "camera on a chip," which would have a full digital interface.

ABS 05:

Distributed Control Systems

S.Vinay Kumar (07931A0444)K.P.Sreedhar (07931A0408) N.Harika (07931A0407)
Email id:vinaykumar.somi@gmail.com, Sridhar_kp@ymail.com, nharika07@gmail.com

Abstract

A distributed control system (DCS) refers to a control system usually of a manufacturing system or process, in which the controller elements are not central in location (like the brain) but are distributed throughout the system with each component sub-system under the control of one or more controllers. The entire system may be networked for communication and monitoring. A DCS typically uses computers (usually custom designed processors) as controllers and use both proprietary interconnections and protocols for communication. Input & output modules form component parts of the DCS. DCS is a very broad term that describes solutions across a large variety of industries and also in Electrical power grids and electrical generation plants, Environmental control systems,Traffic signals, Water management systems, Refining and chemical plants, Pharmaceutical manufacturing

ABS 06:

Designing Real-Time and Embedded Systems with the COMET/UML method

[1] Sunilkumar H T, [2]Vasanth kumar S
[1]sunilht1990@gmail.com, [2]vassankit@gmail.com

Abstract

Most object-oriented analysis and design methods only address the design of sequential systems or omit the important design issues that need to be addressed when designing real-time and distributed applications [Bacon97, Douglas99, Selic94]. It is essential to blend object-oriented concepts with theconcepts of concurrent processing [MageeKramer99] in order to successfully design these applications.This paper describes some of the key aspects of the COMET method for designing real-time and embedded systems [Gomaa00], which integrates object-oriented and concurrent processing conceptsand uses the UML notation [Booch98, Rumbaugh99]. Examples are given from an Elevator Control System [Gomaa00].

ABS 07:

Embedded Systems

S.Vinay Kumar (07931A0444) K.Varun Vihar (07931A0443)
Email id:vinaykumar.somi@gmail.com Email id:v_drazeel@ymail.com
C.Swathi (07931A0445)
Email id:pandu.20500@gmail.com

Abstract

Embedded System is combination of hardware and software that forms the component of larger system. Hardware is normally unique to given application. Computer chips are embedded into control electronics to manage the product’s functionality. The embedded system can be categories into four categories viz.; stand alone system, real-time system, network appliance system and mobile devices.

The new development tools available today make the task easy. Also the production cost is decreasing with increase in complexity. All these developments are leading to an era of invisible computing, or hidden computing where in computer does a job without a ubiquitous and physical presence. Thus embedded devices are becoming smaller, smarter and more integrated. So needless to say, embedded software development is a very lucrative business these days.

This paper deals with definition, requirement issues, characteristics, different categories, embedded system development tools, various applications, Interactive Voice Response (IVR) (one of the application), advantages and pitfalls.

ABS 08:

Embedded Systems

C.Seskala
Sheshi.chilari@gmail.com

P.Tejaswini
ponnakanti.tejaswinireddy@yahoo.in

T.Swetha
Swetha.kalakonda@yahoo.com

K.Neeraja
neerajakatamoni@yahoo.com

Abstract

Many embedded systems have substantially different design constraints than desktop computing applications. No single characterization applies to the diverse spectrum of embedded systems. However, some combination of cost pressure, long life-cycle, real-time requirements, reliability requirements, and design culture dysfunction can make it difficult to be successful applying traditional computer design methodologies and tools to embedded applications. Embedded systems in many cases must be optimized for life-cycle and business-driven factors rather than for maximum computing throughput. There is currently little tool support for expanding embedded computer design to the scope of holistic embedded system design. However, knowing the strengths and weaknesses of current approaches can set expectations appropriately, identify risk areas to tool adopters, and suggest ways in which tool builders can meet industrial needs

ABS 09:

Future Nano CMOS and Nano Technology

Vinay .M, N.Rajath Kornaya
5th sem student, ECE Dept.KIT.Tiptur.
vin.m1990@gmail.com, rajath.10101@gmail.com

Abstract

NANOTECHNOLOGY IS OFTEN REFERRED TO AS THE NEXT INDUSTRIAL REVOLUTION.CMOS technology has been developed into the sub-100 nm range. It is expected that the nano-CMOS technology will governed the IC manufacturing for at least another couple of decades. With the device continuing scaling down, the transistor physical gate length will reach about 10nm late this decade, and will ultimately arrive at about 6nm at 14nm technology node as end of the roadmap.Though there are many challenges ahead, further down-sizing the device to a few nanometers is still on the schedule of International Technology Roadmap for Semiconductors (ITRS).Several technological options for manufacturing nano-CMOS microchips have been vailable or will soon be available. This paper reviews the challenges of nano- CMOS downsizing and manufacturing. We shall focus on the recent progress on the key technologies for the nano-CMOS IC fabrication in the next fifteen years and also on the future applications viz., medicine,house hold appliance,textile, Anti-abrasive, Weather resistant easy cleaning nano-coating .

ABS 10:

Integration Of 'IT' In Machine Tools

Md Adil Ali [1]
adilalidawood@gmail.com

Akhil Pasha [2]
mdakhil27@yahoo.com

Nisar Ahmed [3]
nisar_gfec@yahoo.com

Al-Habeeb College of Engg. & Tech., Hyd

Abstract

Abstract- Today's buzzword "IT" has revolutionized every aspects of our day today working lives. Automation of industries is one of its main contributions. Automation can be defined as technology concerned with the application mechanical, electronic and computer-based systems to operate and control production. . The main advantage of automation is increased labor productivity at reduced labor

cost. The heart of machine tool is a CNC (Computerized Numerical Control) system, which coordinates with the displays, PLC, drives system, and feed back systems. This paper deals with the data acquisition, program management and remote diagnosis of the system.

Numerical control is for of programmable automation in which the mechanical actions of a mechanical tool or other equipment are controlled by a program containing coded alphanumeric data. Computerized numerical control (CNC) is defined as an NC system who's MCU is based on a dedicated microcomputer rather than a hardwired controller. Remote diagnosis or Tele service offers remote support for CNC systems, making it a cost-effective alternative to service calls. Lot of time can be saved by this.

ABS 11:

Molecular Logic Array and Memory Device in Nano technology

T.Valli Gayathri
JCET, HYD, AP.
Gayathri.thangirala@gmail.com

P.Sindhusha
JCET, HYD, AP
sindhusha_pinnoju@yahoo.com

Abstract

This document surveys a number of future molecular nanotechnology capabilities of Computer applications, launch vehicle improvements, and active materials appear to be of particular interest. A new self-assembly technique is proposed to fabricate sub-10nm wide nanotech array, which is an improved selfassembledprocess the direction of surfactant template. For the realization of high density logic and memory device in nanotechnology field, many researchers have applied functional organic rectifying molecules for molecular logic gate and switching molecules for molecular memory device by using self-assembled monolayers (SAMs). We have been worked on a development of 3x3 molecular AND/OR logic gates. To implement the molecular logic gates, we synthesized rectifying molecules having high rectification ratio and fabricated 3x3 array device with 9 nano-pores whose diameter is about 100nm. Finally, we can successfully implement the AND/OR logic gates. In addition, we have worked on a fabrication of molecular memory device. To implement single organic monolayer device in the vertical structure of metal-molecule-metal electrode, it is required to solve fundamental problems, that is, an electrical short. A yield of the molecular device using self assembled single monolayer is less than 5 % even in nano-pore device. For the realization of molecular memory device using single monolayer, we introduce new way by using organic conducting electrode. The yield of the newly developed molecular device is 50-60%, which is improved over 10 times. Furthermore, we like to report a device fabrication by using a nano-imprinting lithography technique

ABS 12:

Motion Tracking For 2d-Mesh Video Object Using Low Power VLSI Design

Gururaju .K.B. **Prasana kumar .B.R.** **Channakeshava .V.N.**
Electronics and Communication
Kalpataru Institute of Tehnology Tiptur-572002
e-mail id: prasanna_kumar_19@yahoo.com

Abstract

This paper presents a low power VLSI architecture for video object motion tracking. Power has been reduced at both algorithmic and arithmetic levels. The video object is modeled as a 2D hierarchical structured mesh, where the deformation of the mesh represents the dynamics of the object across the video sequence. The algorithm benefits from the small number of bits that describes the mesh topology. Low power has been achieved in the algorithm level by: (1) modeling the mesh into independent triangular patches that can be processed in parallel, (2) each patch is hierarchical triangulated using structured technique, which can be pipelined using simple

unit, and (3) and using the three steps motion estimation algorithm to simplify the motion estimation of the mesh nodes. On the arithmetic level, low power has been achieved by using multiplication-free affine transformation because of the followed triangle topology. A VLSI architecture is developed based on the proposed algorithm. The architecture consists of two main parts, a mesh-based motion estimation unit and a mesh-based motion compensation unit. The first unit is based on parallel block matching motion estimation to optimize the latency. The second uses parallel threads. Each thread implements a pipelined chain of scalable multiplication-free affine units.

ABS 13:

MEMS

S.Naziya Yasmin, and K.Rekha,
Gates Institute of Technology,
E-mail:shaik.naziyayasmin@gmail.com

Abstract

A novel high speed, high capacity electron-beam recording technique using nano technology in a Hard Disk Drive form factor is described. The e- beam source is a carbon nanotube (CNT) emitter and can be gated at rates up to several gigahertzes. The. The planned recording media is Phase Change with sub-nanosecond response times, and data read-out by Secondary Electron Emission is anticipated. The key parameters for generating the recording beam are described and a preliminary design are discussed in which the CNT based Read/Write head replaces the magnetic head in a standard Hard Disk Drive (HDD). The technique sidesteps limits associated with HDD technology and potentially provides far higher recording densities and higher data rates than possible with conventional magnetic recording. The NS3 NanoTech Disk (NTD) approach may provide a path forward for HDD's to the low nanometer mark scale.

ABS 14:

Palm Vein Authentication Technology and Its Applications

V.Raghu Pavan and Ch.Brahma Vahini
v.raghupavan@gmail.com

Abstract

This paper discusses the contactless palm vein authentication device that uses blood vessel patterns as a personal identifying factor. The vein information is hard to duplicate since veins are internal to the human body. The palm vein authentication technology offers a high level of accuracy, and delivers the following results: a false rejection rate (FRR) of 0.01%, and a false acceptance rate (FAR) of 0.00008% or lower, based on Fujitsu research using the data of 140,000 palms. Several banks in Japan have used the palm vein authentication technology for customer identification since July2004. In addition, Fujitsu

has integrated the technology into the access control of electronic door lock systems.Fujitsu plans to further expand applications for this technology by downsizing the sensor and improving the verification speed.

ABS 15:

Floating-Point FPGA: Architecture and Modeling

Abhijith Y.P[1] , Pradeep S Agasanahalli[2]
vishalabhi456@gmail.com[1] , pradeepsa.143@gmail.com[2].

Abstract

This paper presents architecture for a reconfigurable device that is specifically optimized for floating-point applications. Fine-grained units are used for implementing control logic and bit-oriented operations, while parameterized and reconfigurable coarse-grained units incorporating word-oriented lookup tables and floating-point operations are used to implement data paths. In order to facilitate comparison with existing FPGA devices, the virtual embedded block scheme is proposed to model embedded blocks using existing field-programmable gate array (FPGA) tools. This methodology involves adopting existing FPGA resources to model the size, position, and delay of the embedded elements. On selected floating-point benchmark circuits, our results indicate that the proposed architecture can achieve four times improvement in speed and 25 times reduction in area compared with a traditional FPGA device and bit-oriented operations.

ABS 16:

Era of Robotics

Sumanth B S , Sharath H
5th sem students, ME Dept
Kalpataru Institute of Technology,Tiptur.
Sumanthra.bs@gmail.com,sharu.sharathh076@gmail.com

Abstract

Robotics is a technology with a future, and it is a technology for the future. If present trends continue, and if some of the laboratory research currently underway is ultimately converted into practicable technology. Recent developments in the field of robotics and associate systems for precision operations are discussed here. Research trends in robotics have been towards minimizing the need for a human presence in the hazardous and uncomfortable working area. The different applications perform a series of pre programmed tasks with accuracy and precision.

A variety of mission payloads e.g. a manipulator, welding attachment, cleaning device, camera, crack detection probe, etc to be carried on board. After completing the task, the vehicle should be free to begin another task or to return to the base station. In order to achieve these objectives, it is necessary for the robot vehicle to have precision control systems, drives and dexterous manipulator systems on board. A self contained, intelligent, decision, autonomous robot is the goal of current research in ROBOTICS. A self contained, intelligent, decision making autonomous applications is the goal of “ROBOTICS”.

ABS 17:

Robotics on Agriculture

V.Raghu Pavan Ch.B.Vahini, K.Chandra Sekhar , U.V.Surendra Reddy
v.raghupavan@gmail.com , vahinichitella@gmail.com,chandu.nav@gmail.com
king4decade@gmail.com

Abstract

Current methods for off-road navigation using vehicle and terrain models to predict future vehicle response are limited by the accuracy of the models they use and can suffer if the world is unknown or if conditions change and the models become inaccurate .In this paper, an adaptive approach is presented that closes the loop around the vehicle predictions. This approach is applied to an autonomous vehicle known as field robots used in agriculture. Agricultural Robotics is the logical proliferation of automation technology into bio systems such as agriculture, forestry, Presently a number of researches are being done to increase their applications. Some of the scientist contributions are mobile robot, flying robot, forester robot, Demeter which are exclusively used for agriculture. A brief discussion is being done about the types of robots which increase the accuracy and precision of the agriculture. Experiments are being done on newly proposed world's smallest, weightless robot for using them as scouts in fields. Even in developing countries, such as India and Brazil, farmers are interested in using robots to tend fields of crops, pick fruit, or even maintain animal. At the present time, agriculture robots must have human interaction in order to compensate for programming complexity issues.

ABS 18:

Robotics

S.Naziya Yasmin and K.Rekha
Shaik.naziyayasmin@gmail.com and rekhakarakala@gmail.com

Abstract

"ROBOTICS" is a fast growing field which deals with manmade machines designed to execute the desired jobs. The present project is an attempt to highlight the utility of robot in the present days. As the microcontroller is compact and heart of our robot , this controls every movement of it .This will get actuated by the signals coming from the computer through the transmitter of frequency 433MHz which can send these signals in the range 200-300 mtrs. The receiver of frequency 433MHz detects these signals and will give it to the microcontroller which will control the motion of robot. In future, if extra peripherals are added to it, then this will have many more applications that are needed in the present days.

ABS 19:

Robotics

C.Seskala **P.Tejaswini**
Sheshi.chilari@gmail.com **ponnakanti.tejaswinireddy@yahoo.in**
T.Swetha **K.Neeraja**
Swethareddy8067@yahoo.com

Abstract

In order to be truly robust, deployed robot systems must be capable of adaptation. Robot programming is never truly finished. It is always necessary to modify, tweak and tune our systems in the field. Environmental conditions in the real world are very different from the ones in our lab, causing sensors and actuators to perform differently. In this paper, we present construction of control programs for mobile robots based on a simple natural language description of task to be performed. We also describe about the robot photographer system and give some observations based on some deployments. We outline the overall architecture of **Lewis**, a robot photographer system, which navigates through the environment, opportunistically taking photographs of people and describe how the various components inter-relate. We thus conclude by presenting a wide variety of robot's applications used in different fields.

ABS 20:

RTL synthesis of zigzag scanning for MPEG applications

B.Hari Krishna[1] B.Dwarakanadh[2] I.Lakshmi Priya[3] K.Prasad[4]
harikrishnab422@gmail.com dwarakanadhb@gmail.com inakollupriya@gmail.com
prasadece463@gmail.com
Audisankara college of engineering and technology
Nh-5 bypass, Gudur, Spsr Nellore dist. Andhra pradesh

Abstract

Communication is playing a vital role in today's life. For effectively communicating through long distances amount of data should be less and band width, transmission rates should be as large as possible. In order to reduce the amount of data, we go for several compression techniques. In transmission of videos, lossless compression is of much importance as the video has to be restored at the receiver end efficiently. The main objective of this project is to develop a HDL coding to align the pixels of image in zigzag order or an alternate order, which is used in the implementation of Variable Length Coding (VLC). Variable Length Coding (VLC) is the final lossless stage of the MPEG video compression unit.VLC is done to further compress the quantized image. VLC consists of three steps: zigzag scanning, Run Length Encoding (RLE), and Huffman coding. In zigzag scanning the quantized coefficients are read out in a zigzag order. By arranging the coefficients in this manner, RLE and Huffman coding can be done to further compress the data. The scan puts the high-frequency components together. These coefficients are then coded as a run-length pair where run is the number of occurrences of a value and the length is the amplitude

ABS 21:

WiMAX Technology

S. Shanwaz , m.veeresh kumar , g.rakesh, s.threenath
Sri Kottam Tulasi Reddy Memorial College of Engineering
Department of electronics and communication engineering

Abstract

WiMAX is a telecommunications technology that provides wireless transmission of data using a variety of transmission modes, from point-to-multipoint links to portable and fully mobile internet access. The technology provides up to 72 Mbit/s symmetric broadband speeds without the need for cables. The technology is based on the IEEE 802.16 standard also called broadband wireless access. The name"WiMAX" was created by the WiMAX forum, which was formed in June 2001 to promote conformity and interoperability of the standard. The forum describes WiMAX as "a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL"

ABS 22:

WiMax: The Next Frontier Broadband Wireless

Sesikala N.Harika P.Tejaswini
Electronics & Communication Engineering
Sri Kottam Tulasi Reddy Memorial College of Engineering, Kondair ,
Mahaboob Nagar

Abstract

A wireless revolution is seeping into our daily lives never before. Sooner or later we are all going to go wireless. Broadband Wireless Access has occupied a niche in the market for about a decade. The recently developed **Blue tooth** wireless technology is a low power, short-range technology for ëad hocí cable replacement and it enables people to wirelessly combine devices wherever they bring them. Due to the short-range limitations of Blue tooth, the recent emergence of Wifi has replaced it. Wifi popularly known as **802.11** is a moderateñrange, moderateñspeed technology based on Ethernet. It allows people to wirelessly access throughout a location. Although the technologies share a **2.4GHz** band, they have potentially overlapping applications. As more and more people use Wifi, more and more people are getting frustrated with its coverage limitations. The demand for more coverage has opened a door for WiMax. **WiMax** built on **IEEE 802.16** standards is a wireless technology that provides high throughput broadband connections over long distances .Due to itís high security, robustness and mainly huge data rates is soon to replace/support existing wireless access technologies like Wifi, Bluetooth, etc and thus believed to be the next generation of wireless access technology. When commercially available, Wimax will offer fixed, nomadic and mobile wireless broadband connectivity without needing direct line-of-sight access to a base station.

ABS 23:

Wireless sensor networks

K.Neeraja and T.Swetha K.Swetha, and P.Tejaswini
Electronics & Communication Engineering
Sri Kottam Tulasi Reddy Memorial College of Engineering,
Kondair, Mahaboob Nagar

Abstract

A **wireless sensor network** (WSN) consists of spatially distributed autonomous sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants.[1][2] The development of wireless sensor networks was motivated by military applications such as battlefield surveillance. They are now used in many industrial and civilian application areas, including industrial process monitoring and control, machine health monitoring[3], environment and habitat monitoring, healthcare applications, home automation, and traffic control.

ABS 24:

Haptics

Malashree. C Leela. B. N Arpana. R
Kalpataru Institute of Technology, Tiptur, KA, INDIA

Abstract

Haptic is a “Science of applying the tactile sensation to human interaction with computer”. In our paper we have discussed the basic concept behind Haptic along with the Haptic devices and how these devices are interacted to produce sense of touch and force feedback mechanism also the implementation of this mechanism by means of Haptic rendering and content detection were discussed.

ABS 25:

An optimization approach in VLSI Architecture of DVB Symbol Deinterleaver

Praveen A H B.Sajidha Thabassum
Department of Electronics & Communication Engineering
Dr.Ambedkar Institute of Technology, Bangalore-560056, INDIA.
nandan.praveen@gmail.com *tabumustaq@gmail.com*

Abstract

Abstract—In this paper, an efficient symbol-deinterleaver architecture compliant with the digital-video-broadcasting (DVB) standard is proposed. By partitioning the entire symbol buffer into four separate parts with a special low-conflict access control strategy, the symbol deinterleaver can be implemented with four-bank single-port on-chip memory blocks with slight overhead. The experimental result shows that 30% savings of hard- ware cost can be achieved compared with the conventional double-buffer approach. In addition, a lookahead online circuit of a symbol permutation-address generator is also proposed, which can provide the required permutation addresses every cycle to avoid either the use of a lookup table or an extra temporary buffer. Being the major part of the entire DVB forward-error-correction decoder, the proposed symbol deinterleaver can contribute a great saving of the overall decoder cost.

ABS 26:

Fuzzy logic based content protection for Image resizing by seam carving

Chaitra R Ananth **P.V. Kavya** **V.Naveen kumar**
ananth.chaitra49@gmail.com kavya.goodchum@gmail.com Naveenv6@gmail.com
Sri kottam tulasi reddy memorial college of engineering
Kondair, itikyala (mandal), mahaboobnagar-509129(a.p)

Abstract

It is often required that image resizing be done intelligently in order to preserve important content. Most image resizing techniques fail to identify and protect important objects, or produce non-photorealistic images. In this paper, a novel low computation-cost method to protect human features during resizing is presented. Seam carving, an effective image resizing algorithm, fails to protect important objects in images, when either the energy content of the object is low with respect to its surroundings, or, the number of seams removed is very large. A fuzzy logic based protection for seam carving using fuzzy segmentation coupled with neural network skin detection is introduced. Using the above tools the energy image produced in seam carving is manipulated, in order to solve the problems in seam carving while simultaneously achieving content aware resizing with protection of human features.

ABS 27:

An Efficient Utilization of a Power In VLSI Design

S. Vinay Kumar (07931A0444), G. Balaji (07931A0430) , V. Naveen (08935A0408), P.Kiran Kumar(07931A0431)
Email id:vinaykumar.somi@gmail.com, balurock430@gmail.com, naveenv6@gmail.com,
kiranben431@gmail.com
Sri kottam tulasi reddy memorial college of engineering
Electronics and communication engineering

Abstract

Efficient power utilization is a major problem in today's electronics industry. The need for low power has caused a major paradigm shift where power dissipation has become as important a consideration as performance and area. This paper reviews various strategies and methodologies for designing low power circuits and systems. It describes the many issues facing designers at architectural, logic, circuit and device levels and presents some of the techniques that have been proposed to overcome these difficulties. The article concludes with the future challenges that must be met to design low power, high performance systems. Another crucial driving factor is that excessive power consumption is becoming the limiting factor in integrating more transistors on a single chip or on a multiple-chip module. Unless power consumption is dramatically reduced, the resulting heat will limit the feasible packing and performance of VLSI circuits and systems. With this understanding, we can now consider how to reduce physical capacitance. From the previous discussion, we recognize that capacitances can be kept at a minimum by using less logic, smaller devices, fewer and shorter wires.

ABS 28:

Image Segmentation

Y.Swathi priyadarshini
E.C.E Department
E-Mail:swathipriyadarshini1990@gmail.com

B.Shireesha
E.C.E Department
E-Mail:kiranshiree@gmail.com

Gates Institute of Technology, Gooty
(Affiliated to JNTU, Anantapur)

Abstract

In the earlier days there were no techniques in identifying the defects during the manufacturing phase of the machines nor there were techniques in identifying the objects from the Satellite Images. As a result there were difficulties in identifying the reasons for the failures of those machines. Since the human eye cannot identify or modify the smallest components of the images, digital image processing has been a major study. This technique provided solutions for many of them. Differentiating two images which looked homogeneous was a great challenge; this is where Image segmentation took its way. In today's world the need raised to the extent that each pixel should be distinguished from the other. The need for image segmentation is evolving at a pace bringing in new techniques lying stress on the efficiency. The study of image at pixel level burgeoned its use in many areas and this paper advocates the methods for distinguishing images right from the pixel level. This technique of image segmentation proves that the images which appear homogeneous may not exactly be the same.

ABS 29:

Brain Fingerprinting Technology

T. Naveen kumar Reddy
Naveentanagala448@gmail.com

B.Krishna Kishore Reddy
vishalnkishore@gmail.com

Sri kottam tulasi reddy memorial college of engineering
Kondair, itikyala (mandal), mahaboobnagar-509129 (A.P)

Abstract

Brain Fingerprinting is a new computer-based technology to identify the perpetrator of a crime accurately and scientifically by measuring brain-wave responses to crime-relevant words or pictures presented on a computer screen. Brain Fingerprinting has proven 100% accurate in over 120 tests, including tests on FBI agents, tests for a US intelligence agency and for the US Navy, and tests on real-life situations including felony crimes.

ABS 30:

Performance Improvement in Mobile Communications Using Antenna Array -A modern approach in wireless communication

V.Hashmi , B.Lehana reddy, B.Swetha reddy
EmailID:_vallipallihashmi@yahoo.com , bswetha.reddy@yahoo.com, lahari.sony@gmail.com
Sri kottam tulasi reddy memorial college of engineering
kondair, itikyala (mandal), mahaboobnagar-509129(A.P)

Abstract

The demand for wireless mobile communications services is growing at an explosive rate, with the anticipation that communication to a mobile device anywhere on the globe at all times will be available in the near future.An array of antennas mounted on vehicles, ships, aircraft, satellites, and base stations is expected to play an important role in fulfilling the increased demand of channel requirement for these services. It provides a comprehensive treatment, at a level appropriate to non-specialists, of the use of an antenna array to enhance the efficiency of mobile communications systems. It presents an overview of various mobile communications systems, including land-mobile, indoor-radio, and satellite-based systems. It discusses advantages of an array of antennas in a mobile communications system, highlights improvements that are possible by using multiple antennas compared to a single antenna in a system, and provides details on the feasibility of antenna arrays for mobile communications applications.

ABS 31:

Worldwide Interoperability for Microwave Access (WIMAX)

Deepak Kumar. M - myle.deepak@gmail.com
Gavish .P .N - gavish.pn@gmail.com
Kiran .S - sskiran51@yahoo.com
Dept of ECE, BITM Bellary

Abstract

This paper presents the features of the Worldwide for Microwave Interoperability Access (WiMAX) technology. The broadband connectivity requires lots of cabeling. This reduces the flexibility of the broadband connection, the connectivity to the outside world without the cable has come, but the connectivity is not broadband. **The WiMax**-Worldwide Interoperability for Microwave Access and it also goes by **IEEE** name as **802.16**. WiMAX "a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL". In this paper we discuss the comparisons between Wireless Fidelity (Wi-Fi) and WiMAX. Several references have been

Included at the end of the article for those willing to know in detail about certain specific topics

ABS 32:

A Solution to Remote Detection of Illegal Electricity Usage via Power Line Communications

G.nandeesh, Goutham.k, M.mahiboob, Venkat
Dept: E.CE BITM, Bellary
[1]Gautham.K and [2]Nandeesh
[1]Gouthamk.28@gmail.com , [2]nandeesh.ece@rediffmail.com

Abstract

Power line communication (PLC) presents an interesting and economical solution for Automatic Meter Reading (AMR). If an AMR system via PLC is set in a power delivery system, a detection system for illegal electricity usage may be easily added in the existing PLC network. In the detection system, the second digitally energy meter chip is used and the value of energy is stored. The recorded energy is compared with the value at the main kilo Watt-hour meter. In the case of the difference between two recorded energy data, an error.

ABS-33:

Mobile Ad Hoc Networking: An Essential Technology for Pervasive Computing

Megha.B.Patil , Niveditha.S
Dept. of Electronics and Communication Engineering,
Ballari Institute of Technology and Management,
Bellary, Karnataka, India.

Megha27patil@gmail.com

Abstract

In the near future, a pervasive computing environment can be expected based on the recent progresses and advances in computing and communication technologies. Next generation of mobile communications will include both prestigious infrastructure wireless networks and novel infrastructure less mobile ad hoc networks (MANETs). A MANET is a collection of wireless nodes that can dynamically form a network to exchange information without using any pre-existing fixed network infrastructure. The special features of MANET bring this technology great opportunity together with severe challenges. This paper describes the fundamental problems of ad hoc networking by giving its related research background including the concept, features, status, and applications of MANET. Some of the technical challenges MANET poses are also presented, based on which the paper points out some of the key research issues for ad hoc networking technology that are expected to promote the development and accelerate the commercial applications of the MANET technology. Special attention is paid on network layer routing strategy of MANET and key research issues include new X-cast routing algorithms, security & reliability schemes, QoS model, and mechanisms for interworking with outside IP networks.

ABS-34:

4G Cellular Networks - Its Scope and Challenges Ahead

Ashok Kumar .m , Rishika R. Soni
Dept. of Electronics & Communications Engineering,
Ballari Institute of Technology & Management,
Bellary, Karnataka. (India).
ashmax24@gmail.com, deeps.26dec@gmail.com

Abstract

Since the inception of mobile communications in the early 1980's we have witnessed an ever-growing increase in the development of mobile communication technology. Analog wireless communication systems were replaced by digital ones ,voice services are being complemented with data services, supported data transfer speeds have increased by more than a thousand-fold, network coverage has been stretched to cover virtually entire countries and continents & many other remarkable achievements have taken place in a relatively short period. As we are moving to the next generation, we are still lacking the precise definition of the architecture and the successful deployment path of the 4G technology. Several theories have been developed looking at different standards and aiming to select and develop the most promising one. A sincere attempt is thus been made in this paper to define 4G standards with the help of available 3G standards, conceptualize a flexible and expandable architecture to provide multiple possibilities for current, future services and applications within a single terminal. The paper also aims at exploring the various techniques for the successful deployment of 4G and focuses on the technological challenges involved in mobile telephones as the wireless cellular networks evolves further to offer a multitude of services.

ABS-35:

Micro-electro mechanical system: an overview of technology

Kevin. V. Trada, Abul Faiz Iqbal
Dept. Of electronics and communication,
Ballari Institute of Technology and management
Bellary, Karnataka, India
Coolkevin729@yahoo.com

Abstract

Imagine a machine so small that it is imperceptible to the human eye. Imagine working machines no bigger than a grain of pollen. Imagine thousands of these machines batch fabricated on a single piece of silicon, for just a few pennies each. Imagine a world where gravity and inertia are no longer important, but atomic forces and surface science dominates. Imagine a silicon chip with thousands of microscopic mirrors working in unison, enabling the all optical network and removing the bottlenecks from the global telecommunications infrastructure. You are now entering the micro domain, a world occupied by an explosive technology known as MEMS. A world of challenge and opportunity, where traditional engineering concepts are turned upside down, and the realm of the "possible" is totally redefined The paper outlines the fabrication techniques, benefits, applications and challenges of MEMS. MEMS, often referred to as micro systems technology, are fabricated using modified silicon and non-silicon fabrication technology. It reduces cost and increases reliability of the system. MEMS--scale accelerometers, geophones and gyros are replacing some of the standard size precursors and are establishing new markets of their own. Commercial applications are inertial sensors, pressure sensors. This technology is also used in industrial, consumer and auto motive marketing. MEMS can have on the commercial and defence markets, industry and the federal government have both taken a special interest in their development.

ABS-36:

Nanorobotics

RakeshVarma, Piyush Kumar
Dept. Of electronics and communication,
Ballari Institute of Technology and management
Bellary, Karnataka, India
rakeshvarma078@gmail.com, pkpiyushprofessional@gmail.com

Abstract

Nanotechnology is so new that no one is really sure what will come out of it. Even so, predictions range from the ability to reproduce things like diamonds and food to the world being devoured by self-replicating nanorobots. Many new nanotechnology research fields require a high degree of precision in both observing and manipulating materials at the atomic level[1]. The advanced nanorobotics technology needed to manipulate materials at this scale, a million times smaller than a grain of sand, is being developed .The integration of different technologies to act as simultaneous real-time nanoscale `eyes´ and `hands´, including the advanced nanorobotics, high-resolution ion/electron microscopy, image processing/vision control and sophisticated sensors, will be the key to realizing such nanomanipulation.This paper presents the major aspects of nanorobotics which are at the verge of implementation and would be no less than revolution in the field of medicine if brought into reality.

ABS-37:

Cutting Edge Trends in Industrial IGBT Module Technology

Latha S. and G. Madhavi
Department of Electronics and Communication Engineering
Ballari Institute of Technology and Management, Bellary -583 104, KA, INDIA
latha.14389@gmail.com

Abstract

More than ten years have elapsed since IGBT modules first emerged as the preferred power semiconductor device for a wide range of industrial applications. During this time IGBT chip and module packaging technology has evolved through multiple generations each with incremental improvements in performance and reliability. At the same time optimized processing techniques and improved yields have reduced the cost of chip manufacturing. As a result, there are often attractive opportunities to upgrade the performance of industrial power conversion equipment while simultaneously reducing cost. This paper will summarize the latest advances in IGBT technology and the implications for future applications.

ABS-38:

Signal and Image Processing: It's Methodology and Research areas

Nivedita. K and Geetha G
Dept. of Electronics and Communication Engineering,
Ballari Institute of Technology and Management,
Bellary, Karnataka, India.
nivesona@gmail.com

Abstract

In the age of multimedia, images has become an integrated part of human life. No other medium can offer the expressive power of the image or a video sequence, because images contain an enormous amount of information. Drawing or photo of a complex technical would dramatically change the clarification and simplification of the task of description. At a glance one can grasp much more from an image than from a descriptive text. Clearly, images contain and convey information in the form, which human beings can easily process. Visual system allows handling an enormous amount of data in a very short time. It is now possible to integrate anything into an image using image processing techniques. Image processing has taken rapid strides since the invention of computer. Image processing is no longer confined to large industries and large scientific labs. One can sit at his own desk in front of his personal computer and process images. This paper aims in gaining the basic concept involved in digital image processing without much of mathematics involved. Starting from the definition of image it briefly covers the structure of digital image, steps and tools involved in digital image processing and applications.

ABS-39 :

3G and 4G cellular networks: their impact on today's mobility solutions…and future mobility strategies

Sreekanth. J, Prashant Pandey,
Dept. of Electronics and Communication Engineering,
Ballari Institute of Technology and Management,
Bellary, Karnataka, India.
sree9739458772@gmail.com

Abstract

The procurement of wireless communications systems are uniquely identified by "generation" designations. Introduced in the early 1980s, first-generation (1G) systems were marked by analog-frequency modulation and used primarily for voice communications. The introduction of digital systems led the researchers and scientists to develop digital data communication systems popularly known as Second Generation (2G) communication networks in 1990’s supporting low-band communications like Global Systems for Mobile communications (GSM) and Personnel Digital Communications (PDC) developed by the combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA), serving over 248 million users today. The wireless system in widespread use today goes by the name of 2.5G—an "in-between" service that serves as a stepping stone to 3G. Whereby 2G communications is generally associated with GSM service, 2.5G is usually identified as being "fueled" by General Packet Radio Services (GPRS) along with GSM. The legacy continued with third-generation (3G) systems, making their appearance in late 2002 and in 2003, is designed for voice and paging services, as well as interactive-media use such as teleconferencing, Internet access, and other services. Segue to fourth-generation (4G), the "next dimension" of wireless communication. The 4g wireless uses Orthogonal Frequency Division Multiplexing (OFDM), Ultra Wide Radio Band (UWB), and Millimeter wireless and smart antenna. This paper presents an overview of current technology trends in the wireless technology market, a historical overview of the evolving wireless technologies, 3G wireless technology standards to address the growing demand for wireless multimedia services. The paper also introduces 4G technologies and throws light on the strengths and weakness’ of 3G and 4G cellular networks. This paper also shows some of the possible scenarios that will benefit the 4th generation technology

ABS-40:

The Impact of Eye Gaze on Communication using Humanoid Avatars

Ankita.G, Anusha.K
Dept. of Electronics and Communication Engineering,
Ballari Institute of Technology and Management,
Bellary, Karnataka, India.
bhumeei@gmail.com ,
anushakarur@gmail.com

Abstract

In this paper we describe an experiment designed to investigate the importance of eye gaze in humanoid avatars representing people engaged in conversation. We compare responses to dyadic conversations in four mediated onditions: video, audio-only, and two avatar conditions. The avatar conditions differed only in their treatment of eye gaze. In the random-gaze condition the avatar’s head and eye animations were unrelated to conversational flow. In the informed-gaze condition, they were related to turn-taking during the conversation. The head animations were tracked and the eye animations were inferred from the audio stream. Our comparative analysis of 100 post-experiment questionnaires showed that the random-gaze avatar did not improve on audio-only communication. The informed-gaze avatar significantly outperformed the random-gaze model and also outperformed audio-only on several response measures. We conclude that an avatar whose gaze behaviour is related to the conversation provides a marked improvement on anavatar that merely exhibits liveliness.

ABS-41:

Moment Based Fingerprint Matching

Bhagya lakshmi M[1], K.M.Sadyojatha[2], Juber M.A[3]
Dept. of Electronics and Communication Engg
Bellary Institute of Technology and Management, Bellary.
[1] bhagya.ch.26@gmail.com ,[2] saddukm@gmail.com, [3] ma.juber@gmail.com

Abstract

Biometric recognition systems offer greater security and convenience than traditional methods of personal recognition. Along with the rapid growing of this emerging technology, the system performance, such as accuracy and speed, is continuously improved. At the same time, the complexity of the biometric system itself is increasing. Fingerprint Identification is the method of identification using the impressions made by the minute ridge formations or texture patterns found on the fingertips. Fingerprints offer an infallible means of personal identification. A large number of fingerprint matching algorithms are proposed till date. These algorithms are mainly based upon image correlation, filter banks for feature extraction or texture descriptors. In this work statistical approach of fingerprint matching is employed so that it is simple, cost effective, and computationally less complex. The fingerprint image which is extracted from a biometric device is divided into smaller areas say 3 x 3, 5x 5 etc. The size of this small area is selected heuristically. The background of this image (other than the image) intensity values are made zero. Then image matrix of nonzero elements is extracted. The Mean, Variance, Std, Mean (mean), Mean (Variance), Mean (Std), Variance (mean), Variance (Variance), Variance (Std), Std (mean), Std (Variance) & Std (Std) are calculated. Finally obtained results are analyzed to find similarities between different images. It is found from the study that it is possible to match and identify two fingerprints for similarities. Some of the sample results are presented in the text

ABS-42:

I^2C Implementation on Risc

Prabha. K
Dept. of Electronics and Communication Engineering,
Ballari Institute of Technology and Management,
Bellary, Karnataka, India.
Prabhade8@gmail.com

Abstract

Providing a proper means of communication between IC’s for exchange of information or signals, is major aspect in the embedded systems, which led to the I^2C bus. Embedded systems getting faster, smaller and more power efficient with more features. It can be easily interfaced with the outside world. When connecting multiple devices to a embedded systems, the address and data lines of each of the devices were conventionally connected individually. This would take up precious pins of the microcontroller results is lot of traces on the PCB and require more components to connect everything together. This made these systems expensive to produce and susceptible to interference and noise. The need for a cost effective inter-IC bus for use in consumer, telecommunications and industrial electronics, has led to the development of the Philips I^2C bus. Today the I^2C bus is implemented in a large number of peripherals and microcontrollers, making it a good choice for low speed applications. The goal of this application is to implement an I^2C communications software interface for devices which have I^2C peripheral. The software of this application performs I^2C

master transmitter and master receiver functions.I^2C protocol is a widely used and uses only 2 pins to communicate and the size of the IC’s are pretty small and increases the performance of the process. In

this project we have shown how the I^2C can work with two different RISC processors, namely PIC and the ARM microcontroller. The PIC is interfaced to the RTC (real time clock) which the ARM is interfaced to the memory both of which work on the I^2C as it interfaces the same using this protocol.

ABS-43:

Design and implementation of a three dimensional CNC machine

Venkata Krishna Pabolu, Prof. Nenkkanti Venkata Rao, K.N.H. Srinivas, & Dr. A.S.C.S Sastry
paboluvenkat@gmail.com, venkatnekkanti@rediffmail.com, **knh.tridents@gmail.com,**
Indiaascssastry@rediffmail.com
ECE Dept, SVEC, Tadepalligudem
Pedatadepalli, Andhra Pradesh, India

Abstract

This paper discusses the design and implementation of low cost three dimensional computerized numerical control (CNC) machines for Industrial application. The primary function of this microcontroller based CNC machine is to cut the metal in to required shape. This discuss is focused on communication between Personal computer (PC) and a numerical control machine. The objective to devise a computer controlled cutting machine arose from increasing demand for flexibility and cutting with respect to edge quality. The system has an 8 bit microcontroller based embedded system to achieve cost effectiveness and also maintains the required accuracy and reliability for complex shapes. The backbone of the system is a cleverly designed mechanical system along with the embedded system resulting in accuracy. The system uses C# as a programming language and .NET platform for user interface.

ABS-44:

Characterization and Optimization of Automatic Image Registration Algorithms

T.Radha Krishna and Mahesh
MITS College Madanapall
vka4mahesh@yahoo.co.in

Abstract

In many image-processing applications it is necessary to register multiple images of the same scene acquired by different sensors, or images taken by the same sensor but at different times. Mathematical modeling techniques are used to correct the geometric errors like translation, scaling and rotation of the input image to that of the reference image, so that these images can be used in various applications like change detection, image fusion etc. In the conventional methods, these errors are corrected by taking control points over the image and these points are used to establish the mathematical model. The objective of this paper is to implement and evaluate a set of automatic registration algorithms to correct the geometric errors of the input image with respect to the reference image, by increasing the accuracy level of the registration and reducing the RMS error to less than a pixel. Various algorithms such as Wavelet transformation method, Fast Fourier transformation method, Morphological Pyramid approach and Genetic Algorithms are developed and compared. These algorithms are capable of considering the transformation model to sub-pixel accuracy. The benefits of these methods are accuracy, stability of estimation, automated solution and the low computational cost.

ABS-45:

Asynchronous Congestion Control in Multi-Hop Wireless Networks with Maximal Matching-Based Scheduling

R.Bhargavi and Sri G.V.R. Sagar.
GPREC, KURNOOL .
bhargavi_rallabandi@yahoo.co.in.

Abstract

We consider a multi-hop wireless network shared by many users. For an interference model that constrains a node to either transmit to or receive from only one other node at a time, and not to do both, we propose an architecture for fair resource allocation that consists of a distributed scheduling algorithm operating in conjunction with an asynchronous congestion control algorithm. We show that the proposed joint congestion control and scheduling algorithm supports at least one-third of the throughput supportable by any other algorithm, including centralized algorithms. Index Terms—Congestion Control, Fair Resource Allocation, Totally Asynchronous Algorithm, Distributed Scheduling, Wireless Networks.

ABS-46:

Hardware Synthesis of Modules Encountered in JPEG Algorithm

Prashob nair, ameya kuvelkar, shanmon philip, ravikrishna, havalikar
& prof. Nitesh n.naik
Goa college of engineering
Farmagudi-GOA

Abstract

The dependence on electronic devices in our daily lives and their interactive multimedia features makes data-compression techniques indispensable, for data-storage optimization (efficiency). JPEG algorithm is the most significant compression technique used for image processing applications in cell phones , PDAs, digital cameras etc. Hence digital circuit realization of such techniques has to be performed. In this paper we digitalize those subsystems of the algorithm which require more computations and which has challenging hardware synthesis then others. The digitalization of the subsystems is incorporated on a reconfigurable hardware in this case FPGA.

ABS-47:

Media processor architectures for video and Imaging on camera phones

Vidhya and Pallavi.S
ECE kalpataru Instiute Of Technology, Tiptur , Karnataka
guddie.vidhya1@gmail.com and pallavipinky.1@gmail.com

Abstract

A dedicated media processor is used in many camera phones to accelerate video and image processing. Increased demand for higher pixel resolution and higher quality image and video processing necessitates dramatically increased signal processing capability. To provide the increased performance at acceptable cost and power requires redesign of the traditional architecture. By wisely partitioning algorithms across programmable and fixed function blocks, the performance goals can be met while still maintaining flexibility for new feature requirements and new standards. In this paper we provide an overview of media processor architectures for camera phones and describe the system architecture, power, and performance. We also address the challenges in supporting new imaging trends and high resolution video at low power and cost

ABS-48:

Fault Secure Encoder and Decoder for Nanomemory Applications

C.Nalini & G.Sunil
SVPCET, R.V.S. Nagar,Puttur

Abstract

Memory cells have been protected from soft errors for more than a decade; due to the increase in soft error rate in logic circuits, the encoder and decoder circuitry around the memory blocks have become susceptible to soft errors as well and must also be protected .This paper propose a new approach to design fault-secure encoder and decoder circuitry for memory designs. The key novel contribution of this paper is identifying and defining a new class of error-correcting codes whose redundancy makes the design of fault-secure detectors (FSD) particularly simple.

ABS-49:

Cost-Efficient SHA Hardware Accelerators

Asiya sulthana[1], V.Sudheer raja[2],Praveen[3] and Shravan [4]
ECE Dept.Vaagdevi College of Engineering, Jawaharlal Nehru Technological University, Hyderabad.
asiya806@yahoo.co.in, sudheerrajav@yahoo.com, praveenandsuma@gmail.comand
adepushravan406@gmail.com

Abstract

Hash functions are among the most widespread cryptographic primitives, and are currently used in multiple cryptographic schemes and security protocols, such as IPSec and SSL. This paper presents a new set of techniques for hardware implementations of Secure Hash Algorithm (SHA) hash functions. These techniques consist mostly in operation rescheduling and hardware reutilization, therefore, significantly decreasing the critical path and required area. Throughputs from 1.3 Gbit/s to1.8 Gbit/s were obtained for the SHA implementations on a XilinxVIRTEX II Pro. Compared to commercial cores and previously published research, these figures correspond to an improvement in throughput/slice in the range of 29% to 59% for SHA-1 and 54% to 100% for SHA-2. Experimental results on hybrid hardware/software implementations of the SHA cores have shown speedups up to 150 times for the proposed cores, compared to pure software implementations.

ABS-50:

An Automated Microcontroller Based Liquid Mixing System

Dr.A.S.C.S.Sastry, Mr. K.N.H.Srinivas, Mr.CH.V.S.R.G.Krishna, & Mr. CH.S.Kiran Kumar
Kakinada institute of engineering& technology and Sri Vasavi engineering college
ascssastry@gmail.com, knh.tridents@gmail.com, Krish251@gmail.com, & kiranurs4ever@gmail.com

Abstract

Mixing of two or more liquids is somewhat difficult in many situations particularly in industries. Especially in such situations where the ratio of volume and temperature of liquids which are to be mixed has to be changed at regular instances. This becomes a hectic task to maintain and reconfigure entire system. This paper introduces a systematic approach to design and realize a dynamically reconfigurable liquid mixing system.

ABS-51:

Self-Checking Carry-Select Adder Design

J. Chandana Priya 2. Sarath Babu

Abstract

Arithmetic operations are frequently used in many VLSI-based systems. The design of faster and highly reliable adders is of major importance in such systems. Carry-select adders are one of the faster types of adders. This project paper proposes a scheme that encodes the sum of bits using two rail codes. Self-Checking Checkers checks the encoded sum bits. The multiplexers used in the adder are also totally self-checking. This scheme is illustrated with the implementation of a 2-bit Carry select adder that can detect all single stuck-at faults on line. The detection of double faults is not guaranteed. Adders of arbitrary size can be constructed by cascading the appropriate number of such 2-bit adders. A range of adders from 4 to 32 bits is designed using this approach employing a 0.12µm CMOS technology. The transistor overhead in implementing these self-checking adders varies from 14.27 % to 19.67 %, and the area overhead is 31.2 % Compared to the adders with out built-in self-checking capability.

ABS-52:

Design and Verification of Low Power SDRAM Controller

Vijaya Prakash.A.M, SindhuraPrakash, and Dr.K.S.Gurumurthy.
Dr.MGR University,Bangalore-04
drksgurumurthy@gmail.com

Abstract

The fast growth of mobile computing and pocket computer has increased the effect of energy management in hardware design greatly. Memories are such a common feature in any system that sometimes it is barely mentioned and memory chips occupy a greater part of power consumption in embedded system. This being the case, we try to present an effective power mode scheme for SDRAM controllers which reduce the power, making it suitable for energy sensitive applications. The problem of minimizing power may be seen as hardware or software optimizations. High-level language optimization also appears as an alternate technique to achieve low power consumption when coding embedded systems. The SDRAM controller architecture consists all basic and few advanced features of a controller, with added code optimization techniques. Four coding techniques are adopted in reducing the power consumed by the controller, Latch based clock gating for gating the flip flops, one hot encoding for sequential logic, gray code encoding for counters and Resource sharing to remove redundant logic elements. Minimizing the power contributes to maximizing the life of the battery, a major demand in portable electronic devices.

ABS-53:

Hardware Optimized Implementation of Low Pass, High Pass, Band Pass and Band Stop FIR Filters

M.Mohammed Irshad ,and D.V.Sri Hari Babu,
KTRMC of Engineering. Mahabubnagar, A.P
irshad_422@yahoo.com

Abstract

The Finite Impulse Response (FIR) filters are widely used in signal processing applications due to their stability and linear phase properties. The FIR filters are most commonly used in consumer devices

including, portable music and video players, mobile telephones, home theatre and docking stations to mention a few. Therefore, optimized implementation of FIR filters in hardware has gained significant attention for the researchers. The design techniques for minimizing the area, throughput and power are of paramount importance in the field of high speed FIR filter design for Application Specific Integrated Circuits (ASICs) and FPGAs. Previously researchers have focused on reducing area by replacing multiplier operations with addition, subtraction and shift operations. In this paper we propose a novel methodology to reduce the hardware required in the process of FIR filter design by using the quantization of filter coefficients.

Moment Based Fingerprint Matching

Bhagya lakshmi M[1], K.M.Sadyojatha[2], Juber M.A[3]
Dept. of Electronics and Communication Engg
Bellary Institute of Technology and Management, Bellary.
[1] bhagya.ch.26@gmail.com ,[2] saddukm@gmail.com, [3] ma.juber@gmail.com

AB-54

Abstract

Biometric recognition systems offer greater security and convenience than traditional methods of personal recognition. Along with the rapid growing of this emerging technology, the system performance, such as accuracy and speed, is continuously improved. At the same time, the complexity of the biometric system itself is increasing.

Fingerprint Identification is the method of identification using the impressions made by the minute ridge formations or texture patterns found on the fingertips. Fingerprints offer an infallible means of personal identification. A large number of fingerprint matching algorithms are proposed till date. These algorithms are mainly based upon image correlation, filter banks for feature extraction or texture descriptors.

In this work statistical approach of fingerprint matching is employed so that it is simple, cost effective, and computationally less complex. The fingerprint image which is extracted from a biometric device is divided into smaller areas say 3 x 3, 5x 5 etc. The size of this small area is selected heuristically. The background of this image (other than the image) intensity values are made zero. Then image matrix of nonzero elements is extracted. The Mean, Variance, Std, Mean (mean), Mean (Variance), Mean (Std), Variance (mean), Variance (Variance), Variance (Std), Std (mean), Std (Variance) & Std (Std) are calculated. Finally obtained results are analyzed to find similarities between different images. It is found from the study that it is possible to match and identify two fingerprints for similarities. Some of the sample results are presented in the text.

AB-55

Designing Real-Time and Embedded Systems with the COMET/UML Method

Sunilkumar H T, Vansanth Kumar.S ,sunilht1990@gmail.com,vassankit@gmail.com

Abstract

Most object-oriented analysis and design methods only address the design of sequential systems or omit the important design issues that need to be addressed when designing real-time and distributed applications [Bacon97, Douglas99,Selic94]. It is essential to blend object-oriented concepts with theconcepts of concurrent processing [MageeKramer99] in order to successfully design these applications.This paper describes some of the key aspects of the COMET method for designing real-time and embedded systems [Gomaa00], which integrates object-oriented and concurrent **processing** concepts and uses the UML notation [Booch98, Rumbaugh99]. Examples are given from an Elevator Control System.

Prioritization of Strategic Initiatives Using AHP

Mr.Vinay.S,
Assistant Professor
Department of ISE,
NMAMIT, Nitte.
vinaymanyan@gmail.com

Veena Desai
Senior Lecturer
Department of E&C
SJEC, Mangalore
veenadesai.2010@gmail.com

Doing the right things, the right way, right on target and achieving more with less requires formulating and deploying sound strategies. Today's fierce global competition demands excellence both in strategy and in its execution by senior management in order to meet the challenges of tomorrow. The act of balancing strategy and operations, and continual worry about the future, always push the top management to the helm. Unless there is a common and mutually agreed rational framework that helps align the various units at work in an enterprise with vision from conception till declaration of results, it is not possible to build long-lasting balanced organizations.

1. Introduction to Balanced Score board

Balanced scorecard (BSC)[1], originally developed in the early 1990s by Robert Kaplan and David Norton, is one such framework that helps achieve the required balance. It helps translate the strategy into actions from four perspectives:

- *Financial*: Traditional measures of profitability, revenue, and sales growth.
- *Customer*: Customer retention, customer satisfaction and market research.
- *Internal business processes*: Processes to meet or exceed customer expectation.
- *Learning and growth*: How the organization and its people grow and meet new challenges?

BSC is different from the traditional performance management system, because:

- It also includes non-financial measures for evaluating the overall performance of an organization.
- It brings in the concept of defining leading and lagging performance indicators/drivers to compare past and plan future performance targets.
- It includes indicators from both internal and external stakeholder perspectives and helps build a balance between them.
- It acts as an effective communication vehicle, translates the strategy into focused measurable actions and aligns the entire organization with the vision.

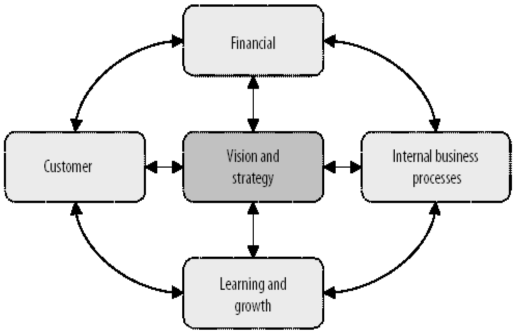


Figure1.1: Balanced Score Board

1.2. A framework to Align Strategic Initiatives with Vision

Major changes take a long time in big organizations and involve many people across the hierarchies. It is thus important to have a framework that fits the following description:

1. It is consistent, robust and stable. It helps the analyst remain focused on the implementation of any changes without getting deflected by internal or external chaos.
2. It evaluates strategies over its life cycle and realigns them whenever there is a shift.
3. It acts as a common communication vehicle across the enterprise from top to bottom.
4. It takes less effort, and enables quick and accurate delivery of key information while aligning various layers and locations of an organization on an end result.

The following steps describe a framework to align initiatives for the Educational Organization:

Step 1: Mission, vision and values are articulated and communicated across the organization to ensure the constant purpose of existence and progress towards excellence in performance. Mission describes the purpose of existence of an organization, vision defines the results to be achieved and values define the guiding principles or the code of ethics for the organization.

Step 2: Senior management assesses (a) the external environment (current and future market opportunities, competition etc.) and (b) the internal environment (current strengths and weaknesses), and brainstorms to formulate strategies. Strategy is the approach (based on the assessment) to accomplish the mission and implement the vision. Formulation of strategies is a step forward towards the execution of vision.

Step 3: Define strategic objectives that are measurable. Performance measures are assigned to these strategic objectives. Such measures represent the performance outcome objectively. Targets are set to drive the performance outcome of the organization. For instance, if the strategy is “learning and growth”, the performance measures could be (a) to assure the overall personality growth and (b) to enhance the quality level of the knowledge acquired. And targets could be (a) >90% (b) 100%. Tying performance measures to the objective is the most critical step towards organizational alignment. It is advisable to involve all stakeholders when the performance measures are being defined. It helps building buy-in and incorporating their feedback at an early stage. However, it is management’s responsibility to ensure consistency in the definition and deployment of the measures across the organization.

Step 4: Initiatives are identified, implemented and managed by metrics and targets to ensure that they are successful. The success of each initiative is measured based on the performance outcome. This performance outcome can be benchmarked against the best-in-class organizations.

Step 5: Initiatives are prioritized; resources are committed and run the initiatives. The initiatives are tracked on a periodic basis to check the progress and percentage results rate. The rewards and compensation are tied to the success of these initiatives to motivate the departments, teams and individuals. Implementing such a framework is itself a big change initiative. Having a framework that helps manage many important initiatives comes with its own overhead to maintain and manage a disciplined approach to decision.

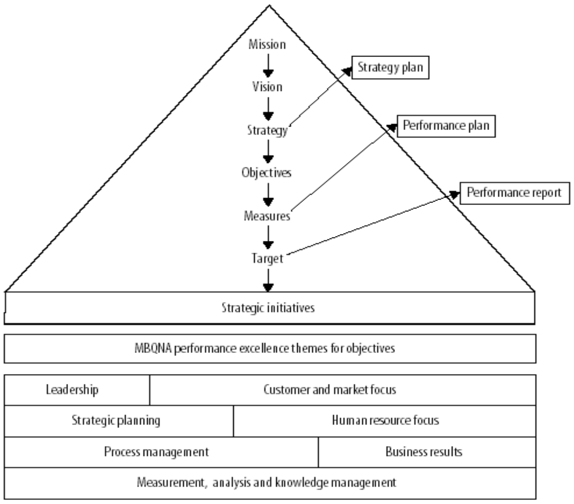


Figure 1.2: Aligning initiatives: a frame work

Step 6: Initiatives are prioritized; resources are committed and run the initiatives. The initiatives are tracked on a periodic basis to check the progress and percentage results rate. The rewards and compensation are tied to the success of these initiatives to motivate the departments, teams and individuals. Implementing such a framework is itself a big change initiative. Having a framework that helps manage many important initiatives comes with its own overhead to maintain and manage a disciplined approach to decision-making. Implementation of a framework is ineffective if the management culture is to keep rushing to judgments and bypassing the framework. As a practice, the framework is never rolled out across the organization until it gets deeply rooted into the management culture and significant signs of continuous maintenance and respect are visible. Only then can strict adherence to the framework by middle and line management be expected.

1.3. Prioritization of Initiatives

The AHP [5] has been applied in complex, real-world; multi-criteria decision making problems to evaluate strategic alternatives. Let us consider an example of Educational organization with major four objectives and a set of initiatives driving those objectives as shown in Table 1.1. The first step is to valuate how much these four objectives help to fulfill the vision considering the situational circumstancesof the organization. Paired comparisons of these objectives are obtained from the top management in terms of relative importance to the vision where objective “Business growth (O1)” is rated marginally strong when comparing it with the “Internal evaluation process (O2)” and rated very strong when comparing it with the “Learning satisfaction (O3)” . Referring to the Table 1.2 the comparison of object “Learning satisfaction (O3)” with the “Defined career path (O4)” is given a value 1/5 indicating the management vision is to provide students “Defined career path” which intern help in maintaining brand value of the organization.

Objectives	O1	O2	O3	O4	Local rating
O1	1	3	7	5	0.05415
O2	1/3	1	5	1/3	0.1662
O3	1/7	1/5	1	1/5	0.0497
O4	1/5	3	5	1	0.241

Table 1.2: Comparison Matrix of Objectives

Each of these objectives has 2–3 initiatives that have been identified by the management as shown in Table 1.1. The paired comparisons of these initiatives are shown in Tables 1.3, 1.4, 1.5 and 1.6 respectively, corresponding to objectives O1, O2, O3 and O4 based on their capability of execution to meet the targets.

Theme	Objectives	Measures	Target	Initiatives
Management / Financial Perspective	Business growth & Brand Equity (O1)	• Profitability • Advertise world wide	100% >80%	• Enter new geographies (I_{11}) • Publicize organization brand (I_{12}) • State of art facility (I_{13})
Faculty / Internal Process	Internal Evaluation Process (O2)	• Academic performance • Tracking extracurricular activity	>80% >20%	• Transparency in evaluation (I_{21}) • Motivating students (I_{22}) • Well defined evaluation process (I_{23})
Student / Learning & growth	Improved skill set with learning satisfaction (O3)	• Knowledge acquired • Overall personality growth	>80% >60%	• Strong academic base (I_{31}) • Interpersonal Skill (I_{32}) • Personal development opportunity (I_{33})
Parent / Return on investment	Well defined academics with defined career path (O4)	• Career in good organization • Responsible individuals	>90% 100%	• Soft skill training (I_{41}) • Improved academic performance (I_{42}) • Industry (I_{43}) Exposure

Table 1.1: Initiatives of Educational organization

Initiatives	I11	I12	I13	Local rating
I11	1	1/7	1/5	0.0738
I12	7	1	3	0.644
I13	5	1/3	1	0.2828

Table1.3: Comparison Matrix of Objective O1

In the Table 1.3 the initiative "Publicize organization brand (I12)" is rated very strong when comparing it with the "Enter new geographic location (I11)" and it is considered to be marginally strong when compared with "Investment in state of art facility (I 13)".

Initiatives	I21	I22	I23	Local rating
I21	1	3	1/3	0.333
I22	1/3	1	3	0.336
I23	3	1/3	1	0.333

Table 1.4 Comparison Matrix of Objective O2

In the Table 1.4 "Well defined evaluation Process (I23)"is considered to be marginally strong when compared with "Transparency in evaluation (I21)" indicating if the initiative I23 is satisfied it intern results in I21.

Initiatives	I31	I32	I33	Local rating
I31	1	7	1/3	0.3316
I32	1/7	1	1/5	0.080
I33	3	5	1	0.587

Table 1.5 Comparison Matrix of Objective O3

The Table1.5 is in favor of student where "Strong academic base (I31)" is considered to be very strong when compared with "Interpersonal Skill (I32)".

Initiatives	I41	I42	I43	Local rating
I41	1	1/7	1/5	0.071
I42	7	1	7	0.724
I43	5	1/7	1	0.2045

Table 1.6 Comparison Matrix of Objective O4

Referring to Table 1.5 "Personal development opportunity (I33)" is considered to be strong when compared with "Interpersonal Skill (I32)". When the initiatives are compared with itself it results in equal and the value 1 is given i.e. to the diagonal elements of the matrix.

Table 1.6 lists out comparison matrix of initiatives of parents view point with the education. Here the initiative "Improved academic performance (I42)"is considered to be very strong when compared with the "Soft skill training (I41)".

Final ranking of initiatives with global priorities with respect to the vision come out to be as shown in Table 1.7 . Global rating is then calculated by multiplying these local rating with their corresponding objective weights.

Objective	Initiatives	Local rating	Global Rating
O1: Business growth & Brand Equity (w1=0.5415)	I11: Enter new geographies I12: Publicize organization brand I13: State of art facility	0.0738 0.644 0.2828	0.0399 0.348 0.1531
O2: Internal Evaluation Process (w2= 0.1662)	I21:Transparency in evaluation I22: Motivating students I23:Well defined evaluation process	0.333 0.336 0.333	0.0553 0.0558 0.0553
O3: Improved skill set with learning satisfaction (w3=0.0497)	I31:Strong academic base I32:Interpersonal Skill I33:Personal development opportunity	0.3316 0.080 0.587	0.0165 0.0039 0.029
O4: Well defined academics with defined career path (w4=0.241)	I41:Soft skill training I42:Improved academic performance I43:Industry Exposure	0.071 0.724 0.2045	0.017 0.174 0.049

Table 1.7: Final ranking of Initiatives

1.4. Evaluation Summary

According to the above analysis, the organization needs to deploy its major effort in Publicizing organizational brand (global priority = 0.348), and then focus on Improved academics (global priority = 0.174) and on State of art facilities (global priority= 0.1531).

In the top level management the main strategy is to get "Return on Investment". The main contribution for this is from the international students and to do so establishing the brand value of the organization plays a major role. Establishing and publicizing the brand value can be done by providing good infra structure which is a basic need for any organization, along with

Improved academics which can be provided with the good quality staff and with global tie-ups with the industries which intern publicize the organization .

1.5. Conclusion

Most organizations face massive resistance in getting the objectives agreed and finalized for inclusion in the scorecard. Even after that, there is a lack of adequate buy-in of the decision, which tends to be considered personal and subjective. There is also a tendency to believe that there are hidden agendas, which are not discussed in the open. Apart from the difficulty in reaching a consensus, some of the other challenges include: defining performance measures objectively, funding long-term versus short-term initiatives and procedures for measuring the success of initiatives etc.

References

[1]. Saaty TL, Vargas LG (2001) “Models, Methods, Concepts and Applications of the Analytic Hierarchy Process”. Kluwer, Dordrecht.

[2] J. P. Cavano and J. A. McCall. A Framework for the Measurement of Software Quality. In *Proc. of ACM Software Quality Assurance Workshop*, pages 133–139, 1978.

[3] IEEE Recommended Practice for Software Requirements Specifications, 1998. IEEE Std. 830-1998.

[4] A. Davis. *Software Requirements*. Prentice-Hall, 1990.

[5] Navneet Bhushan, Kanwal Rai (2004),“Strategic Decision Making Applying the Analytic Hierarchy Process” Springer-Verlag London Limited.

ABS-56 LASER PACEMAKER: A LIGHT TO MOVE THE HEART

[1]Vishnu Karwa, [2]Naseeruddin
[1]AMIE(ECE), [2]Lecturer Dept of ECE
BITM, Bellary
[1]vishnukarwa@indiatimes.com

Infrared LASER pacemakers that can optically synchronize the beat of an embryonic heart shows great promise for developmental biology, and perhaps ultimately for use as a pacemaker in humans. Most people think of light as merely something to see with. However, a growing number of scientists are now harnessing the power of light to control biological processes, with light-sensitive probes such as caged compounds, photoactivatable green-fluorescent proteins or photoswitchable molecules being used to modify living samples in a controlled manner.

ABS-57 BRAIN FINGERPRINTING

Abhilash K K,
Ballari Institute of Technology & Management,
Bellary, Karnataka, INDIA
E-mail address: akksrk@gmail.com

The fundamental difference between the perpetrator of a crime and an innocent person is that the perpetrator, having committed the crime, has the details of the crime stored in his memory, and the innocent suspect does not. This is what Brain Fingerprinting testing detects scientifically, the presence or absence of specific information. Electronics is the main back bone of this new and very helpful concept in crime detection and in counter terrorism. Brain fingerprinting is based on finding that the brain generates a unique brain wave pattern when a person encounters a familiar stimulus Use of functional magnetic resonance imaging in lie detection derives from studies suggesting that persons asked to lie show different patterns of brain activity than they do when being truthful. Issues related to the use of such evidence in courts are discussed. The author concludes that neither approach is currently supported by enough data regarding its accuracy in detecting deception to warrant use in court. In the field of criminology, a new lie detector has been developed in the United States of America. This is called “brain fingerprinting”. This invention is supposed to be the best lie detector available as on date and is said to detect even smooth criminals who pass the polygraph test (the conventional lie detector test) with ease. The new method employs brain waves, which are useful in detecting whether the person subjected to the test, remembers finer details of the crime. Even if the person willingly suppresses the necessary information, the brain wave is sure to trap him, according to the experts, who are very excited about the new kid on the block.

ABS-58 VIDEO ADAPTATION OF THE JPEG 2000 WITH MJPEG

Sanjeevakumar Harihar., N. Manja Naik, Asst. Prof, Manjula.N.Harihar, Santosh Nejakar
Sanjeevkumar.harihar@rediffmail.com, manjunn3@yahoo.com, manjulaharihar@gmail.com

Introduction

The JPEG-E core is a JPEG encoder that forms a high performance solution for image and video compression applications. Probably the fastest core in market, the JPEG-E can encode over 30 frames/sec of 4:3 HDTV, 1440x1152, 4:2:0 even on FPGA devices. Compliance with the baseline ISO/IEC 10918-1 JPEG standard makes the JPEG encoder core ideal for any cross platform application such as consumer digital cameras, camcorders, copiers, printers, scanners and remote surveillance systems.